Available online at www.sciencedirect.com
ScienceDirect
Journal of Memory and Language 57 (2007) 210–231
Journal of Memory and Language
www.elsevier.com/locate/jml

Use of word length information in utterance planning ☆

Antje S. Meyer [a,*], Eva Belke [b], Christine Häcker [a], Linda Mortensen [c]

[a] *School of Psychology, Behavioural Brain Sciences Centre, University of Birmingham, Edgbaston, Birmingham, B15 2TT, UK*
[b] *Department of Linguistics, University of Bielefeld, Germany*
[c] *Psychology Department, Rice University, Houston, TX, USA*

Received 9 June 2006; revision received 20 October 2006
Available online 15 December 2006

Abstract

Griffin [Griffin, Z. M. (2003). A reversed length effect in coordinating the preparation and articulation of words in speaking. *Psychonomic Bulletin & Review, 10*, 603–609.] found that speakers naming object pairs spent more time before utterance onset looking at the second object when the first object name was short than when it was long. She proposed that this reversed length effect arose because the speakers' decision when to initiate an utterance was based, in part, on their estimate of the spoken duration of the first object name and the time available during its articulation to plan the second object name. In Experiment 1 of the present study, participants named object pairs. They spent more time looking at the first object when its name was monosyllabic than when it was trisyllabic, and, as in Griffin's study, the average gaze-speech lag (the time between the end of the gaze to the first object and onset of its name, which corresponds closely to the pre-speech inspection time for the second object) showed a reversed length effect. Experiments 2 and 3 showed that this effect was not due to a trade-off between the time speakers spent looking at the first and second object before speech onset. Experiment 4 yielded a reversed length effect when the second object was replaced by a symbol (x or $+$), which the participants had to categorise. We propose a novel account of the reversed length effect, which links it to the incremental nature of phonological encoding and articulatory planning rather than the speaker's estimate of the length of the first object name.

Keywords: Picture naming; Word length effects; Eye movements; Speech planning

Introduction

The present research concerns the way speakers coordinate their speech planning with the articulation of utterances. To produce even the simplest utterances, speakers must generate several types of plans determining the conceptual content of the utterances, their semantic and syntactic structure, and their morphological, phonological, and phonetic form (e.g., Levelt, 1989). Speakers, quite sensibly, plan successive utterance fragments sequentially. They often begin to speak as soon

☆ We thank Joana Cholin for providing us with the syllable frequency counts and Julia Holzgrefe for conducting Experiment 2. We are grateful to Zenzi Griffin, Steven Lupker, and Julio Santiago for their constructive comments on an earlier version of the paper. The research was supported by a grant from the Friedrich Ebert Stiftung to C. Häcker.
* Corresponding author. Fax: +44 121 414 4897.
E-mail address: a.s.meyer@bham.ac.uk (A.S. Meyer).

0749-596X/$ - see front matter
doi:10.1016/j.jml.2006.10.005

as they have a plan for the first utterance fragment and plan the remainder of the utterance while they are talking. This minimises their memory load and, since speech planning and articulation occur simultaneously, the time required for conveying a message.

Many studies have investigated the characteristics of the planning units speakers use (for a review see Smith & Wheeldon, 1999). One important result of this research is that speakers, instead of consistently using the same types of units, appear to have some flexibility in their choice of planning units (e.g., Ferreira & Swets, 2002; Schriefers & Teruel, 1999). Presumably this allows them to adjust to the communicative situation by using small planning increments when it is important to be quick to begin to speak and larger ones when it is important to produce utterances fluently. However, flexibility may come at a price: unless the criteria governing the choice of the planning units are quite simple, the time and processing resources required for choosing appropriate planning units might cancel out any benefits of flexibility.

The present research was largely motivated by an observation first made by Griffin (2003), which suggests that speakers carry out surprisingly complex computations in deciding when to initiate simple noun phrase conjunctions such as "cat, chair." Before turning to Griffin's results and our own experiments, we briefly review how speakers plan words and phrases and how they co-ordinate speech planning and articulation.

An overview of single word production

Most models of word production distinguish between the selection of word units (sometimes called lemmas) from the mental lexicon and the retrieval of the associated word forms (e.g., Garrett, 1980; Levelt, 1989). Evidence for this distinction comes from a variety of sources, including the occurrence of "tip of the tongue" states, where speakers have a strong feeling of knowing a word, have access to its meaning and syntactic properties but cannot retrieve the complete phonological form (e.g., Brown & McNeill, 1966), neuropsychological evidence for selective impairments of access to the semantic or phonological properties of words (e.g., Caramazza, Papagno, & Ruml, 2000; Dell, Schwartz, Martin, Saffran, & Gagnon, 1997; Foygel & Dell, 2000), and experimental evidence from healthy participants, demonstrating that the semantic and syntactic properties of words become available slightly before their phonological forms (e.g., Indefrey & Levelt, 2004; Jescheniak, Schriefers, Garrett, & Friederici, 2002; Schriefers, Meyer, & Levelt, 1990; Van Turennout, Hagoort, & Brown, 1998).

According to all models of word production, speakers generate the phonological forms of words out of phonological segments rather than retrieving them as units from the mental lexicon. Such a process of phonological encoding must be postulated to account for the occurrence of sound errors, such as "inner at date" (instead of "dinner at eight") and "stedal peel guitar" (instead of "pedal steel guitar," Fromkin, 1973, p. 247), where single segments or segment clusters are incorrect, and for the fact that the pronunciation of words in connected speech often differs from their citation forms. For instance, segments may be associated to new syllables (as in "demand it," spoken as "de.man.-dit"), they may be deleted (as in "got to" spoken as "go.to"), or changed (as in "hand bag" pronounced as "ham.bag"; see Levelt, Roelofs, & Meyer, 1999, for further discussion). Speakers can only generate these connected speech forms if individual segments are available as planning units.

In most models, metrical structures, which specify the stress patterns of words, are stored in the mental lexicon for all words, often as hierarchical structures consisting of stressed and unstressed syllables and syllable constituents (e.g., Dell, 1986; Stemberger, 1985). During phonological encoding, speakers retrieve segments and metrical structures and combine them by assigning segments to positions in metrical structures. However, all words can be syllabified by applying universal and language-specific syllabification rules, and in many languages stress can also be assigned by rule to the majority of the words. Therefore, Levelt (1992, see also Levelt et al., 1999) proposed that the lexical entries for words with regular stress patterns only specify the words' segments and their order, and that syllabification and stress assignment are done by applying appropriate rules. For words with irregular stress, the stress pattern is specified in the lexical entries.

Current models of word production describe the process of phonological encoding within spreading activation frameworks; i.e., they assume that the segments of a word receive activation from superordinate nodes and eventually become selected and associated to the metrical structure. The models differ in their assumptions about the time course of segmental activation, i.e., in whether segments within a syllable or a word are activated in parallel or sequentially (e.g., Dell, 1986; Dell, Burger, & Svec, 1997; Hartley & Houghton, 1996; Levelt et al., 1999; Roelofs, 1997a, 1997b; Sevald & Dell, 1994; Vousden, Brown, & Harley, 2000), but they all postulate that segmental activation is followed by a sequential segment selection or syllabification process, which proceeds from the beginning to the end of each word. Evidence supporting this sequentiality assumption comes from studies using different types of priming and interference paradigms (e.g., Cholin, Schiller, & Levelt, 2004; Meyer, 1990, 1991; Meyer & Schriefers, 1991; Roelofs, 1996, 1998, 2002, 2004; Sevald & Dell, 1994; for a review see Wilshire & Saffran, 2005).

The word form representation generated during phonological encoding consists of discrete and contextindependent phonological segments. It must be transformed into a phonetic representation, which determines the movements of the articulators. Crompton (1982) and Levelt (1992) proposed that speakers had access to a mental syllabary, which is a store of preassembled articulatory gestures for frequent syllables (see also Cholin, Levelt, & Schiller, 2006; Cholin et al., 2004; Levelt & Wheeldon, 1994). Low-frequency syllables are assembled out of the gestural scores corresponding to individual segments. In the models proposed by Levelt et al. (1999) and by Roelofs (1997a, 1997b), the mental syllabary is accessed during phonetic encoding. As soon as the first phonological syllable of a word has been generated, activation spreads from the phonological segments to all syllables that include these segments. An activated syllable is selected and the corresponding articulatory code is placed in an output buffer if the syllable includes the required segments in the appropriate order.

If phonological encoding encompasses a sequential component, it should take speakers longer to initiate the articulation of long words than of short words. Several studies have confirmed this prediction (e.g., Eriksen, Pollack, & Montague, 1970; Klapp, Anderson, & Berrian, 1973; Santiago, MacKay, Palma, & Rho, 2000; Wheeldon & Lahiri, 2002). However, when Bachoud-Lévi, Dupoux, Cohen, and Mehler (1998) asked speakers of French and English to name objects with monosyllabic or disyllabic names, they failed to obtain an effect of name length on the object naming latencies. Meyer, Roelofs, and Levelt (2003) asked Dutch speakers to name objects with monosyllabic or disyllabic names. They found shorter naming latencies for objects with monosyllabic names than for objects with disyllabic names when the two sets of objects were presented in separate test blocks, but not when the objects with long and short names were shown together in mixed blocks (as was done in the study by Bachoud-Lévi et al.). Bachoud-Lévi and colleagues suggested that speakers generated the phonological forms of the words sequentially, but initiated the pronunciation of the disyllabic words as soon as the phonological code for the first syllable had been retrieved (see also Schriefers & Teruel, 1999). Therefore, no word length effect was observed. However, other studies demonstrated that speakers usually generate the complete phonological code for at least one word before speech onset (e.g., Costa & Caramazza, 2002; Meyer, 1990; Meyer & Schriefers, 1991; Roelofs, 1998; Wheeldon & Lahiri, 1997). Therefore, Meyer et al. (2003) suggested that speakers always generated the complete phonological codes of monosyllabic and disyllabic words before speech onset, but that in the mixed blocks they were likely to begin to speak as soon as they had retrieved the articulatory code for the first syllable of the words. This allowed them to initiate the production of disyllabic object names as quickly as the production of monosyllabic words (see also Roelofs, 2002).

In sum, studies of single word production have demonstrated that speakers generate, rather than retrieve, the phonological forms of words, that they do this sequentially, proceeding from the beginning to the end of a word, and that they can initiate the articulation of a word before it has been completely planned on all levels.

Planning and initiating phrases and sentences

When speakers produce phrases or sentences, they must carry out the encoding processes described above for each content word and they must generate the syntactic and prosodic structure of the utterance (for a review see Bock & Levelt, 1994). How these planning processes are timed relative to each other and relative to the articulation of the utterance is not fully understood. However, analyses of speech errors, hesitations, and pauses and experimental studies have shown that speakers plan successive utterance fragments in sequence and that they often begin to speak as soon as they have planned the first part of the utterance. Which planning units speakers prefer at each level is not entirely clear, but there is strong evidence that they tend to use broad planning units, corresponding roughly to clauses, when they determine the content and syntactic structure of their utterances, and smaller units when they determine the phonological form of the utterance (e.g., Ferreira & Swets, 2002; Levelt, 1989; Schriefers, 1993; Smith & Wheeldon, 1999).

For obvious practical reasons, much of the experimental research into the production of multi-word utterances has concerned the generation of simple object descriptions. Several studies have investigated how speakers produce noun phrases such as "the car" or "the green car." Their results converge to show that speakers tend to retrieve all lemmas of a noun phrase before speech onset (e.g., Jescheniak, Schriefers, & Hantsch, 2003; Schriefers, de Ruiter, & Steigerwald, 1999). The evidence concerning the generation of the corresponding phonological representations is less clear. Schriefers and Teruel (1999) found that speakers often began to speak as soon as they had retrieved the first syllable of the first word of a noun phrase. By contrast, results obtained by Costa and Caramazza (2002; see also Miozzo and Caramazza, 1999) and by Jescheniak et al. (2003) suggest that all words of a complex noun phrase such as "the big car" are activated to some extent before speech onset, though the first word is likely to be more strongly activated than the following words.

Other studies have examined how speakers generate utterances referring to two or more distinct objects, such as "the cat and the chair." Meyer (1996) asked participants to name pictures of object pairs in noun phrase conjunctions such as "the cat and the chair" or in sentences such as "the cat is next to the chair." The pictures were accompanied by auditory distractor words that were semantically or phonologically related to the first or second object name or unrelated to both. The speech onset latencies were longer when the distractor was semantically related to the first or second noun than when it was unrelated to both, implying that both lemmas were selected before speech onset. However, in a follow-up study using very similar materials, a semantic interference effect was only obtained when the distractors were related to the first noun, but not when they were related to the second noun (Meyer, 1997). Though the reasons for the discrepancy between the results of the two studies are not clear, it can be concluded that speakers do not always select the second noun lemma in noun phrase conjunctions before speech onset. In Meyer's (1996) study, distractors that were phonologically related to the first object name sped the responses relative to unrelated ones, whereas distractors that were phonologically related to the second object slightly delayed the responses. These results are consistent with the proposal by Jescheniak et al. (2003) that before speech onset the phonological form of the first word is likely to be more strongly activated than the phonological forms of the following words.

In several recent multiple-object naming experiments, the participants' eye movements were recorded along with their speech. These studies have yielded a number of key findings (for reviews see Griffin, 2004; Meyer & Lethaus, 2004; see also Griffin & Bock, 2000): First, speakers typically fixate upon the objects they name in the order of mention. Given the strong link between eye gaze and visual attention (e.g., Deubel & Schneider, 1996; Irwin, 2004), this implies that speakers attend to the objects in the order of mention. Second, the speakers' eye gaze usually runs slightly ahead of their overt speech. When speakers name two or more objects, they typically look at the first object for about 500–700 ms, then start to inspect the second object, and then, about 150–350 ms later, initiate the name of the first object. Thus, speakers spend most of the time before speech onset looking at the first object, but they also spend some time looking at the second object.

Finally, the gaze duration for an object, i.e. the time speakers spend looking at it before turning to another object, depends on the total time they require to identify the object (Meyer, Sleiderink, & Levelt, 1998), to select its name (Griffin, 2001; Griffin & Oppenheimer, 2006), and to retrieve the corresponding morphological and phonological form (Griffin, 2001; Meyer & van der Meulen, 2000; Roelofs, 2007). Thus, speakers only initiate the shift of gaze to a new object after they have planned the name of the current object to the level of phonological form. Together with the results of the reaction time studies described above, these results suggest that speakers generate the names of the objects they refer to in noun phrase conjunctions in a highly sequential fashion (but see Morgan & Meyer, 2005) and begin to speak when they have a complete plan for the first object name but not necessarily for the second object name.

The present study

The study by Meyer et al. (2003) mentioned above, which investigated the effects of name length on object naming latencies, included an eye tracking experiment in which participants named object pairs in noun phrase conjunctions. The length of the name of the first object (the left one on the screen) was varied (e.g., "cat and chair" vs. "camel and chair"). The items were presented in pure blocks, in which all left objects had monosyllabic names or they all had disyllabic names, or in mixed blocks. The results for the speech onset latencies (i.e., the onset latencies of the name of the left object) closely matched those observed for the speech onset latencies in the single-object naming experiment: In pure blocks the mean speech onset latency was shorter when the first object had a short name than when it had a long name, but in mixed blocks no word length effect was found.

The analysis of the participants' eye movements showed that, as in earlier multiple-object naming experiments, the participants looked at the objects in the order of mention and that the shift of gaze from the first to the second object usually occurred slightly before speech onset. The length of the name of the left object and the presentation of the items in pure or mixed blocks had similar effects on the gaze durations for the left objects as on the speech onset latencies. In pure blocks participants looked longer at the left objects with long names than at the left objects with short names, but in mixed blocks the gaze durations for objects with long and short names did not differ from each other. The similarity of the results obtained for the speech onset latencies and gaze durations shows that the participants' decision when to begin to speak and the decision when to direct their gaze to a new object were governed by similar criteria (see also Roelofs, 2007; Roelofs & Lamers, 2007).

However, the temporal relationship between the shift of gaze from the left to the right object and the onset of speech was slightly different for utterances beginning with monosyllabic and disyllabic words: The time interval between the shift of gaze and the onset of speech (the gaze-speech lag, hereafter) was significantly longer (by 29 ms) when the first object had a short name than when it had a long name (see Fig. 1). In other words, the participants spent slightly more time before speech onset looking at the right object when the left object had a short name than when it had a long name.

A similar observation had been made earlier by Griffin (2003). She asked speakers of American English to

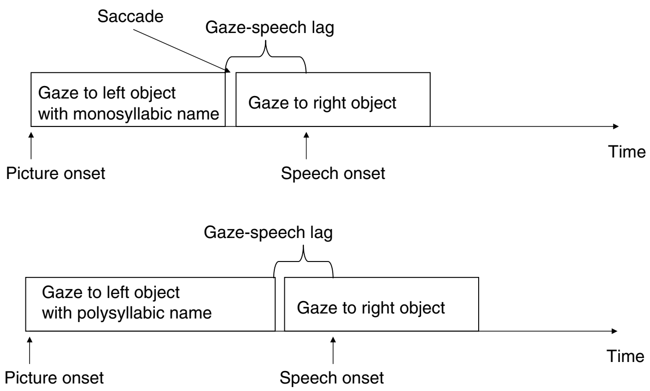

Fig. 1. Co-ordination of eye movements and speech onset for utterances beginning with monosyllabic and polysyllabic object names.

name pairs of objects in bare nouns. The name of the left object, which was to be named first, was monosyllabic or polysyllabic (e.g., "wig" or "windmill"). Monosyllabic and polysyllabic items were tested in mixed blocks. The gaze duration for the left object was not affected by the length of its name, paralleling the results obtained for mixed blocks by Meyer et al. However, when the first object name was monosyllabic, the speakers spent more time looking at the right object before they began to speak than when the first object name was polysyllabic.

Griffin's account of these findings, called the length monitoring hypothesis hereafter, was that speakers aimed to produce the utterances as fluently as possible and estimated how much time they would have to plan the second object name while articulating the first object name. When the first object name was short, speakers had less time to plan the second object name *after* speech onset, and therefore planed the second object name more extensively *before* speech onset than when the first object name was long. When the speakers named the object pairs in phrases, inserting "next to" between their names (e.g., "wig next to bear"), the utterances were initiated sooner, the right objects were inspected for a shorter time before speech onset than in the bare noun condition, and the pre-speech inspection times for the right objects did not depend any more on the length of the name of the left object. Griffin concluded that this pattern arose because the fixed insertion "next to" provided speakers with ample time to plan the name of the second object after speech onset.

These results are interesting and theoretically important because they imply that speakers can estimate the length of the first word of an utterance and use this information in deciding when to initiate an utterance and how extensively to plan the following words before speech onset. The length estimate must be based on properties of the phonological or phonetic representation of the first object name. Studies of speech monitoring have shown that speakers can access the phonological representation of words they are about to say. This sometimes allows them to interrupt themselves and to repair faulty utterance plans before they are executed (e.g., Hartsuiker & Kolk, 2001; Postma, 2000; Slevc & Ferreira, 2006). Given these findings, the availability of word form representations for the purposes of speech planning is not surprising. The novel observation is that speakers use this information to fine-tune the temporal co-ordination of their speech planning and the onset of articulation.

The goal of Experiment 1 of the present study was to replicate, with new materials and in a different language, the pattern of results seen in the blocked condition of the study by Meyer et al. (2003), i.e., the co-existence of a regular word length effect for the duration of the gazes to the left object (the first object to be named) and a reversed length effect for the gaze-speech lag. Given that this pattern had only been seen in one earlier study (Griffin did not find a word length effect on the left-object gaze durations), a replication seemed important. Replicating the effect of name length on the left-object gaze duration was of some interest in its own right because the effect would indicate, first, that the phonological encoding of polysyllabic words takes more time than the phonological encoding of monosyllabic words, and, second, that the time required to generate the phonological form of an object name is reflected in the duration of the speaker's gaze to the object. This would further support the view that speakers attend to the objects they name until they have retrieved the phonological form of their names.

Experiment 1 yielded the predicted results, i.e., a word length effect on the gaze durations for the left object and a reversed length effect on the gaze-speech lag. The length-monitoring hypothesis links the reversed length effect on the gaze-speech lag to the speakers' estimate of the spoken duration of the first object name. However, monosyllabic and polysyllabic words differ not only in their spoken durations but also in the complexity and duration of the phonological encoding processes, which was reflected in the gaze durations for the left objects. Therefore, it can be argued that the reversed length effect on the gaze-speech lag was not related to the time required to *say* the object names but to the time required to *plan* the names: The longer speakers need to plan the first object name, the less time they devote before speech onset to planning the second object name. This trade-off hypothesis was assessed in Experiments 2 and 3, in which the time required to plan the object names was varied while their length was held constant. In Experiment 4, a different task was used: Participants first named an object with a monosyllabic or trisyllabic name and then determined whether a symbol shown to the right of the object was the letter *x* or a plus sign. They had to fixate upon the symbol to identify it. In this task, the gaze duration for the objects should depend on the length of their names, as in the dual-object naming task. However, if the reversed length effect on the gaze-speech lag arises because speakers aim to produce their utterances as fluently as possible, this effect should be absent since there is only one object to name on each trial.

Experiment 1

The experiment included three sub-experiments. In the first sub-experiment, a word-picture matching task was used to determine whether the objects with monosyllabic and trisyllabic names were similar in ease of object recognition (for the use of this task see also Jescheniak & Levelt, 1994; Santiago et al., 2000). In the second sub-experiment, participants practiced naming the objects individually. This also allowed us to examine whether there would be a word length effect for the single-object naming latencies. Finally, in the third sub-experiment, the participants named object pairs. The left object, which had to be named first, either had a monosyllabic or a trisyllabic name. The name of the right object was always disyllabic. The displays featuring objects with monosyllabic and trisyllabic left objects were tested in separate blocks.

Based on the results obtained by Meyer et al. (2003), we expected that the participants would inspect the two objects in the order of mention, that they would look longer at the left objects with trisyllabic names than at those with monosyllabic names, that they would usually begin to say the name of the left object slightly after the shift of gaze to the right object, and, finally, that the gaze-speech lag would be shorter when the left object had a long name than when it had a short name.

Method

Participants

The participants in all experiments were undergraduate students of the University of Birmingham. They were native speakers of English and completed the experiment for course credits or payment. Experiment 1 was carried out with 20 participants.

Apparatus

The experimental software package NESU (Max Planck Institute for Psycholinguistics, Nijmegen) was used to display the stimuli and record the speech onset latencies. The stimuli were presented on a 19 in. Samtron 95P Plus colour monitor. The participants' speech was recorded using a Sony ECM-MS907 microphone and a Sony TCD-D8 DAT recorder. Speech onset latencies were measured using a voice-key (Hasomed GmbH, Magdeburg, Germany). Eye movements were recorded using an SMI EyeLink 2D head-mounted eye-tracking system. The system estimates the positions of both eyes every four milliseconds with a spatial accuracy of about 0.1° of visual angle.

Materials

We used line drawings of 18 objects with monosyllabic names, 18 objects with disyllabic names, and 18 objects with trisyllabic names. The pictures stemmed from Snodgrass and Vanderwart (1980) and from a picture gallery available in the first author's lab. Sixteen items of each length were used on experimental trials and two on practice trials (see Appendix A for a listing of the experimental items). There were equal numbers of object names starting with vowels, plosives, and fricatives in the monosyllabic and trisyllabic set. The three sets were matched as closely as possible for mean name frequency (means: 16.39 ($SD = 11.87$), 11.45 ($SD = 7.29$), and 10.10 ($SD = 7.29$) occurrences per million words in the COBUILD data base for the monosyllabic, disyllabic, and trisyllabic set, respectively), but the frequency difference between the monosyllabic and trisyllabic sets approached significance ($F(1,30) = 3.41$, $p < .10$, 95-%-confidence interval: CI = 6). The pictures were scaled to fit into frames of 6 by 6 cm, corresponding to 5.7° by 5.7° of visual angle when viewed from the participant's position.

Since name agreement norms were not available for all of the drawings, a norming study was carried out with 30 participants. They received a booklet showing the items in a random order and were asked to write down the first name that came to mind for each object. Name agreement was 87% ($SD = 17\%$), 96% ($SD = 11\%$), and 85% ($SD = 18\%$) for the items with monosyllabic, disyllabic, and trisyllabic names, respectively. The 2%-difference between the monosyllabic and trisyllabic items set was not significant ($F(1,15) < 1$, CI = 11%).

For the object recognition task, the objects were presented individually, in the centre of the screen. They were preceded by printed words (shown in font Arial, 36 point, lower case) likewise centred on the screen. For the single-object naming task, the objects were shown individually, centred as before. For the dual-object naming task, object pairs were presented, one centred in the left and one in the right half of the screen. The midpoint-to-midpoint distance between the objects was 19 cm (18° of visual angle).

Design

In the object recognition sub-experiment, each picture was presented twice, once preceded by its name (matching condition, requiring a *yes*-response) and once preceded by an unrelated word of the same length, which was the name of another experimental item (mismatching condition, requiring a no-response). The monosyllabic, disyllabic, and trisyllabic items were tested in separate blocks. Each item appeared in two successive blocks, once in the matching and once in the mismatching condition. Each block began with four practice trials and included the same number of matching and mismatching trials. Ten participants saw the monosyllabic items first, then the trisyllabic items, and finally the disyllabic ones. For the remaining ten participants the order of testing the monosyllabic and trisyllabic items was reversed. The disyllabic items, which were used as right objects in the dual-object naming task, were always tested last. They were only included in the object recognition and single-object naming sub-experiment to familiarise the participants with the complete set of materials. Within each group of participants, the order of testing the two blocks of items of the same length alternated. Each participant saw the items in a different random order.

In the single-object naming sub-experiment the participants saw the same objects as in the object recognition sub-experiment but without any preceding words. The assignment of objects to blocks and the order of the blocks for each participant was the same as in the object recognition sub-experiment, but a different random order of the items within blocks was used.

In the dual-object naming sub-experiment, the participants saw object pairs. An object with a monosyllabic or trisyllabic name was presented on the left side of the screen and an object with a disyllabic name on the right side. Each object with a monosyllabic or trisyllabic name was combined with two semantically and phonologically unrelated items from the disyllabic set (e.g., "axe-candle" and "axe-orange"), and each disyllabic item was combined with two monosyllabic and two trisyllabic items.

Each object pair was shown twice. There were four test blocks featuring left objects with monosyllabic names, followed or preceded by four blocks featuring left objects with trisyllabic names. For each participant, the order of testing the items with monosyllabic and trisyllabic names was the same as in the object recognition and single-object naming sub-experiment. In each block, each monosyllabic or trisyllabic left object was tested once, in the first and third block in combination with one of the associated right objects and in the second and fourth block in combination with the other right object. The order of the items within blocks was random and different for each block and participant.

Procedure

Participants were tested individually in a quiet room. They completed the three sub-experiments in a single session. At the beginning of each trial of the object recognition sub-experiment a fixation mark was shown in the centre of the screen for 800 ms followed by a blank interval of 300 ms. Next, a word was presented for 1000 ms, followed first by a blank interval of 200 ms and then by a picture, which was shown for 800 ms. The next trial began after a blank interval of 700 ms. Participants had to indicate as quickly as possible whether or not the word corresponded to the picture name. They pressed the right button for a *yes*-response, and the left button for a *no*-response.

In the single-object naming sub-experiment, the fixation mark and blank interval were directly followed by a picture, which was shown for 1000 ms. The participants were asked to name the objects as quickly as possible, using bare nouns (e.g., "bicycle").

Before the beginning of the dual-object naming sub-experiment, the head band of the eye tracker was placed on the participant's head and the system was calibrated. The system was recalibrated after each block. The participants were asked to name the object pairs as quickly as possible using bare nouns (e.g., "bicycle, dragon"). They were specifically instructed to speak fluently and to avoid pausing between the two object names. The trial structure was the same as in the single-object naming sub-experiment, except that the fixation mark appeared in the centre of the left half of the screen and that the pictures were presented for 2200 ms.

Analysis of eye movements

The EyeLink software determines the average position and duration of fixations between saccades. We used the Cognitive Parsing algorithm of the software package, which defines saccades as eye movements covering a minimum of 0.15° of visual angle at a minimum velocity of 30°/s with an acceleration of minimally $8000°/s^2$. The data from the right eye were analysed.

The participants were instructed to look at the fixation mark at the beginning of each trial. Sometimes the first fixation of a trial was located slightly below or to the left or right of the fixation mark. Typically, this was seen on several successive trials. We corrected for such drifts by manually aligning the first fixation with the fixation mark. The positions of the remaining fixations of the trial were recomputed accordingly. Fixations were categorised as being on one of the objects when they fell within a virtual frame of 6 by 6 cm (5.7° by

5.7°) enclosing the object. We determined the order in which the participants inspected the two objects, the first-pass gaze duration for the left object and the gaze-speech lag. The first-pass gaze duration (gaze duration hereafter) was defined as the time interval between the onset of the first fixation on the left object and the offset of the last fixation before the shift of gaze to the right object. The gaze-speech lag was defined as the time interval between the end of the first-pass gaze to the left object and the onset of speech (i.e., the onset of the name of the left object).

Statistical analyses

Separate statistical analyses were carried out for each sub-experiment. Analyses of variance were conducted on the error rates (after arcsine-transformation, see Winer, Brown, & Michels, 1991), the object recognition latencies, the speech onset latencies for the single- and dual-object naming, and the left-object gaze duration and gaze-speech lag of the dual-object naming sub-experiment. We report F_1, using participants as random variable, F_2, using items as random variable, and $minF'$ (Clark, 1973). Variability is reported in 95-%-confidence interval half-widths (CI) based on the mean squared error of the relevant comparison from the participant analysis (Masson & Loftus, 2003).

Results

Object recognition

Table 1 displays the mean response latencies and error rates for the pictures with monosyllabic, disyllabic, and trisyllabic names in the matching and mismatching condition. Only the items with monosyllabic and trisyllabic names were included in the statistical analyses because these items were used as left objects in the dual-object naming task, whereas the objects with disyllabic names were merely used as right objects.

The rate of missing and incorrect responses was low (2.64% of the responses) and did not differ significantly across the experimental conditions. The responses were significantly faster (by 46 ms) on matching than on mismatching trials ($F_1(1,19) = 19.83$, $F_2(1,30) = 16.63$, $minF'(1,48) = 9.04$, all $p < .01$, CI = 17 ms). The main effect of word length was significant in the analysis by items only ($F_1(1,19) = 2.09$, $F_2(1,30) = 6.25$, $p < .05$, $minF'(1,32) = 1.57$, CI = 22 ms), with the latencies being longer (by 23 ms) for monosyllabic than for trisyllabic items. The interaction of word-picture match and length was not significant.

Table 1
Mean error rates (%), response latencies (ms), left-object gaze durations (ms), and gaze-speech lags (ms) in Experiment 1

	Type of object name		
	Monosyllabic	Trisyllabic	Disyllabic
Object recognition			
Error rate			
Word-picture match	2.19	1.25	2.50
Word-picture mismatch	4.38	2.81	0.90
Recognition latency			
Word-picture match	545	528	571
Word-picture mismatch	597	570	540
Single-object naming			
Error rate	14.69	11.56	8.91
Object naming latency	619	655	637
Dual-object naming			
Error rate	8.51	7.92	
Speech onset latency	792	795	
Left-object gaze duration	532	608	
Gaze-speech lag	260	186	

Single-object naming

Incorrect and missing object names, responses beginning with filled pauses or hesitations, and responses with latencies exceeding 1800 ms were categorised as errors. The error rates for monosyllabic and trisyllabic items did not differ significantly. In contrast to the results found for the word-picture matching task, the mean response latency was significantly shorter, by 36 ms, for monosyllabic than for trisyllabic items ($F_1(1,19) = 5.43$, $p < .05$, $F_2(1,30) = 4.78$, $p < .05$, $minF'(1,48) = 2.54$, CI = 23 ms, see Table 1).

Dual-object naming

Utterances in which one of the two objects was named incorrectly, utterances which included an audible pause or hesitation, and utterances with latencies exceeding two seconds were categorised as errors. The error rates were lower than in the single-object naming task, most likely because the participants had already named the objects several times. The error rates for utterances beginning with monosyllabic and trisyllabic object names were very similar (see Table 1).

On most of the trials, the participants first fixated upon the left and then upon the right object. We excluded 17 trials on which they only fixated upon the left object and 20 trials on which they did not fixate upon the two objects in the order of mention. Six additional trials were excluded because of technical problems. In total (including the naming errors mentioned earlier) 10.30% of the data from the monosyllabic condition and 9.54% of the data from the trisyllabic condition were excluded from the further analyses.

For the remaining trials we determined the speech onset latency, the gaze duration for the left object, and the gaze-speech lag. The mean gaze duration was significantly longer, by 76 ms, for left objects with trisyllabic names than for left objects with monosyllabic names

($F_1(1,19) = 20.33$, $F_2(1,30) = 19.24$, $minF'(1,48) = 9.88$, all $p < .01$, CI = 25 ms). By contrast, the gaze-speech lag was significantly *shorter*, by 74 ms, when the left object had a trisyllabic name than when it had a monosyllabic name ($F_1(1,19) = 36.09$, $F_2(1,30) = 53.20$, $minF'(1,42) = 21.50$, all $p < .01$, CI = 18 ms). Consequently, the mean speech onset latencies (which correspond to the means of the sum of left-object gaze duration and gaze-speech lag for each trial) for monosyllabic and trisyllabic items differed by only 3 ms (all $F < 1$, CI = 24 ms).

Meyer et al. (2003) reported first-pass gaze durations and gaze-speech-lags, as we do here, whereas Griffin (2003) reported pre-speech gaze times for the left and right object, which were defined as the ''amount of time that speakers spent gazing at the objects before speech onset'' (p. 606). The pre-speech gaze time for the left object included the duration of first-pass gazes and the duration of any later gazes to the left object occurring before speech onset. In the present experiment, the pre-speech gaze durations for left objects with monosyllabic and trisyllabic names (528 vs. 600 ms) were very similar to the first-pass gaze durations (532 vs. 608 ms). This is because the participants' gaze rarely (on 1.7% of the trials) returned to the left object before speech onset and because the first-pass gaze to the left object usually (on 94.3% of the trials) ended before speech onset. The trial-by-trial correlation between first-pass gaze duration and pre-speech gaze time was $r = .96$ ($df = 2282$).

The gaze-speech lag was longer (by 77 ms) than the pre-speech gaze time for the right object because the gaze-speech lag included the duration of the saccade from the left to the right object. The trial-by-trial correlation between the gaze-speech lag and pre-speech gaze time for the right object was $r = .95$ ($df = 2282$). The size of the reversed length effect was almost identical for both variables (74 and 73 ms).

Discussion

In the single-object naming task, the speech onset latencies were longer for objects with trisyllabic names than for objects with monosyllabic names, supporting the prediction made by current models of word production (e.g., Levelt et al., 1999) that trisyllabic words should take longer to plan than monosyllabic ones. In the dual-object naming task, longer gaze durations were obtained for the left objects with long names than for those with short names. For the gaze-speech lag a *reversed* length effect was observed. When the left object had a long name, participants spent less time before speech onset looking at the right object than when the left object had a short name. Since the regular word length effect on the gaze durations and the reversed length effect on the gaze-speech lag had almost the same absolute size, the speech onset latencies for utterances beginning with long and short object names were very similar.

One account of these findings is that speakers anticipated that they would have less time to prepare the name of the right object after speech onset when the name of the left object was short than when it was long and adjusted the pre-speech planning of the name of the right object accordingly. However, the left objects with long and short names differed not only in the time required to *articulate* their names, but also in the time required to *plan* their names. This was evidenced by the difference in speech onset latencies in the single-object naming task and by the difference in left-object gaze durations in the dual-object naming task. Most likely these effects arose because the object names differed in length, though the difference in the frequency of their names may also have contributed to the effects. Because the gaze durations for the left objects with long and short names were different, an alternative account of the difference in the gaze-speech lags suggests itself: It could be due to a trade-off between the times speakers devoted to planning the names of the first and second object before speech onset. The more time they needed to retrieve the name of the first object, the less time they spent before speech onset planning the name of the second object. Such a trade-off relationship could arise if speakers aimed to initiate their utterances before some self-imposed response deadline.

This trade-off hypothesis predicts that any variable, not just name length, that affects the time required to retrieve the name of the left object should affect the gaze-speech lag (in the opposite direction) as well. We re-examined results from earlier experiments to determine whether this was true, but the evidence was inconclusive. Therefore, Experiment 2 was specifically designed to test the trade-off hypothesis. This was done by varying the age of acquisition of the names of the left objects instead of their length.

Experiment 2

Experiment 2 included the same tasks as Experiment 1. We selected 16 objects with early acquired names and 16 objects with late acquired names and tested them first in a word-picture matching task, then in a single-object naming task, and finally in a dual-object naming task.

The age of acquisition (AoA) affects the speed of responding to words and of producing them in many tasks, including picture naming. In word naming and lexical decision, the effects of word frequency and AoA are often similar in size and highly correlated, suggesting a shared basis for both effects (e.g., Lewis, Gerhand, & Ellis, 2001). In picture naming, AoA effects are typically stronger than predicted on the basis of the frequency effects for the items, suggesting that there is a frequency-independent AoA effect (e.g., Barry, Hirsh, Johnston, & Williams, 2001; for a review see Brysbaert & Ghyselinck, 2006). Different accounts of AoA effects in picture naming have been proposed (for reviews see Johnston & Barry, 2006; Juhasz, 2005), allocating them at the conceptual-semantic level (Brysbaert, Van Wijnendaele, & De Deyne, 2000, but see Izura & Ellis, 2004) or the phonological level (e.g., Brown & Watson, 1987, but see Monaghan & Ellis, 2002). Belke, Brysbaert, Meyer, and Ghyselinck (2005) proposed that the frequency-independent AoA-effect observed in picture naming arises during lexical selection, i.e., when speakers select a lexical unit from a set of competing candidates.

Based on the existing evidence we expected shorter single-object naming latencies and shorter gaze durations for the objects with early acquired names than for those with late acquired names. The most important question was whether there would be a reversed AoA effect on the gaze-speech lag in the dual-object naming task, as predicted by the trade-off hypothesis.

Method

Participants

The experiment was conducted with 20 participants.

Materials

The same 16 objects with disyllabic names were used as in Experiment 1, but the objects with monosyllabic and trisyllabic names were replaced by objects with early acquired and late acquired names. We selected the new items using the naming norms collected by Morrison, Chappell, and Ellis (1997). The early acquired items had a mean age of acquisition (measured as the age in months at which 75% of the children could name the picture) of 38.5 months ($SD = 9.9$ months), whereas the late acquired items had an average age of acquisition of 103.5 months ($SD = 19.5$ months, $F(1,30) = 140.36$, $p < .01$, CI = 18.2 months). Thirty items were disyllabic and two were trisyllabic. The average number of segments in the early and late acquired set was the same (5.1 segments). The items were matched pairwise for their onset segments. They were also matched for average name agreement (90% ($SD = 10\%$) for both groups), object familiarity (3.3 ($SD = 1.0$) and 2.5 ($SD = 0.7$) for early and late acquired items) and visual complexity (3.0 ($SD = 1.0$) and 3.0 ($SD = 0.8$)). AoA correlated positively with name frequency ($r = .37$, $p < .05$, $n = 32$), and the items in the early acquired set had a higher average word frequency than those in the late acquired set (32.1 ($SD = 44.5$) vs. 10.4 ($SD = 9.5$) occurrences per million words in the COBUILD data base for early and late acquired names, respectively; $F(1,30) = 3.63$, $p < .07$). This confound is not problematic because our goal was not to test claims about specific effects of AoA or frequency, but merely to create a relatively difficult and an easier object set.

Apparatus, design, and procedure

The apparatus, design, and procedure were the same as in Experiment 1, except that the sets of objects with early and late acquired names replaced the sets with monosyllabic and trisyllabic names.

Results

Object recognition

The error rates were low and very similar for early and late acquired items (means: 2.1 vs. 3.0%). The response latencies were significantly shorter (by 58 ms) for matching than for mismatching word–picture pairs ($F_1(1,19) = 30.79$, $p < .01$, $F_2(1,30) = 26.08$, $p<.01$, $minF'(1,48) = 14.12$, $p < .05$, CI = 22 ms, see Table 2). The mean response latency was shorter by 24 ms for early acquired than for late acquired names, but this difference was not significant ($F_1(1,19) = 1.88$, $F_2(1,30) = 3.50$, $p < .10$, $minF'(1,39) = 1.22$, CI = 39 ms).

Single-object naming

In the single-object naming task, the error rates for early and late acquired items did not differ significantly from each other. As expected, the mean response latency was shorter, by 86 ms, for objects with early acquired names than for objects with late acquired names ($F_1(1,19) = 91.59$, $F_2(1,30) = 20.40$, $minF'(1,42) = 16.68$, all $p < .01$, CI = 14 ms).

Table 2
Mean error rates (%), response latencies (ms), left-object gaze durations (ms), and gaze-speech lags (ms) in Experiment 2

	Type of object name		
	Early acquired	Late acquired	Disyllabic
Object recognition			
Error rate			
Word-picture match	3.12	4.28	1.67
Word-picture mismatch	1.08	1.63	2.20
Recognition latency			
Word-picture match	508	531	491
Word-picture mismatch	564	591	553
Single-object naming			
Error rate	4.22	6.41	3.91
Object naming latency	640	726	647
Dual-object naming			
Error rate	7.19	8.82	
Speech onset latency	791	818	
Left-object gaze duration	551	584	
Gaze-speech lag	235	233	

Dual-object naming

The error rates for utterances beginning with early acquired and late acquired names did not differ significantly from each other. As in Experiment 1, the participants usually looked first at the left object and then at the right object. We excluded 44 trials on which participants only fixated upon the left object and nine trials on which they did not fixate upon the two objects in the expected order. Overall, 8.83% of the data from the early acquired condition and 11.33% of the data from the late acquired condition were excluded from further analyses.

The mean gaze duration was shorter by 33 ms for objects with early acquired names than for objects with late acquired names. This difference was significant in the analysis by participants ($F_1(1,19) = 8.41$, $p < .01$) and approached significance in the analysis by items ($F_2(1,30) = 3.07$, $p < .10$, $minF'(1, 47) = 2.25$, CI = 14 ms). The mean gaze-speech lag for the two types of utterances differed only by two milliseconds (all $F < 1$, CI = 13 ms). The speech onset latency was shorter by 27 ms for the utterances beginning with early acquired names than for the utterances beginning with late acquired names. This difference was significant in the analysis by participants ($F_1(1,48) = 7.62$, $p < .05$) and approached significance in the item analysis ($F_2(30) = 3.37$, $p < .08$, $minF'(1,29) = 2.34$, CI = 14 ms).

Discussion

The object recognition latency was shorter for the objects with early acquired names than for those with late acquired names. Though this difference was not significant, it suggests that the two item sets were not perfectly matched for ease of object recognition. For the present purposes this is of little consequence because the goal of the experiment was not to investigate the origin of age of acquisition effects, but to use differences in age of acquisition to create a set of objects that could be named relatively quickly and a set of objects that would be slower to name. The single-object naming latencies showed that we accomplished this aim.

In the dual-object naming task, the left-object gaze duration and the speech onset latency were shorter in the early acquired than in the late acquired set, reflecting the difference in the time required to retrieve the early vs. late acquired object names, and perhaps a difference in the time required to recognise the objects. By contrast, the gaze-speech lag for the two item sets was almost identical. This pattern does not support the trade-off hypothesis discussed above, according to which any effect on the left-object gaze duration should be accompanied by an inverse effect on the gaze-speech lag.

However, the age of acquisition effect on the gaze durations was small (28 ms) and not significant in the item analysis, perhaps because the participants had already seen and named the objects several times in the preceding blocks of the experiment (e.g., Barry et al., 2001; Belke et al., 2005). The word length effect on the gaze durations in Experiment 1 was much stronger (76 ms). Therefore, one could conclude that only large differences in left-object gaze durations are accompanied by reversed differences in gaze-speech lag. The goal of Experiment 3 was to assess this possibility by inducing more substantial differences in the left-object gaze durations.

Experiment 3

Participants saw triplets of objects arranged in a triangle (see Fig. 2) and named them in the order left–right–bottom object. As in the preceding experiments, the main dependent variables were the speech onset latency, the left-object gaze duration, and the gaze-speech lag. For reasons which are not related to the concerns of the present paper, we were also interested in comparing the gaze durations and gaze-speech lags for objects named as first vs. second object in longer utterances. This is why object triplets were used.

The objects were either presented as intact line drawings or in a degraded version, in which approximately 50% of the pixels of their contours were deleted. Contour deletion should affect the ease of object recognition and this should be reflected in the duration of the speakers' gazes to the objects (Meyer et al., 1998). Each object triplet was accompanied by an auditory distractor word that was phonologically related or unrelated to the name of the left object (e.g., "birch/church" or "task/church"). Compared to the unrelated distractors, the related distractors should facilitate the phonological encoding of the object names (e.g., Damian & Martin, 1999; Schriefers et al., 1990; Wilshire & Saffran, 2005) and therefore the gaze durations for the left objects should be shorter in the related than in the unrelated condition (Meyer & van der Meulen, 2000). By combining contour deletion and phonological priming, we expected to create differences in left-object gaze durations between conditions that would be at least as large as the effect of name length

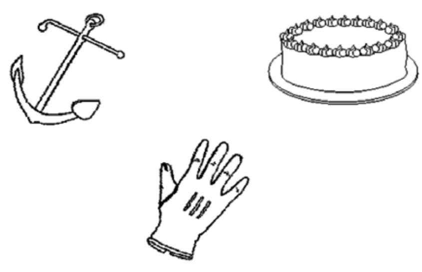

Fig. 2. Arrangement of objects in Experiment 3.

in Experiment 1. The most important question was whether contour deletion and phonological priming would also affect the gaze-speech lag.

Method

Participants

The experiment was carried out with 16 participants.

Apparatus

The same set-up was used as in Experiments 1 and 2. The auditory stimuli were presented using Sony MDR-E819 earphones.

Materials

On each trial we presented line drawings of three objects and an auditory distractor word. There were 55 experimental and four practice items. The length of the object names varied from one to three syllables (see Appendix A for a listing). The names of the three objects shown together were semantically and phonologically unrelated. The drawings were scaled to fit into frames of 7 by 7 cm (6.7° by 6.7°) and were arranged as shown in Fig. 2. The midpoint-to-midpoint distance between the objects was 17 cm (16° of visual angle). In the degraded version 50% of the contours of the objects were made invisible by a superimposed diagonal grid, which had the same light grey colour as the background.

For each object triplet a distractor word was selected that was phonologically related to the name of the left object, which the participants named first. Target and distractor had the same number of syllables. Monosyllabic targets and distractors shared the vowel and the following consonant or consonant cluster (e.g., "bag–rag"). Disyllabic and trisyllabic target–distractor pairs shared at least the final syllable and some pairs shared all segments except the initial consonant or consonant cluster (e.g., "finger–singer"). The distractor words were spoken by a female native speaker of English with no obvious regional accent.

The mean length of the auditory distractors was 736 ms ($SD = 123$ ms). Following Meyer and Schriefers (1991), they were presented slightly before picture onset (mean SOA = −207 ms, $SD = 113$ ms) such that the first segment that the distractor shared with the target began at picture onset.

Design

The experiment included four experimental conditions corresponding to the combinations of object quality (intact or degraded) and distractor type (related or unrelated). In the related condition, we used the target–distractor pairs described above. In the unrelated condition, the same distractor words were used, but they were assigned to new, unrelated targets.

The effects of degradation and priming were tested within participants and within items. Four experimental blocks were presented, each of which included all practice and experimental items. In two successive blocks, the pictures were shown in the intact version, and in two blocks they appeared in the degraded version. The order of testing degraded and intact objects was counterbalanced across participants.

Within each block, each object triplet appeared once. Half of the items within each block were tested in the related and half in the unrelated condition. Each item was tested twice with a related distractor (once in the intact and once in the degraded version) and twice with an unrelated distractor. Each participant saw the items in a different random order.

In addition to the experimental blocks there was a practice block in which all objects were presented individually without a distractor in the degraded version, and a practice block in which the objects were presented without a distractor in the intact version. These practice blocks preceded the corresponding experimental blocks.

Procedure

The participants were asked to name the objects as quickly and as accurately as possible, using bare nouns only. They should try to ignore the distractor words. At the beginning of the session, they received a booklet displaying the objects and the names they should use to refer to them. The booklets showed the objects in the intact or degraded version depending on the type of test blocks to follow. When the participants had familiarised themselves with the objects and their names, they were given a second picture booklet that did not include the object names and were asked to name all objects in the presence of the experimenter. Any naming errors were corrected. Then the practice block began. At the beginning of each practice trial, a fixation point was shown for 500 ms in the centre of the screen. After a blank interval of 100 ms, one of the objects was presented for 2000 ms. Participants named the objects as fast as possible. The inter-trial interval was 1500 ms.

The practice block was followed by two experimental blocks showing all pictures in the same version (degraded or intact). On experimental trials a fixation point appeared for 500 ms in the centre of the top left quadrant of the screen, where the left object would appear a little later. After a blank interval of 100 ms, an object triplet was presented for 4000 ms, accompanied by an auditory distractor word. After a blank interval of 900 ms, the next trial began. There were short pauses between the test blocks. After the second experimental block, the participants saw a booklet showing the pictures in the version to be tested in the third and fourth experimental block. When they had familiarised themselves with the materials, they named them in the presence of the experimenter. Then the second practice block and the third and fourth experimental block followed. The experimental session took approximately 90 min.

Results

The rate of naming errors was low (3.89%) and independent of the experimental conditions (all Fs < 1). 53 trials were excluded because participants failed to fixate upon one of the three objects and 241 trials because they did not inspect the objects in the expected order. The overall rate of missing observations was 12.24%.

The mean gaze duration for the left object was shorter by 54 ms in the intact than in the degraded condition (see Table 3), but this difference was only significant in the analysis by items ($F_1(1,15) = 2.75$, $F_2(1,54) = 17.84$, $p < .001$, $minF'(1,19) = 2.38$, CI = 67 ms). There was a significant effect of distractor type ($F_1(1,15) = 29.06$, $F_2(1,54) = 27.55$, $minF'(1,49) = 14.40$, all $p < .01$, CI = 23 ms), with the mean gaze duration for the left objects being shorter, by 60 ms, when related than when unrelated distractors were presented. The effects of stimulus quality and distractor type did not interact with each other (all Fs < 1).

A similar pattern of results was obtained for the speech onset latencies: The mean speech onset latency was shorter by 77 ms in the intact than in the degraded condition ($F_1(1,15) = 15.05$, $F_2(1,54) = 59.34$, $minF'(1,23) = 12.01$, all $p < .01$, CI = 42 ms), and it was shorter by 61 ms when phonologically related than when unrelated distractors were presented ($F_1(1,15) = 23.71$, $F_2(1,54) = 27.78$, $minF'(1,43) = 12.79$, all $p < .01$, CI = 27 ms). The interaction between the two effects was not significant.

Most importantly, there was no evidence for a trade-off between the left-object gaze durations and the gaze-speech lags. The difference in gaze-speech lag in the related and unrelated distractor condition was 3 ms (all Fs < 1, CI = 24 ms). The gaze-speech lag was *longer* by 28 ms in the degraded than in the intact version, but this difference was only significant in the analysis by items ($F_1(1,15) = 1.25$, $F_2(1,55) = 27.00$, $p < .01$, $minF'(1,17) = 1.19$, CI = 49 ms).

Discussion

Visual degradation and phonological priming affected the gaze durations for the left objects. These results replicate findings by Meyer et al. (1998), who also presented intact and visually degraded line drawings, and by Meyer and van der Meulen (2000), who also investigated the effects of phonological priming. Visual degradation affects primarily the visual–conceptual processing of the objects, whereas phonological priming affects the ease of access to their names. Therefore the results of Experiment 3 support the view that speakers naming several objects look at each of the objects until they have identified it and have planned its name to the level of phonological form. The effects of contour deletion and phonological priming were additive, as predicted by serial stage models of lexical access (e.g., Levelt et al., 1999).

The main goal of the experiment was to examine whether the effects of the experimental variables on the left-object gaze durations would be accompanied by reversed effects on the gaze-speech lag. This was clearly not the case. There was no effect of distractor type on the gaze-speech lag. There was a 24-ms effect of visual degradation, but it was only significant in the by-items analysis and it was in the same direction as the effect obtained for the gaze durations: In the degraded condition the speakers spent slightly more time before speech onset looking at the right object than in the intact condition. Since all objects of a triplet were presented in the intact or in the degraded version, the effects of the degradation of the first and second object cannot be separated. Perhaps participants spent more time looking at the degraded than the intact right object before speech onset because they required more time to process the right object to a criterion they aimed to reach before initiating the utterance.

The effect of phonological priming on the gaze durations most likely arose during the retrieval of the phonological form of the name of the left object. The same is true for the effect of word length observed in Experiment 1. The two effects were similar in size—the word length effect was 76 ms and the phonological priming effect was 60 ms. However, only the word length effect was accompanied by a reversed effect on the gaze-speech lag. This pattern argues against the trade-off hypothesis and links

Table 3
Mean error rates (%), speech onset latencies (ms), left-object gaze durations (ms), and gaze-speech lags (ms) in Experiment 3

	Intact objects		Degraded objects	
	Related distractors	Unrelated distractors	Related distractors	Unrelated distractors
Error rate	3.64	3.07	4.21	4.66
Speech onset latency	892	962	978	1029
Left-object gaze duration	670	733	728	784
Gaze-speech lag	221	227	252	252

the reversed word length effect on the gaze-speech lag specifically to the length of the object names rather than the time required to plan them.

Experiment 4

A plausible account of the occurrence of the reversed length effect on the gaze-speech lag is that speakers aim to produce the two object names as part of one fluent utterance and therefore take the estimated spoken duration of the first object name into account when deciding how extensively they should prepare the second object name before speech onset (Griffin, 2003). If this hypothesis is correct, there should be a reversed length effect whenever speakers aim to produce two object names as part of one fluent utterance. One would not expect to see such an effect when speakers only name one object and then react to a visual stimulus using a non-verbal response. This prediction was tested in Experiment 4.

The experiment included three sub-experiments, the first two of which were identical to the word recognition and single-object naming sub-experiments of Experiment 1. They were carried out to familiarise the participants with the materials in the same way as in Experiment 1. In the third sub-experiment, the participants named the same left objects with monosyllabic or trisyllabic names as in Experiment 1. However, instead of naming another object, they categorised a symbol, shown to the right of the object, as either a plus-sign or the letter *x* by pressing a button on a push-button panel. Since the symbol was quite small, the participants had to fixate upon it to identify and categorise it.

Roelofs (2007) used a similar task. In his Experiment 1, participants first named a picture, which was accompanied by a written distractor, and then categorised an arrow shown to the right of the picture as pointing leftwards or rightwards. The distractors were semantically related, unrelated, or identical to the target names or, in a control condition, consisted of a string of Xs. As expected, the distractor type affected the target naming latencies, which were shorter in the control and identity condition than in the unrelated condition, and shorter in the unrelated than in the semantically related condition. Importantly, parallel effects were seen for the durations of the gazes to the target objects and for the symbol categorisation latencies.

Roelofs's results indicate that in the naming + categorisation task, as in the dual-object naming task, the gaze durations for the objects depend on the time required for object identification and name retrieval. Therefore the word length effect on the gaze durations seen in Experiment 1 should be replicated in Experiment 4. But since the participants only produced one object name per trial and therefore did not have to consider the fluency of their utterances, the gaze-speech lag should now be independent of the length of the object names.

Method

Participants

The experiment was carried out with 18 participants.

Materials and design

Materials and design were identical to those of Experiment 1, except that the dual-object naming task was replaced by a naming + categorisation task. For this task, the same objects with monosyllabic and trisyllabic names were shown on the left side of the screen as in the dual-object naming task of Experiment 1. In the centre of the right half of the screen, either a plus-sign or the letter *x* was presented (in font Arial, 24 point). Half of the right objects of Experiment 1 were replaced by the plus-sign, and half by the letter *x*.

Procedure

Before the beginning of the naming + categorisation sub-experiment, the participants learned that on each trial they would see a picture and a symbol, and that they should first name the object and then categorise the symbol. They should press the left button on the push-button panel when they saw the letter *x* and the right button when they saw a plus-sign. Otherwise, the procedure was the same as in Experiment 1.

Table 4
Mean error rates (%), response latencies (ms), left-object gaze durations (ms), and gaze-speech lags (ms) in Experiment 4

	Type of object name		
	Monosyllabic	Trisyllabic	Disyllabic
Object recognition			
Error rate			
Word-picture match	3.47	2.78	2.43
Word-picture mismatch	5.90	5.56	2.47
Recognition latency			
Word-picture match	549	546	567
Word-picture mismatch	603	584	568
Single-object naming			
Error rate	8.67	9.72	6.42
Object naming latency	624	659	665
Naming + categorisation			
Error rate	10.67	11.46	
Speech onset latency	719	733	
Left-object gaze duration	461	51	
Gaze-speech lag	258	223	
Decision latency	1219	1283	

Results

Object recognition

Table 4 shows the error rates and response latencies for the object recognition task. The error rate was significantly higher for mismatching than for matching word-object pairs (5.73 vs. 3.12%, $F_1(1,17) = 6.58$, $p < .05$, $F_2(1,30) = 5.37$, $p < .05$, $minF'(1,46) = 2.96$, CI = 0.7%). The response latencies were significantly longer, by 47 ms, on matching than on mismatching trials ($F_1(1,17) = 11.80$, $p < .01$, $F_2(1,30) = 17.76$, $p < .01$, $minF'(1,38) = 7.09$, $p < .01$, CI = 28 ms). There was no significant effect of word length on error rates or response latencies.

Single-object naming

The error rates for monosyllabic and trisyllabic items (8.67 vs. 9.72%) did not differ significantly. The mean naming latency was significantly shorter, by 35 ms, for monosyllabic than for trisyllabic items ($F_1(1,17) = 11.27$, $p < .01$, $F_2(1,30) = 5.93$, $p < .05$, $\mathrm{minF}'(1,47) = 2.54$, CI = 15 ms, see Table 4).

Naming + categorisation

The rates of naming errors (missing, incorrect, or repaired object names, and naming responses with latencies exceeding two seconds) for monosyllabic and trisyllabic object names were 10.67 and 11.46%, respectively. Error trials were excluded from further analyses. An additional 5.4% of the trials were excluded because participants did not classify the symbol correctly.

As expected, the participants usually looked first at the object and then at the symbol. Four trials were excluded because participants did not look at the symbol at all, and 83 trials because they looked first at the symbol and then at the object. In total, 14.93% of the data from the monosyllabic condition and 14.23% of the data from the trisyllabic condition were excluded from further analysis.

The results for the remaining trials are shown in Table 4. The mean gaze duration was significantly longer, by 49 ms, for left objects with trisyllabic names than for left objects with monosyllabic names ($F_1(1,17) = 8.05$, $F_2(1,30) = 14.65$, $minF'(1,35) = 5.20$, all $p < .01$, CI = 26 ms). The gaze-speech lag was significantly shorter, by 35 ms, when the left object had a trisyllabic name than when it had a monosyllabic name ($F_1(1,17) = 4.24$, $p < .056$, $F_2(1,30) = 25.14$, $p < .01$, $minF'(1,23) = 3.63$, CI = 26 ms). The mean speech onset latency was shorter by 14 ms for monosyllabic than for trisyllabic items, but this difference was not significant ($F_1(1,17) = 1.76$, $F_2(1,30) < 1$, $minF'(1,33) < 1$, CI = 16 ms). Finally, the mean push-button latency was significantly shorter, by 64 ms, when the name of the object shown on the same trial was monosyllabic than when it was trisyllabic ($F_1(1,17) = 7.77$, $F_2(1,30) = 16.85$, $minF'(1,32) = 5.32$, all $p < .01$, CI = 34).

Discussion

The participants of Experiment 4 were faster to initiate their responses than those of Experiment 1 (means: 726 vs. 793 ms, $F_1(1,36) = 4.79$, $p < .05$, $F_2(1,30) = 128.00$, $p < .001$, $minF'(1,39) = 4.62$, $p < .05$ CI = 44 ms), and they spent less time looking at the left object (means: 483 vs. 570 ms, $F_1(1,36) = 7.46$, $p < .05$, $F_2(1,30) = 157.12$, $p < .001$, $minF'(1,39) = 7.12$, $p < .05$, CI = 46 ms). The reasons for these differences are not clear, but they might be related to the fact that the syntactic and prosodic planning of the utterances was less complex for the shorter utterances of Experiment 4. Alternatively, it is possible that the speakers of Experiment 1 already began to process the right object while they were fixating upon the left object and that this interfered with the processing of the left object (see Morgan & Meyer, 2005). Extrafoveal processing of the symbols replacing the right objects in Experiment 4 was less likely because the symbols were too small to be identified prior to fixation.

As expected, the gaze duration for the left object was significantly shorter when the left object had a monosyllabic name than when it had a trisyllabic name. This word length effect was numerically weaker than in Experiment 1 (49 vs. 76 ms), but the interaction of experiment and word length was not significant ($F_1(1,36) < 1$, $F_2(1,30) = 2.58$, $minF'(1,36) < 1$, $p < .05$, CI = 18 ms). The word length effect on the left object gaze durations indicates that the participants initiated the saccade to the second stimulus after they had retrieved the phonological form of the name of the left object (see also Roelofs, 2007). There was a substantial effect of word length on the push-button latencies. This effect arose because on most trials the participants categorised the symbol after they had articulated the object name. It therefore reflects the difference in the spoken durations of the monosyllabic and trisyllabic object names.

Most importantly, the gaze-speech lag was significantly shorter for trisyllabic than for monosyllabic object names. This effect was smaller than in Experiment 1 (35 vs. 74 ms), and the interaction of experiment and length approached significance ($F_1(1,36) = 3.40$, $p < .10$, $F_2(1,30) = 8.85$, $p<.01$, $minF'(1,36) = 2.46$, CI = 16 ms). However, the word length effect on the gaze durations for the left objects was also smaller in Experiment 4 than in Experiment 1. In both experiments, the same pattern was seen: The length effect on the left object gaze durations was accompanied by a reversed length effect of approximately the same size on the gaze-speech lags. Given that the participants of Experiment 4 only named one object, the reversed length effect on the gaze-speech lag cannot be directly related to their aim to produce the utterances fluently. We will return to this finding in the General discussion.

General discussion

Griffin (2003) and Meyer et al. (2003) asked speakers to name object pairs and found that they spent more time before speech onset looking at the second object when the first one had a monosyllabic name than when it had a polysyllabic name. The aim of the experiments reported in the present paper was to replicate this effect and to determine whether it was related to the spoken duration of the object names or to the time required to plan the names.

In Experiment 1, the participants named pairs of objects, the first of which (shown on the left side of the screen) had either a monosyllabic or a trisyllabic name. There was a significant word length effect on the mean gaze duration for the left object and a significant reversed length effect on the gaze-speech lag, which is the time between the end of the inspection of the left object and the onset of speech. It corresponds to the duration of the saccade from the left to the right object plus the pre-speech onset inspection time of the right object. Thus, as in the studies by Griffin and by Meyer et al., the participants spent more time before speech onset looking at the right object when the left one had a short name than when it had a long name.

In Experiments 2 and 3, the ease of identifying the left objects and of retrieving their names was varied, while the length of the names was held constant. This affected the mean duration of the gazes to the left objects but did not yield a reversed length effect on the gaze-speech lag. Therefore, the reversed length effect on the gaze-speech lag seen in Experiment 1 is likely to be related to the difference in the length of the names of the left objects rather than to differences in the time required to identify the objects or to plan their names. In sum, we replicated the reversed length effect and demonstrated that it was an effect of word length rather than word planning time.

In the remainder of this section, we will discuss two hypotheses about the origin of the reversed length effect, called the *length monitoring hypothesis* and the *incremental articulatory planning hypothesis*, respectively. The length monitoring hypothesis is the account put forward by Griffin (2003, p. 608). According to this hypothesis, speakers predict the spoken duration of the first object name and use this information in deciding when to initiate the utterance. When the first object name is short, they initiate the utterance slightly later and spend a little more time preparing the second object name before speech onset than when the first object name is long. Presumably, speakers do this to avoid utterance-internal speech disfluencies.

Under the length monitoring hypothesis, the gaze-speech lag is viewed as the time speakers spend before speech onset preparing the name of the second object. This is, of course, entirely reasonable, given that the second object is inspected during the gaze-speech lag. However, while the speakers are looking at the right object, they are still engaged in the final processing steps leading to the articulation of the name of the left object. As discussed in the Introduction, there is strong evidence that speakers naming several objects look at each object until they have retrieved the phonological form of its name and then initiate the shift of gaze to a new object (e.g., Meyer & Lethaus, 2004; Roelofs, 2007). Therefore, the gaze-speech lag depends on the time required for any phonetic and articulatory programming processes that occur between the completion of phonological encoding of the first object name and the onset of speech. According to the incremental articulatory planning hypothesis, the reversed length effect arises because these processes take more time for monosyllabic than polysyllabic words.

This hypothesis is illustrated in Fig. 3. As described in the Introduction, current models of word production assume that the phonological forms of words are generated sequentially, proceeding from the beginning to the

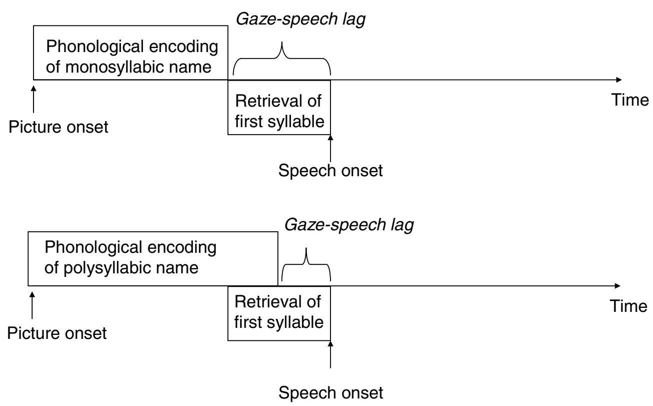

Fig. 3. Co-ordination of phonological and articulatory planning for monosyllabic and polysyllabic object names.

end of each word, and that articulatory encoding begins as soon as the first part of the phonological code has been generated (e.g., Roelofs, 1997a, 1997b, 2002; see also Levelt et al., 1999; Santiago et al., 2000). In Roelofs's (1997a, 1997b) model of word form encoding, phonological segments are activated in parallel but are syllabified (combined into syllables) sequentially, in a left-to-right fashion (see also Levelt et al., 1999). As soon as the first phonological syllable has been created, the corresponding articulatory code is accessed and stored in an output buffer. Therefore, the relative timing of phonological and articulatory encoding is different for monosyllabic and polysyllabic words. In monosyllabic words, articulatory encoding can only be initiated after phonological encoding has been completed. By contrast, in polysyllabic words, the retrieval of the articulatory code of the first syllable occurs in parallel with the generation of the phonological code for the following syllables.

We assume that speakers can initiate an utterance as soon as they have retrieved the articulatory code for a single syllable (Meyer et al., 2003; Roelofs, 2002; Schriefers & Teruel, 1999). If they do so, i.e., if they begin to speak as soon as one syllable has been fully planned, a reversed length effect on the gaze-speech lag can arise. This is because the articulatory code for a monosyllabic word is only retrieved *after* the end of phonological encoding, i.e., after the shift of gaze to the right object has been initiated. By contrast, the retrieval of the articulatory code for a polysyllabic word begins *before* the end of phonological encoding, and before the shift of gaze to the right object is initiated. Thus, if speakers begin to speak as soon as they have retrieved the phonological code for the entire word and the articulatory code for the first syllable of the utterance, the interval between the shift of gaze and the onset of speech—the gaze-speech lag—should be shorter when the first object name is polysyllabic than when it is monosyllabic. This is the result we observed.

This account of the reversed length effect presupposes that the retrieval of the articulatory code for the first syllable of a polysyllabic word takes about as much time as the retrieval of the articulatory code for a monosyllabic word. For a given set of stimuli this may or may not be true. In our Experiments 1 and 4 the monosyllabic object names included significantly more segments than the first syllables of the trisyllabic object names (means: 2.8 ($SD = 0.66$) vs. 2.1 ($SD = 0.57$) segments, $F(1,30) = 11.87$, $p < .01$, CI = 0.23 segments), and they were significantly lower in syllable frequency (CELEX summed frequency per million words: 38 ($SD = 108$) vs. 266 ($SD = 241$), $F(1,30) = 11.96$, $p < .01$). Syllable frequency effects on speech onset latencies have been observed in several word production studies (Cholin et al., 2006; Levelt & Wheeldon, 1994; Perea & Carreira, 1998). Cholin et al. (2004) recently showed that these effects most likely arise during articulatory encoding. However, the reported syllable frequency effects were invariably very small (less than 15 ms), even when the frequency difference between the high and low frequency sets was much larger than in the present materials. Thus, differences in length and frequency between the monosyllables and the first syllables of the trisyllabic words perhaps contributed to the reversed length effect in our study, but they are unlikely to be solely responsible for the effect. It is important to note that a contribution of syllable frequency differences to the reversed length effect would not undermine the essence of the incremental articulatory planning hypothesis, which links the reversed length effect to the time required for the articulatory programming processes occurring between the end of the phonological encoding of the first object name and the onset of speech.

In sum, both hypotheses account for the reversed length effect in dual-object naming experiments. One hypothesis does so by referring to differences in the estimated duration of the first object name and corresponding adjustments of the time taken to plan the second object name, the other by referring to differences in the time required after the shift of gaze for the articulatory encoding of the first syllable of the utterance. As reported above, we found a reversed length effect in Experiment 4 using the naming + categorisation task. Since the participants of this experiment produced only one word, utterance fluency could not have been a concern. The length monitoring hypothesis would have been strengthened if the reversed length effect had been absent in this experiment. Under this hypothesis, one might account for the presence of the effect by postulating that the participants for some reason aimed to press the response button immediately after the offset of the first object name. By contrast, the incremental articulatory planning hypothesis correctly predicts the occurrence of the reversed length effect in the naming + categorisation experiment. This is because this hypothesis ascribes the effect to differences in the temporal coordination of phonological and articulatory planning for long and short object names, which should be unaffected by the nature of the response to the second stimulus.

According to the incremental articulatory planning hypothesis, a reversed length effect on the gaze-speech lag only arises if speakers begin to speak as soon as they have retrieved the articulatory programme for the first syllable of the utterance. The hypothesis does not entail that speakers must always adopt this strategy. If they do so systematically, the word length effect on the gaze duration for the first object and the reversed length effect on the gaze-speech lag should be equal in size, and there should be no word length effect on the speech onset latencies. This was the result observed in Experiment 1 of the present study. However, if speakers retrieve the articulatory code of several syllables of polysyllabic words before speech onset, the reversed length effect on the gaze-speech lag should be attenuated or absent. In this case, articulatory planning should still begin earlier (relative to the completion of phonological encoding and the shift of gaze) for polysyllabic than for monosyllabic words. However, the speakers will need more time to complete the articulatory planning processes for the polysyllabic words, which will counteract the effect of the earlier start of these processes on the gaze-speech lag. A pattern compatible with this scenario was seen in the pure blocks of the dual-object naming experiment carried out by Meyer et al. (2003), where the word length effect on the left-object gaze duration was stronger (64 ms) than the reversed length effect on the gaze-speech lag (34 ms). This suggests that the participants of Meyer et al.'s study were less likely than the participants of the present study to begin to speak as soon as they had retrieved the articulatory code for one syllable, perhaps because they were less thoroughly familiarised with the materials or because they produced phrases such as "kat en stoel" ("cat and chair") instead of bare nouns. Prosodic wellformedness may be more important to speakers when they produce phrases than bare nouns and this might render them less inclined to begin to speak as soon as a single syllable has been fully planned. The length monitoring hypothesis does not entail predictions about the relative size of the length effects on the left-object gaze duration and gaze-speech lag and can therefore easily accommodate these findings.

In addition to the pure blocks discussed so far, in which the names of all left objects had the same length, the study by Meyer et al., included mixed blocks featuring a mixture of left objects with monosyllabic and disyllabic names. In these blocks, there was no length effect on the left-object gaze durations, but there was a reversed length effect on the gaze-speech lag. Under the incremental articulatory planning hypothesis, this pattern can be seen as indicating that stimulus blocking affected how much time the speakers dedicated to the processes preceding the shift of gaze to the right object (i.e., to the visual–conceptual processing of the left object and the planning of its name to the level of phonological form), but that stimulus blocking did not affect the temporal co-ordination of these processes with articulatory encoding. In pure blocks, they dedicated more time to the processing of objects with disyllabic than with monosyllabic names because the phonological encoding processes were more complex for the former than the latter types of names. By contrast, in mixed blocks, they spent approximately the same amount of time on all objects regardless of the length of their names. Effects of stimulus blocking on reaction times have been explained by reference to the response deadlines participants choose (e.g., Kello & Plaut, 2000; Kello, Plaut, & MacWhinney, 2000; Lupker, Brown, & Colombo, 1997; Lupker, Kinoshita, Coltheart, & Taylor, 2003). The results obtained by Meyer et al. suggest that instead of response deadlines the participants used processing deadlines which governed how much time they allowed for the processing of the left object and when they initiated the shift of gaze to the right object (see also Roelofs, 2007). The timing of the speech onset relative to the shift of gaze depended on the time required to complete the articulatory encoding of the first syllable (or perhaps sometimes the entire word) and was independent of these deadlines.

Under the length monitoring hypothesis, the results would indicate that stimulus blocking affected how much time speakers allowed for the planning of the first object name, but that they consistently spent slightly more time preparing the second object name when the first name was short than when it was long. It is not clear why speakers would adopt such a strategy; perhaps they considered the fluency of their utterances to be very important and therefore consistently adjusted the gaze-speech lag depending on the length of the first object name.

In our view, the incremental articulatory planning hypothesis accounts at least as well for the results of the present study and those reported by Meyer et al. as the length monitoring hypothesis. However, Griffin (2003) reported effects of the structure of the utterances the participants produced on the co-ordination of eye gaze and speech that are readily explained under the length monitoring hypothesis but not under the incremental articulatory planning hypothesis: First, speakers spent less time looking at the right object before speech onset and initiated the utterances sooner when they produced phrases, such as "cat next to windmill," than when they produced bare nouns, such as "cat, windmill." Second, the reversed length effect was reduced from a significant 47 ms in the bare noun condition to a non-significant 21 ms in the phrase condition. The length monitoring hypothesis offers an elegant account of these findings: In the phrase condition, the speakers had more time to plan the second object name during the articulation of the preceding words than in the bare noun condition and could therefore afford to initiate the utterance sooner. In addition, they often did not take the length of the name of the left object into account in deciding when to begin to speak because there was always sufficient time to plan the second name after speech onset. This led to the reduction of the reversed length effect.

The incremental articulatory planning hypothesis concerns the temporal relationship between the phonological encoding and articulatory planning of long and short words and can, by its very nature, not explain the effects of utterance format on the co-ordination of eye gaze and speech. However, the hypothesis does not entail that such effects should not exist. As discussed in the Introduction, many studies have shown that speakers naming several objects often plan more than one object name before speech onset (e.g., Smith & Wheeldon, 1999). Therefore, the criteria for speech onset invoked by the incremental articulatory planning hypothesis—the completion of the phonological form of the first object name and the retrieval of the articulatory programme for the first syllable—cannot be the only criteria governing speech onset in all settings. The effects of the utterance format on the speech onset latency and on the gaze-speech lag can be explained in the same way as under the length monitoring hypothesis: When speakers produce utterances using the format "A next to B," they may decide to plan the second object name less thoroughly before speech onset than when they produce the two object names in immediate succession. Thus, as proposed by Griffin, they may aim to produce fluent utterances and adjust their advance planning of the second object name depending on the planning time available after speech onset. Note that in Griffin's experiments, the speakers used the same utterance format on all trials. Therefore, they did not have to predict the length of their utterances on-line on the basis of the developing utterance plan but could draw upon their experience from the preceding trials. An account of the attenuation of the reversed length effect in the phrase relative to the bare noun condition, which is compatible with the incremental articulatory planning hypothesis, was proposed in the above discussion of the results obtained by Meyer et al.: Speakers sometimes choose to plan more than one articulatory syllable of polysyllabic left-object names before speech onset, and they may be more likely to do so when they produce phrases than when they produce bare nouns. This would lead to a smaller reversed length effect in the phrase than in the bare noun condition.

Finally, Griffin found that the speakers of her Experiment 2, who articulated the utterances more slowly than the speakers of Experiment 1, spent more time before speech onset looking at the left object and less time looking at the right object. This pattern can be elegantly accounted for under the length monitoring hypothesis: The slower speakers had more time to prepare the name of the second object during the articulation of the name of the first object and could therefore afford to prepare the second object name less extensively before speech onset. However, it is also possible that the two groups of speakers differed in the temporal co-ordination between phonological and articulatory planning. Since the slower speakers spent more time generating the phonological code of the name of the left object than the faster speakers, more of the articulatory planning could be completed in parallel with the phonological encoding processes, and less remained to be done after the end of these processes, yielding shorter gaze-speech lags. In other words, the incremental articulatory planning hypothesis also offers an account of these differences between the two groups of speakers. More research is required to determine how the co-ordination of speech planning and articulation changes when speakers use different speech rates.

In sum, there is converging evidence from three studies demonstrating that speakers naming several objects spent more time looking at the second object when the first one had a short name than when it had a long name. Two hypothesis have been put forward for this reversed length effect on the gaze-speech lag. The length monitoring hypothesis links the effect to the speakers' estimate of the length of the first object name and the planning for the second object name carried out before speech onset. By contrast, the incremental articulatory planning hypothesis links it to the temporal co-ordination of phonological and articulatory planning of monosyllabic and polysyllabic words. The reason why both hypotheses are viable is that the gaze-speech lag is the time speakers spent before speech onset looking at the second object and presumably planning its name, but it is also the time they require after the shift of gaze to complete the articulatory programming for the first utterance fragment. The length monitoring hypothesis explains the effects of word length, utterance structure, and speech rate by reference to a single principle. The incremental articulatory planning hypothesis is more limited in scope, but it accounts for the reversed length effect without crediting speakers with the ability and willingness to estimate the durations of upcoming words. In our view, this in an advantage, given that it is not clear why speakers would be so concerned with the perfect timing of their utterances. After all, listeners would not be able to hear pauses of the duration of the reversed length effect. However, at present this is a matter of opinion and further research is required to explore the merits of both proposals.

Appendix A

Materials of Experiments 1 and 4

Objects with monosyllabic names: ant, axe, bat, bow, brush, coin, corn, egg, kite, owl, pie, pig, pin, straw, tap, web

Objects with trisyllabic names: aubergine, envelope, banana, bicycle, buffalo, colander, crocodile, elephant, kangaroo, umbrella, parachute, pyramid, potato, skeleton, tomato, violin

Objects with disyllabic names: bucket, candle, coffin, carrot, dragon, ladder, lemon, monkey, orange, pencil, rabbit, robot, spider, toilet, tortoise, wardrobe

Materials of Experiment 2

Objects with early acquired names: apple, button, basket, castle, doctor, kettle, lorry, lion, mermaid, necklace, paintbrush, penguin, piano, scissors, sandwich, window

Objects with late acquired names: anchor, beetle, bullet, cactus, diamond, cannon, lobster, light bulb, medal, needle, pepper, peacock, pliers, syringe, cymbals, waistcoat

Materials of Experiment 3

Related target (left object)/distractor pairs: anchor/vicar, arm/farm, bag/rag, banana/piranha, bear/scare, bell/spell, bone/phone, book/look, broom/zoom, brush/crush, cactus/tortoise, camel/enamel, can/flan, castle/whistle, cheese/breeze, church/birch, cook/hook, cross/boss, curtain/baton, dress/stress, factory/diary, finger/singer, flag/lag, flute/suit, fork/pork, guitar/cigar, gun/fun, hammer/grammar, hand/band, harp/sharp, hat/bat, kettle/nettle, lemon/sermon, mask/task, moon/dune, onion/champion, organ/wagon, pig/gig, plane/strain, rocket/pocket, saw/law, scooter/hooter, seal/deal, ship/clip, snail/mail, sock/wok, square/pear, swing/king, table/cable, tent/rent, top/cop, tree/knee, trousers/scissors, vase/mars, violin/zeppelin

Right objects: ball, cake, wheel, key, iron, giraffe, tire, toaster, nun, palm, watch, lorry, ladder, toilet, pencil, knife, snake, lamp, plant, rake, plug, leaf, cigarette, dog, robot, ring, shuttle, kite, rabbit, well, toe, statue, arch, swan, saddle, sword, pot, ruler, trumpet, desk, bow, door, eye, umbrella, camera, penguin, monkey, pan, sheep, spider, belt, star, bed, stool, boot

Bottom objects: bowl, bread, bus, cap, car, cat, chair, clock, clown, coat, comb, crown, cup, doll, drum, ear, elephant, feather, fence, foot, fox, frog, glass, globe, glove, hair, heart, horse, house, jug, leg, lock, nose, owl, pen, piano, pie, pin, pipe, pump, purse, rope, shirt, shoe, skirt, sled, slide, spoon, stamp, sun, tap, tie, waiter, web, whale

References

Bachoud-Lévi, A.-C., Dupoux, E., Cohen, L., & Mehler, J. (1998). Where is the length effect: a cross-linguistic study of speech production. *Journal of Memory and Language, 39*, 331–346.

Barry, C., Hirsh, K. W., Johnston, R. A., & Williams, C. L. (2001). Age of acquisition, word frequency, and the locus of repetition priming in picture naming. *Journal of Memory and Language, 44*, 350–375.

Belke, E., Brysbaert, M., Meyer, A. S., & Ghyselinck, M. (2005). Age of acquisition effects in picture naming: evidence for a lexical-semantic competition hypothesis. *Cognition, 96*, B45–B54.

Bock, K., & Levelt, W. (1994). Grammatical encoding. In M. A. Gernsbacher (Ed.), *Handbook of psycholingustics* (pp. 945–984). New York: Academic Press.

Brown, R., & McNeill, D. (1966). The 'tip-of-the-tongue' phenomenon. *Journal of Verbal Learning and Verbal Behavior, 5*, 325–337.

Brown, G. D. A., & Watson, F. L. (1987). First in, first out. Word learning age and spoken word frequency as predictors of word familiarity and word naming latency. *Memory & Cognition, 15*, 208–216.

Brysbaert, M., & Ghyselinck, M. (2006). The effect of age-of-acquisition: partly frequency-related, partly frequency-independent. *Visual Cognition, 13*, 992–1011.

Brysbaert, M., Van Wijnendaele, I., & De Deyne, S. (2000). Age-of-acquisition of words is a significant variable in semantic tasks. *Acta Psychologica, 104*, 215–226.

Caramazza, A., Papagno, C., & Ruml, W. (2000). The selective impairment of phonological processing in speech production. *Brain and Language, 75*, 428–450.

The CELEX *Lexical Database* [CD-ROM]. Philadelphia, PA: Linguistic Data Consortium, University of Pennsylvania. (1993).

Cholin, J., Levelt, W. J. M., & Schiller, N. O. (2006). Effects of syllable frequency in speech production. *Cognition, 99*, 205–235.

Cholin, J., Schiller, N. O., & Levelt, W. J. M. (2004). The preparation of syllables in speech production. *Journal of Memory and Language, 50*, 47–61.

Clark, H. (1973). The language-as-fixed-effect fallacy: a critique of language statistics in psychological research. *Journal of Verbal Learning and Verbal Behavior, 12*, 335–359.

COBUILD [Corpus of English Sentences, CD-ROM]. The University of Birmingham. (2000).

Costa, A., & Caramazza, A. (2002). The production of noun phrases in English and Spanish: implications for the scope of phonological encoding in speech production. *Journal of Memory and Language, 46*, 178–198.

Crompton, A. (1982). Syllables and segments in speech production. In A. Cutler (Ed.), *Slips of the tongue and language production* (pp. 109–162). Berlin, New York, Amsterdam: Mouton.

Damian, M. F., & Martin, R. C. (1999). Semantic and phonological codes interact in single word production. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 25*, 345–361.

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

Dell, G. S., Burger, L. K., & Svec, W. R. (1997). Language production and serial order: a functional analysis and a model. *Psychological Review, 104*, 123–147.

Dell, G. S., Schwartz, M. F., Martin, N., Saffran, E. M., & Gagnon, D. A. (1997). Lexical access in aphasic and nonaphasic speakers. *Psychological Review, 104*, 801–838.

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

Eriksen, C. W., Pollack, M. D., & Montague, W. E. (1970). Implicit speech: Mechanisms in perceptual encoding? *Journal of Experimental Psychology, 84*, 502–507.

Ferreira, F., & Swets, B. (2002). How incremental is language production? Evidence from the production of utterances requiring the computation of arithmetic sums. *Journal of Memory and Language, 46*, 57–84.

Foygel, D., & Dell, G. S. (2000). Models of impaired lexical access in speech production. *Journal of Memory and Language, 43*, 182–216.

Fromkin, V. A. (1973). *Speech errors as linguistic evidence*. Mouton: The Hague.

Garrett, M. F. (1980). Levels of processing in sentence production. In B. Butterworth (Ed.). *Language production. Speech and talk* (Vol. 1, pp. 177–220). New York: Academic Press.

Griffin, Z. M. (2001). Gaze durations during speech reflect word selection and phonological encoding. *Cognition, 82*, B1–B14.

Griffin, Z. M. (2003). A reversed length effect in coordinating the preparation and articulation of words in speaking. *Psychonomic Bulletin & Review, 10*, 603–609.

Griffin, Z. M. (2004). Why look? Reasons for eye movements related to language production. In J. M. Henderson & F.

Ferreira (Eds.), *The interface of language, vision, and action: What we can learn from free-viewing eye tracking* (pp. 213–247). New York: Psychology Press.
Griffin, Z. M., & Bock, K. (2000). What the eyes say about speaking. *Psychological Science, 11*, 274–279.
Griffin, Z. M., & Oppenheimer, D. M. (2006). Speakers gaze at objects while preparing intentionally inaccurate names for them. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 32*, 943–948.
Hartley, T. A., & Houghton, G. (1996). A linguistically constrained model of short-term memory for nonwords. *Journal of Memory and Language, 35*, 1–31.
Hartsuiker, R. J., & Kolk, H. H. J. (2001). Error monitoring in speech production: a computational test of the perceptual loop theory. *Cognitive Psychology, 42*, 113–157.
Indefrey, P., & Levelt, W. J. M. (2004). The spatial and temporal signatures of word production components. *Cognition, 92*, 101–144.
Irwin, D. E. (2004). Fixation location and fixation duration as indices of cognitive processes. In J. M. Henderson & F. Ferreira (Eds.), *The interface of language, vision, and action: What we can learn from free-viewing eye tracking?* (pp. 105–133). New York: Psychology Press.
Izura, C., & Ellis, A. W. (2004). Age of acquisition effects in translation production tasks. *Journal of Memory and Language, 50*, 165–181.
Jescheniak, J. D., & Levelt, W. J. M. (1994). Word frequency effects in speech production: retrieval of syntactic information and of phonological form. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 20*, 824–843.
Jescheniak, J. D., Schriefers, H., Garrett, M. F., & Friederici, A. D. (2002). Exploring the activation of semantic and phonological codes during speech planning with event-related potentials. *Journal of Cognitive Neuroscience, 14*, 951–964.
Jescheniak, J. D., Schriefers, H., & Hantsch, A. (2003). Utterance format affects phonological priming in the picture-word task: implications for models of phonological encoding in speech production. *Journal of Experimental Psychology: Human Perception and Performance, 29*, 441–454.
Johnston, R. A., & Barry, C. (2006). Age of acquisition and lexical processing. *Visual Cognition, 13*, 789–845.
Juhasz, B. J. (2005). Age-of-acquisition effects in word and picture identification. *Psychological Bulletin, 131*, 684–712.
Kello, C. T., & Plaut, D. C. (2000). Strategic control in word reading: evidence from speeded responding in the tempo-naming task. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 26*, 719–750.
Kello, C. T., Plaut, D. C., & MacWhinney, B. (2000). The task dependence of staged versus cascaded processing: an empirical and computational study of Stroop interference in speech production. *Journal of Experimental Psychology: General, 129*, 340–360.
Klapp, S. T., Anderson, W. G., & Berrian, R. W. (1973). Implicit speech in reading, reconsidered. *Journal of Experimental Psychology, 100*, 368–374.
Levelt, W. J. M. (1989). *Speaking: From intention to articulation*. Cambridge, MIT: Press.
Levelt, W. J. M. (1992). Accessing words in speech production: stages, processes and representations. *Cognition, 42*, 1–22.
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.
Levelt, W. J. M., & Wheeldon, L. (1994). Do speakers have access to a mental syllabary? *Cognition, 50*, 239–269.
Lewis, M. B., Gerhand, S., & Ellis, H. D. (2001). Re-evaluating age-of-acquisition effects: Are they simply cumulative-frequency effects? *Cognition, 78*, 189–205.
Lupker, S. J., Brown, P., & Colombo, L. (1997). Strategic control in a naming task: changing routes or changing deadlines. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 23*, 570–590.
Lupker, S. J., Kinoshita, S., Coltheart, M., & Taylor, T. E. (2003). Mixing costs and mixing benefits in naming words, pictures, and sums. *Journal of Memory and Language, 49*, 556–575.
Masson, M. E., & Loftus, G. R. (2003). Using confidence intervals for graphically based data interpretation. *Canadian Journal of Experimental Psychology, 57*, 202–220.
Meyer, A. S. (1990). The time course of phonological encoding in language production: the encoding of successive syllables. *Journal of Memory and Language, 29*, 524–545.
Meyer, A. S. (1991). The time course of phonological encoding in language production: phonological encoding inside a syllable. *Journal of Memory and Language, 30*, 69–89.
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. (1997). Conceptual influences on grammatical planning units. *Language and Cognitive Processes, 12*, 859–864.
Meyer, A. S., & Lethaus, F. (2004). The use of eye tracking in studies of sentence generation. In J. M. Henderson & F. Ferreira (Eds.), *The interface of language, vision, and action: What we can learn from free-viewing eye tracking* (pp. 191–211). New York: Psychology Press.
Meyer, A. S., Roelofs, A., & Levelt, W. J. M. (2003). Word length effects in object naming: the role of a response criterion. *Journal of Memory and Language, 48*, 131–147.
Meyer, A. S., & Schriefers, H. (1991). Phonological facilitation in picture-word interference experiments: effects of stimulus onset asynchrony and types 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.
Miozzo, M., & Caramzza, A. (1999). The selection of determiners in noun phrase production. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 25*, 907–922.
Monaghan, J., & Ellis, A. W. (2002). Age of acquisition and the completeness of phonological representations. *Reading and Writing, 15*, 759–788.
Morgan, J. L., & Meyer, A. S. (2005). Processing of extrafoveal objects during multiple object naming. *Journal of Experi-*

mental Psychology: Learning, Memory, and Cognition, 31, 428–442.

Morrison, C. M., Chappell, T. D., & Ellis, A. W. (1997). Age of acquisition norms for a large set of object names and their relation to adult estimates and other variables. *The Quarterly Journal of Experimental Psychology, 50A*, 528–559.

Perea, M., & Carreira, M. (1998). Effects of syllable frequency and syllable neighborhood frequency in visual word recognition. *Journal of Experimental Psychology: Human Perception & Performance, 24*, 134–144.

Postma, A. (2000). Detection of errors during speech production: a review of speech monitoring models. *Cognition, 77*, 97–131.

Roelofs, A. (1996). Serial order in planning the production of successive morphemes of a word. *Journal of Memory and Language, 35*, 854–876.

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

Roelofs, A. (1997b). Syllabification in speech production: evaluation of WEAVER. *Language and Cognitive Processes, 12*, 659–696.

Roelofs, A. (1998). Rightward incrementality in encoding simple phrasal forms in speech production: verb-particle combinations. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 24*, 904–919.

Roelofs, A. (2002). Spoken language planning and the initiation of articulation. *Quarterly Journal of Experimental Psychology—Section A, 55*, 465–483.

Roelofs, A. (2004). Seriality of phonological encoding in naming objects and reading their names. *Memory & Cognition, 32*, 212–222.

Roelofs, A. (2007). Attention and gaze control in picture naming, word reading, and word categorization. *Journal of Memory and Language, 57*, 232–251.

Roelofs, A., & Lamers, M. (2007). Modeling the control of visual attention in Stroop-like tasks. In A. S. Meyer, L. R. Wheeldon, & A. Krott (Eds.). *Automaticity and control in language processing.* Hove: Psychology Press. (in press).

Santiago, J., MacKay, D. G., Palma, A., & Rho, C. (2000). Sequential activation processes in producing words and syllables: evidence from picture naming. *Language and Cognitive Processes, 15*, 1–44.

Schriefers, H. (1993). Syntactic processes in the production of noun phrases. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 19*, 841–849.

Schriefers, H., de Ruiter, J. P., & Steigerwald, M. (1999). Parallelism in the production of noun phrases: experiments and reaction time models. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 25*, 702–720.

Schriefers, H., Meyer, A. S., & Levelt, W. J. M. (1990). Exploring the time course of lexical access in language production: picture-word interference studies. *Journal of Memory and Language, 29*, 86–102.

Schriefers, H., & Teruel, E. (1999). Phonological facilitation in the production of two-word utterances. *European Journal of Cognitive Psychology, 11*, 17–50.

Sevald, C. A., & Dell, G. S. (1994). The sequential cuing effect in speech production. *Cognition, 53*, 91–127.

Slevc, L. R., & Ferreira, V. S. (2006). Halting in single word production: a test of the perceptual loop theory of speech monitoring. *Journal of Memory and Language, 54*, 515–540.

Smith, M., & Wheeldon, L. (1999). High level processing scope in spoken sentence production. *Cognition, 73*, 205–246.

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

Stemberger, J. P. (1985). An interactive activation model of language production. In A. W. Ellis (Ed.). *Progress in the psychology of language* (Vol.1, pp. 143–186). Hillsdale, NJ: Erlbaum.

Van Turennout, M., Hagoort, P., & Brown, C. M. (1998). Brain activity during speaking: from syntax to phonology in 40 milliseconds. *Science, 280*, 572–574.

Vousden, J. I., Brown, G. D. A., & Harley, T. A. (2000). Serial control of phonology in speech production: a hierarchical model. *Cognitive Psychology, 41*, 101–175.

Wheeldon, L., & Lahiri, A. (1997). Prosodic units in speech production. *Journal of Memory and Language, 37*, 356–381.

Wheeldon, L. R., & Lahiri, A. (2002). The minimal unit of phonological encoding: prosodic or lexical word. *Cognition, 85*, B31–B41.

Wilshire, C. E., & Saffran, E. M. (2005). Contrasting effects of phonological priming in aphasic word production. *Cognition, 95*, 31–71.

Winer, B. J., Brown, D. R., & Michels, K. M. (1991). *Statistical principles in experimental design* (3rd ed.). New York: McGraw-Hill.