Journal of Experimental Psychology: Learning, Memory, and Cognition

Effects of Phrase and Word Frequencies in Noun Phrase Production

Zeshu Shao, Jeroen van Paridon, Fenna Poletiek, and Antje S. Meyer

Effects of Phrase and Word Frequencies in Noun Phrase Production

Zeshu Shao and Jeroen van Paridon
Max Planck Institute for Psycholinguistics, Nijmegen,
the Netherlands

Fenna Poletiek
Leiden University

Antje S. Meyer
Max Planck Institute for Psycholinguistics, Nijmegen, the Netherlands, and Radboud University

There is mounting evidence that the ease of producing and understanding language depends not only on the frequencies of individual words but also on the frequencies of word combinations. However, in two picture description experiments, Janssen and Barber (2012) found that French and Spanish speakers' speech onset latencies for short phrases depended exclusively on the frequencies of the phrases but not on the frequencies of the individual words. They suggested that speakers retrieved phrase-sized units from the mental lexicon. In the present study, we examined whether the time required to plan complex noun phrases in Dutch would likewise depend only on phrase frequencies. Participants described line drawings in phrases such as *rode schoen* [red shoe] (Experiments 1 and 2) or *de rode schoen* [the red shoe] (Experiment 3). Replicating Janssen and Barber's findings, utterance onset latencies depended on the frequencies of the phrases but, deviating from their findings, also depended on the frequencies of the adjectives in adjective-noun phrases and the frequencies of the nouns in determiner-adjective-noun phrases. We conclude that individual word frequencies and phrase frequencies both affect the time needed to produce noun phrases and discuss how these findings may be captured in models of the mental lexicon and of phrase production.

Keywords: language production, mental lexicon, multiword frequency, word frequency

A standard assumption in psycholinguistics has long been the "words and rules" view (Dell, 1986; Levelt, 1989; Levelt, Roelofs, & Meyer, 1999; Pinker, 1998; see also Jackendoff, 2002, 2011): Speakers create utterances by combining words according to the grammatical rules of their language, and listeners understand utterances by looking up words in their lexicon and combining them using the same rules. Recently, this view has been challenged on a number of grounds. One important challenge arises from the fact that speakers and listeners are highly sensitive to frequency, not only to the frequencies of individual words but also to the frequencies of word sequences. This means that there must be some memory representation of the co-occurrence patterns of words. One conclusion one might draw from these findings is that the mental lexicon includes stored representations of multiword sequences. This assumption would jeopardize the neat distinction between the lexicon as a store reserved for arbitrary form-meaning mappings (i.e., individual words and some multiword idioms) and a grammar as a set of rules operating on these lexical units. It would also challenge the view that utterances (apart from idiomatic expressions) are necessarily planned by combining individual words. In the present article, we first review the evidence for frequency effects for words and word combinations. We then report two experiments assessing to what extent speech onset latencies for adjective-noun phrases such as *brown shoe* or *the brown shoe* depend on the frequencies of the individual words and the entire phrases.

Word Frequency Effects

There is a large body of research demonstrating that frequent words are processed faster than less frequent ones. This has been shown for various reading and listening tasks, for words presented in isolation and in context, for children and adults, and for speakers using their native or non-native languages (e.g., Balota & Chumbley, 1984, 1990; Broadbent, 1967; Cleland, Gaskell, Quinlan, & Tamminen, 2006; Connine, Mullennix, Shernoff, & Yelen, 1990; Dahan, Magnuson, & Tanenhaus, 2001; Diependaele, Lemhöfer, & Brysbaert, 2013; Kuperman & Van Dyke, 2013; Rayner, 1998; Troia, Roth, & Yeni-Komshian, 1996).

There is also a substantial literature on frequency effects in word production. Frequent words are more likely to be reduced, that is, they tend to be shorter in duration and to feature segment deletions and alterations in vowel quality compared with less frequent words (e.g., Aylett & Turk, 2006; Bell, Brenier, Gregory, Girand, &

Zeshu Shao and Jeroen van Paridon, Max Planck Institute for Psycholinguistics, Nijmegen, the Netherlands; Fenna Poletiek, Institute of Psychology, Leiden University; Antje S. Meyer, Max Planck Institute for Psycholinguistics, and Donders Institute for Brain, Cognition and Behaviour, Radboud University.

Correspondence concerning this article should be addressed to Zeshu Shao, Max Planck Institute for Psycholinguistics, P.O. Box 310, 6500 AH, Nijmegen, the Netherlands. E-mail: zeshu.shao@mpi.nl

Jurafsky, 2009; Griffin & Bock, 1998; Oldfield & Wingfield, 1965; Plaut & Kello, 1999; Pluymaekers, Ernestus, & Baayen, 2005; but see Damian, 2003; Munson, 2007). In addition, frequent words are easier to plan than less frequent ones: Their retrieval is less error-prone (Dell, 1990; Kittredge, Dell, Verkuilen, & Schwartz, 2008; Shi, 2015), less likely to result in tip-of-the-tongue states (Antón-Méndez, Schütze, Champion, & Gollan, 2012), and faster (e.g., Jescheniak & Levelt, 1994; Oldfield & Wingfield, 1965). Though word frequency is correlated with conceptual and lexical variables, such as familiarity with the concept, the ease of recognizing the object mentioned, and the age of acquiring the word and its length, frequency effects cannot be fully accounted for by such influences (e.g., Almeida, Knobel, Finkbeiner, & Caramazza, 2007; Belke, Brysbaert, Meyer, & Ghyselinck, 2005). In short, there is good evidence for a lexical basis of word frequency effects.

Many studies have aimed to pin down the origin of word frequency effects in speech planning more precisely. In a seminal study, Jescheniak and Levelt (1994) examined whether word frequency affected the speed of access to the grammatical representations of words (their *lemmas*) or to their form representations. This was done by comparing word frequency effects in picture naming, which requires full lexical access, to frequency effects arising in object recognition, delayed naming, and gender decision tasks, which do not require word forms access. Jescheniak and Levelt concluded that word forms were the primary locus of the word frequency effect in production. However, more recent evidence suggests that frequency effects, though perhaps most prominent at the word form level, arise at multiple levels of representation (e.g., Caramazza, Costa, Miozzo, & Bi, 2001; Cuetos, Bonin, Alameda, & Caramazza, 2010; Kittredge et al., 2008; Piai, Roelofs, & Maris, 2014; Strijkers, Costa, & Thierry, 2010).

Multiword Frequency Effects in Comprehension

There is strong evidence that the frequency of co-occurrence of words affects language comprehension. Numerous reading studies have shown that the time readers spend processing different words depends not only on the frequencies of the individual words but also on their predictability from the preceding or following context. Most studies have investigated the effects of transitional probabilities (TPs) in word pairs (e.g., Balota, Pollatsek, & Rayner, 1985; Engbert, Nuthmann, Richter, & Kliegl, 2005; McDonald & Shillcock, 2003; Reichle, Pollatsek, Fisher, & Rayner, 1998), but there are also studies demonstrating effects of the frequencies of multiword sequences on word reading times (e.g., Arnon & Snider, 2010; Siyanova-Chanturia, Conklin, & van Heuven, 2011; Tremblay, Derwing, Libben, & Westbury, 2011; Underwood, Schmitt, & Galpin, 2004; Valsecchi et al., 2013). Similarly, there are studies of spoken language comprehension demonstrating processing advantages for words predicted by their contexts (e.g., Huettig, 2015; Kuperberg & Jaeger, 2016). Finally, Jacobs, Dell, Benjamin, and Bannard (2016) showed that recognition memory for adjective noun phrases was affected, albeit in different ways, by the frequencies of the phrases and the frequencies of the individual words (see also Tremblay & Baayen, 2010). In short, the ease of reading, listening, and recognizing word sequences is systematically affected by the probabilities of words occurring together.

Different accounts of phrase effects have been proposed. In the reading and speech processing literature, they are often ascribed to associative, predictive, or integrative mechanisms, which may operate at the conceptual or lexical level of processing. In addition, it has been proposed that very frequent word combinations, sometimes called *lexical bundles*, are stored as units in the mental lexicon (e.g., Bybee & Scheibman, 1999). Tremblay and Baayen (2010) have suggested that lexical storage may not be confined to the most frequent word combination but that word sequences of any frequency may be stored in the lexicon (see also Conklin & Schmitt, 2008).

Multiword Frequency Effects in Language Production

A number of studies have shown that the frequency of word combinations also affects the way speakers produce them. In a seminal study, Bannard and Matthews (2008; see also Matthews & Bannard, 2010) found that 2- and 3-year-old children were more likely to succeed in repeating four-word phrases that appeared often in a corpus of child-directed speech compared with less frequent phrases. Moreover, the 3-year-olds also completed the frequent phrases faster than the less frequent ones. (There were too few data points for the 2-year-olds to analyze utterance speed.)

Multiword frequency effects have also been found in adult word pronunciation. Bybee and Scheibman (1999) showed that *don't* was more likely to be reduced when it appeared in the contexts in which it is used most frequently, namely, following the pronoun *I* (rather than other pronouns) and following a small set of verbs (*know*, *think*, *have*) than in less frequent contexts. Moreover, Arnon and Cohen Priva (2013, 2014) showed that the spoken durations of words depended on the frequencies of the word trigrams they appeared in, even when the frequencies of the individual words and the bigrams of words were taken into account.

These studies concerned the articulation of utterances. The study that directly motivated the current research was a study of utterance planning by Janssen and Barber (2012). The authors carried out two experiments examining the effects of phrase and word frequency on speech onset latencies. The first experiment was a picture description task conducted with adult speakers of Spanish. The participants produced two types of utterances, noun-adjective phrases (e.g., *ancla gris* [gray anchor]), and noun-noun phrases (e.g., *bombero rana* [fireman frog]). Fifty nouns appeared in utterance-initial position in both phrase types. To elicit noun-adjective phrases, each object was shown in one of 10 colors, avoiding prototypical combinations. To elicit the noun-noun phrases, each of the 50 objects was combined with one of 10 further objects. Janssen and Barber estimated the frequencies of the words and phrases using Google web searches. The participants' speech onset latencies for the phrases were recorded.

In mixed effect analyses, the authors determined the independent effects of the frequencies of the two words and the phrasal units, and the phrase type (noun-adjective or noun-noun). They also assessed the effects of other variables that might affect the speech onset latencies, including the familiarity of the objects, the strength of the object-color associations (as determined in a separate association study), the phonological neighborhood size, age of acquisition, imageability, and concreteness of the first word, type of onset consonant, length of the phrase, and trial. Correla-

tions between predictor variables were taken into account in the analyses. Janssen and Barber (2012) found that the speech onset latencies were best predicted in a model that included the familiarity of the object, the length of the phrase and its frequency, but not the frequencies of the individual words.

In their second experiment, Janssen and Barber (2012) asked French speakers to produce noun-adjective and determiner-noun-adjective phrases. Each of 56 objects was shown in two of four colors, yielding 112 trials for each participant. As in the first experiment, the best model of the speech onset latencies included only the effect of phrase frequency but not the effects of the frequencies of the individual words. Furthermore, Janssen and Barber mentioned an unpublished study in which Dutch speakers produced adjective-noun phrases and determiner-adjective-noun phrases, and also in which only phrase frequency but no noun frequency effects were observed. Janssen and Barber argued that their findings were best explained by assuming that speakers retrieved stored phrase-sized units from the mental lexicon.

Janssen and Barber's (2012) results demonstrate speakers' sensitivity to the frequencies of phrases and are consistent with the multiword frequency effects reported in the previous studies. Given the evidence that reading, speech processing, and pronunciation are affected by multiword frequencies, it is not too surprising to see that utterance planning times also depend on the frequency of the group of words being planned. The finding that the speech onset latencies did not depend on the frequencies of the individual words is more surprising, given the strong evidence from comprehension and phonetic studies of word and phrase pronunciation demonstrating that the frequencies of individual words and word sequences jointly determine how words are processed or produced. In addition, earlier phrase production studies had reported word frequency effects on speech onset latencies (e.g., Alario, Costa, & Caramazza, 2002; Griffin, 2001; Griffin & Spieler, 2006; Janssen & Caramazza, 2011). However, Janssen and Barber point out that in these latter studies, the effects of phrase frequency were not determined, and as word and phrase frequencies tend to be correlated, the reported word frequency effects may have been confounded with phrase frequency effects.

The Present Study

Janssen and Barber's (2012) proposal that speakers store form representations of multiword phrases has important implications for models of lexical representations and lexical access. This is because it runs counter to the received view that speakers compose most phrases by selecting and combining individual words. Of course, there must be stored lexical representations for phrases with noncompositional meaning (*yellow fever*, *white tea*; e.g., Partee, 1995), and it is conceivable that some of these phrases are retrieved as units, without being decomposed into individual words. However, it seems implausible that this should also happen for "ordinary" noun phrases such as *red chair*.

Before further considering the implications of Janssen and Barber's (2012) findings, it seems prudent to seek additional evidence about the importance of word and phrase frequencies for phrase production. This was the main goal of the present research. As in Janssen and Barber's study, line drawings were used to elicit phrases including a noun and an adjective. The experiments were carried out in Dutch, in which the adjective precedes the noun. The

utterances varied in phrase frequency and in the frequencies of the two words. We used the same line drawings and colors as Janssen and Barber, and, as in their study, phrase and word frequencies were correlated. In Experiments 1 and 2, the speakers produced adjective-noun phrases, such as *groene schoen* [green shoe], paralleling Janssen and Barber's Experiment 1. In Experiment 3, the participants included the definite determiner in their utterances, as in *de groene schoen* [the green shoe], paralleling Janssen and Barber's Experiment 2. The participants' utterance onset latencies were recorded. In Experiment 1, we also recorded how long the participants looked at the objects.

Based on the results reported by Janssen and Barber (2012), one would predict the speech onset latencies to depend only on the frequencies of the phrases but not on the frequencies of the individual words. In contrast, based on the results of the studies mentioned above that obtained effects of the frequencies of words within phrases on speech onset latencies, one would predict effects of the frequencies of the individual words, perhaps in addition to a phrase frequency effect. In the first two experiments, there should be an effect of the frequency of the adjective, the first word of the utterance. Whether there should also be an effect of the frequency of the noun should depend on the speakers' planning scope and on the basis of word frequency effects. In Dutch, the grammatical gender of the noun (neuter or non-neuter) determines the form of most prenominal color adjectives (e.g., neuter *rood huis* [red house] vs. non-neuter *rode schoen* [red shoe]). Thus, speakers must retrieve the grammatical representation of the noun (its lemma) in order to generate the form of the preceding adjective (Exceptions are the adjectives *roze* [pink] and *oranje* [orange], which do not change their form). If lemma retrieval is frequency sensitive, utterance onset latencies should depend on the frequencies of the nouns as well. By contrast, if lemma retrieval is not strongly frequency sensitive, and if word frequency effects arise primarily during word form retrieval, a noun frequency effect should only arise if speakers retrieve the phonological form of the noun before speech onset. Several studies have used priming or interference paradigms to examine the phonological planning scope for adjective-noun phrases (e.g., Costa & Caramazza, 2002; Damian & Dumay, 2007, 2009; Jescheniak, Schriefers, & Hantsch, 2003; Schriefers & Teruel, 1999). Their results showed that speakers typically retrieved the forms of both words before speech onset, though studies by Michel Lange and Laganaro (2014) and Schriefers and Teruel (1999) indicated that some speakers used a smaller planning scope. If the speakers in the present study retrieve the forms of both words before speech onset, effects of the frequencies of the adjective and the noun should be seen, replicating Alario et al. (2002).

In Experiment 1, we recorded the participants' eye movements in addition to their speech onset latencies. Each display featured a colored object on the left side of the screen and an arrow, pointing left or right, on the right side (for the use of this paradigm, see also Jongman, Roelofs, & Meyer, 2015; Korvorst, Roelofs, & Levelt, 2006; Roelofs, 2008). The participants first described the object in an adjective-noun phrase and then pressed one of two response buttons to indicate the direction of the arrow. We recorded how long they looked at the line drawing before turning to the arrow. This was done because earlier picture description studies had shown that gaze durations depended on the planning times for words and phrases: Speakers tend to look at each object they

describe until they have retrieved the phonological form of the corresponding utterance (Belke & Meyer, 2007; Griffin, 2001). For instance, speakers producing noun-noun conjunctions such as *shoe and comb* looked longer at objects with low-frequent than with high-frequent names (e.g., Meyer, Sleiderink, & Levelt, 1998), and they looked longer at objects with long than with shorter names (Meyer, Roelofs, & Levelt, 2003). Any adjective and phrase frequency effects seen in the speech onset latencies should also be seen in the timing of the gaze shifts from the object to the arrow. In addition, gaze shift latencies might show an effect of noun frequency, even if such an effect is absent in the speech onset latencies. This is because the phonological form of the noun may or may not be retrieved before speech onset but should be retrieved before the shift of gaze to the arrow.

Experiment 1

Method

Participants.   All participants in the studies described here were recruited from the participant pool of the Max Planck Institute for Psycholinguistics. They were native speakers of Dutch and had normal or corrected-to-normal vision. They were paid €8 ($10) for their participation in the study. Permission to conduct the research was provided by the Ethics Board of the Social Sciences Faculty at Radboud University. Experiment 1 was carried out with 24 participants (six men, mean age = 21.44 years, range = 19 to 32 years).

Materials.

Preparation of visual stimuli.   The materials consisted of 460 colored line drawings of common objects. To construct them, 50 drawings were selected from the database of the International Picture Naming Project (Bates et al., 2003). These drawings were also used by Janssen and Barber (2012). The same 10 color categories were used as in their study. The names of the objects and colors are listed in Appendix A.

As noted, Dutch nouns are marked for grammatical gender. Sixteen of the object names had neuter gender and 34 had non-neuter gender. Two of the 10 adjectives ended in *-e* and had gender-invariant forms (*oranje* [orange] and *roze* [pink]). The remaining adjectives ended in a consonant and appeared in the consonant-final form preceding nouns with neuter gender (as in *rood huis* [red house]) and took the suffix *-e* before non-neuter nouns (as in *rode bel* [red bell]).

Our design was different from Janssen and Barber's (2012): Instead of showing each object in one color, we used all color-object combinations, excluding seven combinations (e.g., red heart, green leaf) that we deemed to be extremely common or prototypical. We also excluded pairs sharing the onset consonant, because onset overlap may facilitate word form retrieval (Damian & Dumay, 2007). We opted for this large set of items after a pilot experiment in which each of 50 objects was combined with one color (as in Janssen and Barber's study) yielded no word or phrase frequency effect.

Name agreement estimates for 47 of the 50 pictures were obtained from Severens, Van Lommel, Ratinckx, and Hartsuiker (2005). The low mean H-value of 0.61 ($SD = 0.52$), computed following Snodgrass and Vanderwart (1980), indicates high name agreement. Age of acquisition (AoA) estimates for 35 object names were available from Moors et al. (2013). The average AoA was 4.89 years ($SD = 1.15$ years). The average length of the object names was 4.81 phonemes ($SD = 1.87$ phonemes).

Frequency estimates for the adjectives and nouns can be obtained from Dutch corpora (CELEX or SUBTLEX), but most of the adjective noun phrases are not attested in these corpora. Therefore, we followed Janssen and Barber (2012) and used Google web searches to obtained frequency estimates for the words and phrases (see also Brysbaert, Keuleers, & New, 2011). We restricted the search to Dutch and web pages from the Netherlands. To determine the stability of the estimates over time, we retrieved frequencies for a random sample of 50 color-object phrases twice, 4 months apart, once on January 20, 2015, and once on May 28, 2015. The log-transformed phrase frequency estimates from the two searches were strongly correlated ($r = .99$, $p < .001$).

The H-indices of the objects were weakly correlated with the log-transformed name frequencies ($r = -.13$). H-Indices refer to the drawings as shown in the norming study rather than the colored line drawings we presented. Therefore, we did not include them in the analyses but used the object verification latencies described in the following rating study as indicators for the ease of object recognition. AoA was moderately correlated with noun frequency ($r = .33$, $p < .001$). As AoA estimates were missing for one third of the items, this variable was not included in the analyses.

As one might expect, the log-phrase frequencies were moderately correlated with the log-adjective frequencies and the log-noun frequencies. Both word and phrase frequencies were correlated with phrase length (see Table 1).

Table 1
Correlations of Item Characteristics in Experiments 1 and 2

Item characteristics	Log-adjective frequency	Log-noun frequency	Phrase length	Log-RT of color verification	Log-RT of object verification	Familiarity rating
Log-phrase frequency	.33**	.41**	−.30**	−.13**	.01	.17**
Log-adjective frequency		.11**	−.12*	−.18**	.04	.23**
Log-noun frequency			−.45**	.00	.01	−.02
Phrase length				−.02	.15**	.01
Log-RT of color verification					−.05	.06
Log-RT of object verification						.05

Note.   RT = reaction time.
* $p < .05$.   ** $p < .01$.

The colored line drawings were scaled to fit into virtual frames of 4 × 4 cm (about 4° visual angle for the participant). They appeared on a white background centered in the left half of the computer screen. An arrow with a length of 0.75 cm appeared centered in the right half of the screen. It was flanked on each side by two "x"s so that the participants could not determine the direction of the arrow without making an eye movement toward it. The distance between the midpoints of the two stimuli was 11.5 cm. Two hundred thirty randomly selected pictures were combined with an arrow pointing leftward, and the remaining 230 pictures appeared with an arrow pointing rightward.

A concern one might have about these materials is that any observed frequency effects may be WYeeV93+ caused by differences in the ease of recognition of the objects or in the participants' familiarity with the objects, the colors, or their combinations. Three studies, were conducted to address this issue: A color and an object verification study were conducted to determine the ease of recognition of the colors and the objects, respectively, and a rating study was conducted to assess how familiar participants were with objects appearing in particular colors.

Color and object verification studies. On each trial of the color verification study, participants saw a printed color name followed by a line drawing of an object, and decided as quickly as possible whether or not the written word and the line drawing matched in color (e.g., *rood* [red] followed by a red line drawing) or mismatched (e.g., *rood* followed by a green line drawing; for the use of similar tasks, see Jescheniak & Levelt, 1994; Stadthagen-Gonzalez, Damian, Pérez, Bowers, & Marín, 2009).

The study was carried out with 22 participants (three men, mean age = 22.4 years). The materials consisted of the line drawings used in the main experiment and the printed names of the 10 colors. The color names were shown in Arial font size 20. Each line drawing was combined once with the correct color name to form a matching pair and once with a randomly selected incorrect color name to a mismatching pair. Across the 500 mismatching pairs, all color adjectives were used equally often. Each participant saw all 1,000 word-picture combinations. Four different random orders of the stimuli were used.

On each trial, a fixation cross was presented for 500 ms, followed by a written word. Participants were asked to place their index fingers on the "z" and "m" keys of a keyboard and to press either key to indicate that they had read the word. A line drawing was then presented until the participant responded, maximally for 4 s. Participants pressed the "z" key, labeled "JA" (yes), to indicate a match between word and picture and the "m" key, labeled "NEEN" (no) to indicate a mismatch. After 1.5 s, the next trial began. The experiment took approximately 40 min to complete.

The error rate was 5.06% (*SD* = 3.24) in the match condition and 3.27% (*SD* = 1.94) in the mismatch condition. Incorrect responses and outliers, defined as latencies deviating by more than 2.5 standard deviations from the participant's mean (4.76% of the data), were excluded from the analyses. The average response latency for correct responses was 491 ms (*SD* = 99 ms) on match trials and 503 ms (*SD* = 88 ms) on mismatch trials.

The object verification experiment was identical to the color verification experiment except that the line drawings were preceded by matching or mismatching object names rather than color names. The experiment was carried with 22 participants (three men, mean age = 23.4 years). On match trials, the line drawings were preceded by the written object names and on mismatch trials by the names of randomly selected other objects in the set. The error rate was 4.4% (*SD* = 2.02) in the match condition and 2.89% (*SD* = 1.82) in the mismatch condition. Errors and outliers (2.5% of the data) were excluded from the analyses. The average response latency was 496 ms (*SD* = 99 ms) on match trials and 530 ms (*SD* = 86 ms) on mismatch trials.

As Table 1 shows, the log-transformed color verification latencies from match trials correlated significantly, though not strongly, with the log-transformed frequencies of the color names. The log-transformed object verification latencies did not correlate with the log- transformed frequencies of the objects.

Rating study for color-object familiarity. It is possible that phrase frequencies are correlated with the participants' familiarity with different color-object combinations. For instance, as shoes are more often brown than green, people might hear and say "brown shoe" more often than "green shoe." Alternatively, color-object familiarity and phrase frequency might be independent of each other, as people might not mention expected properties of objects very often. However, color-object familiarity might still affect the ease of phrase production. For instance, self-monitoring processes carried out before speech onset may be completed faster when the planned utterance matches the speaker's experience with the object than when there is a mismatch. In short, color-object familiarity might affect speech onset latencies in two ways, mediated through the frequencies of the corresponding phrases or independently of phrase frequencies.

To assess these possibilities, we conducted a web-based rating study, programmed in LimeSurvey 2.05 (www.limesurvey.org), in which 63 participants (15 men, mean age = 25.6 years) indicated how often they had encountered each of the objects shown in Experiments 1 in different colors. On each trial, they saw the Dutch translation equivalent of the question "In which colors have you ever seen a [noun] (e.g., car)?" at the top of the screen. The names of the 10 colors used in the study appeared in a column on the left side of the screen. Their order varied randomly from trial to trial. On the right side of the screen, there was a row of empty boxes. Participants were asked to drag the name of the most common color of the object to the top box, and then enter the name of the second-most common color into the second box, and so on, until they had selected and ranked all colors in which they had experienced the object. The instructions stressed that participants should indicate in which colors they had actually seen the objects rather than indicating in which colors the objects could potentially appear. Participants could revise their ratings. The study was self-paced. The order of the objects was randomized across participants.

To quantify the participants' familiarity with the color-object combinations, the selected colors were assigned scores from 10 (first rank) to 1 (lowest rank). Colors not included in the ranking received a score of zero. We multiplied the scores by the number of participants assigning it to an object-color combination, added up the resulting values, and divided the sum by the total number of participants responding to the object-color combination. For instance, if, out of 63 participants, 40 had indicated that black was the most common color of a piano and 23 participants had indicated that black was the second most common color, the overall score for black piano would be $(40 \times 10 + 23 \times 9)/63 = 9.64$. The familiarity scores could range between 10 (all participants

chose the same color as the most common color for an object) and 0 (no participant indicated that they had encountered the object in that color). We explored the usefulness of related scoring methods (e.g., only considering the first choices), but using different scoring methods did not change the results of the analyses.

The color-object familiarity scores ranged from 1 to 10, with an average of 5.18 ($SD = 2.78$). The scores correlated with frequency of the phrase ($r = .17$, $p < .001$). Thus, there was a weak tendency for phrases with higher frequencies to refer to more familiar color-object combinations.

Design of the main experiment. In the main experiment, each participant saw all 460 colored line drawings. The order of the drawings was random, with the restriction that the same color or the same object did not appear on successive trials. Four pseudorandomized lists were created. Each list was seen by six participants.

Procedure. Participants were tested individually. They were first given a booklet showing the objects and colors and their names. They were asked to familiarize themselves with the materials and to use only the names given in the booklet to refer to the objects and colors. Then the eye tracker was calibrated and the experiment began. Each trial began with the presentation of a fixation cross. Participants were asked to look at the cross. This allowed for drift correction. WYeeV93+ Seven hundred ms after the onset of fixation, the cross was replaced by one of the colored objects. Participants were asked to describe the object in an adjective-noun phrase and then to press the left or right button on a two-button panel to indicate the direction of the arrow. The trial was terminated by the button press. The intertrial interval was one second. The items were presented in five blocks of 92 trials each. Participants could take short breaks between blocks. At the beginning of each block, the eye tracker was recalibrated.

Apparatus and software. Eye movements were recorded using an Eyelink 1000 Tower Mount eye tracker. The movements of the participant's right eye were tracked throughout the experiment with a sampling rate was 1 kHz. Stimulus presentation was controlled by the software package Experiment Builder (Version 1.10.1241). The stimuli were presented on an Acer TCO03 20-in. monitor. A Sennheiser ME64 microphone was used to record the participants' speech. Speech onset latencies were measured online using a voice key and were later checked and corrected using the speech analysis program Praat (Boersma, 2001). Participants used a custom-made two-button panel to indicate the direction of the arrow.

Data coding and analysis. Responses were coded as errors when participants produced object or color names that were different from the names they had been familiarized with or when responses began with or included repairs or disfluencies. Analyses of response latencies and eye movements were based only on correct responses. We excluded responses with speech onset latencies below 300 ms or deviating by more than 2.5 standard deviations from a participant's mean. The manual responses to the arrows were coded as correct or incorrect. We did not exclude speech onset or gaze shift latencies from trials in which participants carried out incorrect manual responses. However, we did exclude trials in which participants carried out the manual response before speech onset, because on those trials, they did not follow the instruction to prioritize the verbal task. We also excluded trials in which the manual response latencies were longer than 6,000 ms because such long latencies suggest lapses of attention.

To analyze the participants' eye movements, virtual squares of 6×6 cm around the line drawing and the arrow were defined as regions of interest. As the location of the fixation cross coincided with the center of the line drawing and as a gaze to the fixation cross triggered the presentation of the experimental display, the participants usually fixated upon the line drawing when it appeared or shortly afterward. We excluded trials in which no fixation to the line drawing was detected within 500 ms after trial onset, because on these trials, participants had not followed the instruction to look at the fixation cross at trial onset or had blinked. In addition, trials with gazes to the objects that were longer than 3,000 ms were excluded because these long gazes were likely to be indicative of lapses of attention.

After these exclusion criteria had been applied, one participant was excluded because their rate of valid responses was only 52%. In the data set from the remaining participants, there were six items (*oranje hond* [orange dog], *roze hond* [pink dog], *geel oog* [yellow eye], *roze oog* [pink eye], *gele paperclip* [yellow paperclip], *oranje paperclip* [orange paperclip]) with seven or fewer valid responses. These items were also excluded from the analyses.

From the remaining data set, 8.91% of the data were excluded from the speech onset analyses and 16.44% from the gaze shift analyses. Specifically, 0.62% of the trials were excluded from the set of speech onset latencies because the responses to the arrow preceded the naming responses. From the remaining set of speech onset latencies, 5.61% as naming errors, and 2.04% because the latencies were below 300 ms or deviated more than 2.5 standard deviations from a participant's mean. From the set of gaze shift latencies, an additional 6.53% of the data were excluded because gaze onset latencies were above 500 ms, 1.00% because gaze shift latencies were above 3000 ms, and 0.66% because manual response latencies were above 6,000 ms. The proportion of data loss is in line with earlier studies recording eye movements during picture naming (e.g., Roelofs, 2008).

For the valid trials, we computed logit rates of naming errors and log-transformed latencies of the speech onset, gaze shifts from the picture to the arrow, and the manual response latencies. All dependent variables were submitted to mixed effects regression modeling analyses in R Version 3.2.0 (R Core Team, 2015), using the packages LanguageR (Baayen, 2013) and lme4 (Bates, Maechler, Bolker, & Walker, 2014). All variables were centered. By-participant intercepts and random slopes for all fixed effects were included as random effects.

We used a stepwise additive model building procedure to examine the effects of log- adjective, log-noun, and log-phrase frequencies on error rates, speech onset latencies, and gaze durations. For each dependent variable, we first constructed a baseline model that contained variables as fixed effects that were significantly correlated with the dependent variable but not of primary theoretical interest. Specifically, the baseline model for the logit rates of naming errors contained trial (capturing practice and repetition effects) and log-reaction time (RT) of color verification; the baseline model for speech onset latencies additionally contained familiarity rating; and the baseline model for gaze duration additionally contained phrase length. We then added the three fixed effects of log-adjective, log-phrase, and log-noun frequencies. The order of entry corresponded to the strength of the

correlation coefficients between the fixed effects and depend measure, starting with the strongest effect. A frequency effect was treated as significant when including it significantly improved the model using likelihood ratio tests. The Akaike information criterion for each model is reported in Appendix B. As Table 1 shows, the predictors were moderately correlated. However, in all regression models, the tolerance statistics were above 2, and the average variance inflation factor was smaller than 1.5 (Field, 2009). Thus, collinearity should not be problematic for the analyses.

Results

Naming errors. The rate of naming errors was 5.61% (SD = 1.18%). The most common errors were incorrect color names (1.73%); others were hesitations (1.56%), incorrect object names (0.70%), missing responses (0.46%), and other types of errors (1.16%). Results of the baseline model revealed a negative effect of trial (z = −3.40, p < .001), indicating that the error rates decreased across trials, and a positive effect of log-RT of color verification (z = 4.08, p < .001), indicating that the error rates increased with increasing color verification time. Adding log-adjective frequency significantly improved the model (χ^2 = 13.27, p = .02), though log-adjective frequency did not significantly predict the logit-error rates. Neither log-noun frequency nor log-phrase frequency significantly contributed to the model (χ^2s < 11.63, ps > .11). Results of the final models for error rates, speech onset latencies, gaze shift latencies, and manual responses are reported in Appendixes C and D.

Speech onset latencies. Correct responses were initiated with a mean latency of 1,073 ms (SD = 91 ms) after picture onset. In the baseline model, we found a positive effect of log-RT of color verification (t = 2.76) and a negative effect of trial (t = −5.30). Thus, the speech onset latencies increased with increasing log-RT of color verification and decreased across trials. Including log-adjective frequency (χ^2 = 173.52, p < .001) significantly improved the model, as did including log-phrase frequency (χ^2 = 15.30, p = .02). Both log-adjective and log-phrase frequency showed a negative effect, implying that higher frequencies were associated with shorter speech onset latencies. Including log-noun frequency did not improve the model fit (χ^2 = 10.81, p = .15).

Gaze shift latencies. The average latency of the gaze shift from the object to the arrow was 1,747 ms (SD = 127 ms) after picture onset, corresponding to 674 ms after speech onset. The baseline model was similar to the baseline model for speech onset latencies: The effects of log-RT of color verification, familiarity rating, and trial were significant (t = 4.87, t = −4.22, and t = −2.33, respectively). Length in phonemes also had a significant effect (t = 6.12), with slower gaze shifts being associated with longer phrases. The model improved significantly when log-adjective and log-phrase frequency were added (χ^2 = 45.76, p < .001, and χ^2 = 16.56, p = .035, respectively), but there was no further improvement when log-noun frequency was added as well (χ^2 = 4.08, p = .90). Thus, the pattern of results was the same as seen for the speech onset latencies.

Manual response latencies. The manual responses typically followed the shift of gaze to the arrow, with an average latency of 1,981 ms as measured from target onset (SD = 334 ms), corresponding to 234 ms after gaze shift and 908 ms after speech onset. The manual response latencies were highly correlated with the gaze shift latencies (r = .65, p = .001; both measured from picture onset), and showed significant effects of log-RT color verification, familiarity rating, and trial (t = 3.81, t = −2.82, and t = −3.26, respectively). Moreover, the manual response latencies were significantly predicted by log-adjective frequency, log-phrase frequency, and log-noun frequency (t = −2.73, t = −2.94, and t = −3.93, respectively). The model significantly improved with the inclusion of each of these predictors (log-adjective frequency, χ^2 = 43.05, p < .001; log-phrase frequency, χ^2 = 25.99, p = .035; log-noun frequency, χ^2 = 39.59, p < .001).

Discussion

The latencies of the speech onsets and the gaze shifts from the objects to the arrows depended on the frequencies of the phrases and the adjectives: Higher frequencies were associated with faster responses. The frequency effects on the gaze durations and manual responses are important because they show that the frequency effects arise during the planning of the utterances and are not WYeeV93+ caused by differences in the time required for the articulatory preparation of different phrases or in the accuracy of the measurements for different onset consonants. Janssen and Barber (2012) also found phrase frequency effects on speech onset latencies for phrases, but they did not find additional word frequency effects. This held for experiments carried out in Spanish and French and for experiments in Dutch (reported in an unpublished study mentioned by Janssen and Barber). However, our findings are consistent with results of numerous other studies reporting word frequency effects in word and phrase production tasks (e.g., Alario et al., 2002; Cuetos et al., 2010; Griffin, 2001; Jescheniak & Levelt, 1994). These studies suggest that phrases are produced by retrieving and combining individual words and that the time required for these retrieval processes depends, among other things, on the frequencies of the individual words.

On this view, the absence of noun frequency effects in our study is puzzling, especially because a similar phrase production study with English speakers found that the speech onset latencies depended on the frequencies of the adjectives as well as the nouns (Alario et al., 2002). The phonological forms of most of the adjectives used in our study depended on the grammatical gender of the noun. Therefore, the participants had to retrieve the grammatical gender of the noun early enough to affect the phonological encoding of the adjective, and one might therefore expect noun frequency effects to be more likely to arise in Dutch than in English, in which the form of prenominal adjectives does not depend on properties of the following nouns. However, in spite of the marking of adjectives for noun gender, the participants might not always have retrieved the noun before speech onset. Two of the adjectives used in the study (*oranje* and *roze*) had the same form before neuter and non-neuter nouns. Moreover, 16 of the 50 nouns had non-neuter gender and required the base form of the adjective. Thus, changing the form of the adjective from the monosyllabic base to the disyllabic form featuring the suffix *-e* was only necessary on 54% of the trials. Speakers perhaps retrieved information about the adjective and noun in parallel, as, for instance, proposed by Jescheniak et al. (2003), first prioritizing the retrieval of the adjective, initiated the utterance as soon as the phonological form of the adjective stem was available, and adjusted its form, as necessary, when information about the noun gender became avail-

able. In other words, the utterances could be initiated without knowledge of the gender of the noun, with the adjective form being finalized after speech onset. This would explain the absence of a noun frequency effect on the speech onset latencies. This hypothesis is consistent with evidence suggesting that speakers can initiate utterances as soon as they have planned a single onset consonant (Kawamoto, Liu, Lee, & Grebe, 2014).

This hypothesis implies that noun retrieval continued after speech onset. Therefore, one might expect a noun frequency effect on the gaze shift latencies, which should capture the entire time required for the planning of a phrase to the level of the phonological form, and in the manual responses to the arrows. The gaze shift latencies also depended only on the frequencies of the adjectives and the phrases and on their length but not on the frequencies of the nouns. By contrast, for the manual response latencies a noun frequency effect was seen, in addition to adjective and phrase frequency effects. This pattern points to late retrieval of the frequency sensitive representations of the noun: Apparently, noun retrieval was completed after the shift of gaze away from the line drawing but before the initiation of the manual response to the arrow. This conclusion is not entirely consistent with the earlier evidence suggesting that speakers tend to fixate upon the objects they describe until they have completed the phonological encoding of the relevant utterance (e.g., Meyer et al., 1998). However, results obtained by Meyer, Wheeldon, van der Meulen, and Konopka (2012) suggested that speakers may deviate from this default strategy and move their gaze earlier to a new stimulus when the linguistic task is very easy. This may have been the case in Experiment 1, as the materials featured many repetitions of the color adjectives and nouns, and was trivially easy for our adult participants.

Experiment 2

Experiment 1 replicated the phrase frequency effect reported by Janssen and Barber (2012) and, in addition, yielded an effect of the frequency of the first word of the utterance. An important difference between the studies is that only the present Experiment 1 featured a secondary task. In order to facilitate comparisons across the studies, we conducted a second experiment, in which participants only produced adjective-noun phrases without having to carry out a second task.

Method

Participants. Experiment 2 was carried out with 25 native Dutch speakers (four men, mean age = 22.48 years, range = 19 to 28 years).

Materials and design. The same materials were used as in Experiment 1 (see Appendix A), except that the fixation cross and the line drawings now appeared in the center of the screen, and the arrow with flanking "x"s was not presented any more. The design was the same as in Experiment 1.

Procedure. At the beginning of the session, the participants were familiarized with the materials in the same way as in Experiment 1. They were then asked to describe each of the line drawings in an adjective-noun phrase. Two examples were provided. On each experimental trial, a fixation cross was presented for 500 ms and was replaced by a colored line drawing. The trial

was terminated when the voice key detected an utterance onset and, at the latest, 1,500 ms after picture onset. The intertrial interval was 1,000 ms. For the statistical analyses, the voice key latencies were replaced by latencies measured off-line using the software package Praat 5.1.

Apparatus and software. The experiment was controlled by the software package Presentation (Version 14.3; www.neurobs.com). Stimuli were presented on a Samsung screen. A Sennheiser ME64 microphone was used to record the responses.

Data coding and analysis. The same exclusion criteria were applied to the speech onset latencies as in Experiment 1. Five items (*geel oog* [yellow eye], *oranje fles* [orange bottle], *rood sandwich* [red sandwich], *rood vork* [red fork], *roze zaag* [pink saw]) were excluded from the analyses WYeeV93+ because of error rates exceeding 50%. From the remaining data, we removed trials with naming errors and outliers, defined as latencies below 300 ms or more than 2.5 standard deviations above a participant's mean (0.92% of the data). Logit naming error rates and log-transformed onset latencies were submitted to mixed effects regression modeling analyses as in Experiment 1. The baseline model for the rates of naming error and log-transformed speech onset latencies contained the log RT of color verification and trial. Frequencies were added to the baseline model in the order log-phrase frequency, log-noun frequency, and log-adjective frequency. The order of entering the fixed effects corresponded to the strength of their correlations with speech onset latencies (see Appendixes B to D).

Results

Naming errors. The average rate of naming errors was 4.25% (SD = 5.16%). The most common errors were incorrect color names (2.30% of the trials) and hesitations (2.05%). Other error types were incorrect object names (0.28%), missing responses (0.25%), and other types of errors (1.50%). In the baseline model for the error rates, we found a negative effect of trial (z = −2.99, p = .003), and a positive effect of log-RT of color verification (z = 4.50, p < .001), indicating that the error rates decreased across trials and increased with increasing verification time for the colors. The model did not improve significantly with the addition of log-adjective frequency (χ^2 = 0.59, p = .55), log-noun frequency (χ^2 = −1.56, p = .12), or log-phrase frequency (χ^2 = 0.33, p = .74).

Speech onset latencies. The mean speech onset latency was 946 ms (SD = 250 ms). Results of the mixed effect models were in line with the results of Experiment 1. In particular, a positive effect of log-RT of color verification (t = 4.82) and a negative effect of trial (t = −2.32) were found in the baseline model. The model was significantly improved when including log-adjective frequency (χ^2 = 46.61, p < .001) and log-phrase frequency (χ^2 = 19.88, p = .002). By contrast including log-noun frequency did not improve the model significantly (χ^2 = 5.48, p = .60).

Discussion

The results of Experiment 2 closely replicated those of Experiment 1: The speech onset latencies depended on the frequencies of the phrases and adjectives but not on the frequencies of the nouns. Thus, the central finding of Janssen and Barber's (2012) study, the phrase frequency effect, was obtained again. In addition, there was

an effect of the frequency of the first word of the phrases, which had also been seen in Experiment 1 of the present study but not by Janssen and Barber. The two experiments of the current study differed in whether or not the participants carried out a second task after describing the line drawings. Without the second task, the participants responded somewhat faster (with average latencies of 946 ms compared with 1,073 ms), but in both task contexts, they appeared to have used the same planning strategy of retrieving the adjective but not necessarily the noun before speech onset. This explains the presence of an adjective frequency effect and the absence of a noun frequency effect.

An alternative account of the word frequency effects seen in the two studies is that the alleged adjective frequency effects were chance findings or caused by confounding variables. There were only 10 adjectives, and differences in the time required to initiate their production may have stemmed from differences in their phonetic structure, for instance, their onset segments, their phonological neighborhood densities, or semantic features that were not controlled for (e.g., Mehta & Zhu, 2009). In that case, our results would indicate that there was no genuine word frequency effect, but only a phrase frequency effect, as observed by Janssen and Barber (2012). Experiment 3 was conducted to address this issue and to assess whether a noun frequency effect would be seen when speakers were forced to retrieve the noun gender before speech onset.

Experiment 3

Experiment 3 was identical to Experiment 2 except that the participants now included the definite determiners *de* or *het* in their utterances producing phrases such as *de groene schoen* [the green shoe] or *het groene huis* [the green house], depending on the gender of the noun. Because the determiner was the first word of the utterance, the participants now had to retrieve the gender of the noun before speech onset (e.g., Costa, Sebastian-Galles, Miozzo, & Caramazza, 1999). Thus, the experiment allowed us to test the view that no noun frequency effect had been seen in Experiments 1 and 2, because information about the noun gender was not necessary to initiate the utterance but was only needed a little later to finalize the sound form of the adjective. In Experiment 3, it was necessary and, consequently, a noun frequency effect should be seen, provided, of course, that the retrieval of the relevant noun representation is frequency sensitive. A noun frequency effect, if observed, would be less likely to be due to confounded with other variables than the adjective frequency effect seen in Experiment 1 because the set of nouns was much larger than the set of adjectives (50 vs. 10 words). Moreover, all utterances began with the determiner *de* or the determiner *het*, and systematic effects of different onset consonants or different genders on the speech onset latencies could be determined in the statistical analyses and separated from frequency effects.

In addition to assessing the effect of noun frequency, Experiment 3 served to examine again whether the speech onset latencies would depend on the frequencies of the entire phrases, which now included three rather two words. As such, the results of Experiment 3 can be compared with those of Experiment 2 in Janssen and Barber's (2012) study, which also featured phrases starting with definite determiners.

Method

Participants. Experiment 3 was carried out with 22 native Dutch speakers (five men, mean age = 22.25 years, range = 19 to 27 years).

Materials and design. The same line drawings were used as in Experiments 1 and 2 (see Appendix A). As before, we used Google web searches to obtained frequency estimates for the phrases. The frequencies of these three-word phrases were positively correlated with the frequency of the adjective-noun phrases tested in Experiments 1 and 2 ($r = .67$, $p < .001$), and with the frequencies of the adjectives and the nouns and the familiarity rating for the objects. These correlations were slightly weaker than those in the materials of Experiment 1. Note that following the definite determiner, the adjective always takes the suffix *-e*; thus, there were 10 surface forms of adjectives in this experiment compared with 18 in Experiments 1 and 2. There were no significant correlations of the phrase frequencies with the log-RT of color verification or object verification (see Table 2).

Following Janssen and Barber (2012), we computed the forward and backward TPs between the pairs of words (determiner-adjective and adjective-noun), and the TPs of all three words in a phrase, by multiplying these TPs. Forward TP for the three words was correlated with log-adjective frequency ($r = .17$, $p < .001$), log-noun frequency ($r = .35$, $p < .001$), and log-phrase frequency ($r = .61$, $p < .001$). Backward TP for the three words in the phrase was correlated with log-adjective frequency ($r = .19$, $p < .001$) and log-phrase frequency ($r = .46$, $p < .001$) but not with log-noun frequency ($r = -.11$, $p = .02$). We included the TPs in some of the analyses 1 (see below[1]).

Procedure. The procedure was exactly the same as in Experiment 2, except that the participants were now asked to include the appropriate determiners in their utterances. Two examples were provided in the instructions.

Data coding and analysis. Incorrect responses were excluded from the analysis of the speech onset latencies. Latencies that were 2.5 standard deviations beyond a participant's mean as well as latencies below 300 ms were likewise excluded (1.37% of the data). Logit-error rate and log-transformed speech onset latencies were submitted to mixed-effects modeling analyses. All variables were centered. The baseline model for the logit-error rate contained the familiarity rating, trial, and determiner (sum contrast coded with -1 for *de* and $+1$ for *het*). Frequencies were entered into the baseline model in the order log-noun frequency, log-phrase frequency, and log- adjective frequency. The entry order of the fixed effects corresponded to the strength of their correlations with error rate. The baseline model for the log-transformed speech onset latencies contained the log RT of color verification, phrase length, trial, and determiner. Frequencies were added to the baseline model in the order log-phrase frequency, log-noun frequency, and log-adjective frequency. Again, the order of entering the fixed effects corresponded to the strength of their correlations with speech onset latencies (see Appendixes B, C, and F).

---

[1] In Experiment 1, the forward and backward transitional probabilities were highly correlated with the phrase frequencies ($rs > .90$; similar to the correlations in Janssen & Barber's, 2012, study); therefore, including them in the analyses was not useful.

Table 2
Correlations of Item Characteristics in Experiment 3

Item characteristics	Log-adjective frequency	Log-noun frequency	Phrase length	Log-RT of color verification	Log-RT of object verification	Familiarity rating
Log-phrase frequency	.20**	.37**	−.14**	−.09	−.02	.16**
Log- adjective frequency		−.01	.16**	−.18**	−.04	.23**
Log-noun frequency			−.28**	.00	−.01	−.02
Phrase length				−.03	.14*	.14**
Log-RT of color verification					−.05	.06
Log-RT of object verification						.05

Note. RT = reaction time.
* $p < .05$. ** $p < .01$.

Results

Naming errors. WYeeV93+ Because of a technical error, 4.69% of the data, distributed randomly across participants, were lost. The error rate in the remaining data set was 9.14% (SD = 4.40%). The most common errors were hesitations (3.90%), followed by incorrect color names (2.28%), incorrect determiners (1.66%), incorrect object names (0.83%), and other types of errors (0.40%). None of the predictors in the baseline model were significant (for determiner, $z = 1.22$, $p = .22$; for trial, $z = -1.60$, $p = .11$; for familiarity rating, $z = -1.46$, $p = .14$). Including log-noun, log-adjective, or log-phrase frequency did not improve the model ($\chi^2 = 9.35$, $p = .15$, $\chi^2 = 1.18$, $p = .99$, and $\chi^2 = 12.99$, $p = .11$, respectively).

Speech onset latencies. The average speech onset latency for correct responses was 991 ms (SD = 148 ms). In the baseline model for the log-transformed speech onset latencies, there were significant effects of log-RT of color verification ($t = 3.22$), phrase length ($t = 4.96$), and determiner ($t = -4.16$), with longer latencies associated with longer verification RT, longer phrases, and phrases starting with *het* (the less common determiner) rather than *de*. Model fit improved with the addition of the phrase and noun frequencies (log-phrase frequency, $\chi^2 = 31.29$, $p < .001$; log-noun frequency, $\chi^2 = 18.49$, $p = .02$) but not with the addition of the adjective frequency (log-adjective frequency, $\chi^2 = 6.81$, $p = .66$). The log-transformed speech onset latencies decreased with increasing log-noun and log-phrase frequency.

Additional analyses were run to evaluate the effects of the TPs of the words in the phrases. TPs were included in the models in addition to or instead of the phrase frequencies. Analyses were carried out separately for forward and backward TP. All analyses yielded the same results: Including TPs in the models did not improve model fit.

Discussion

Experiments 1 and 2 showed that the speech onset latencies for adjective-noun phrases depended on the frequencies of the adjectives but not on the frequencies of the nouns. One account of this pattern is that frequency-sensitive retrieval processes for the nouns were carried out too late to affect phrase onset latencies. Thus, a noun frequency effect should be obtained when noun information has to be retrieved prior to speech onset. Experiment 3 tested this prediction by eliciting determiner-adjective-noun phrases. Because the form of the determiner depends on the gender of the noun,

participants now had to retrieve noun information, minimally its grammatical gender, before initiating the response (for discussion of determiner selection in Dutch, see Dhooge, De Baene, & Hartsuiker, 2016; Roelofs, 1992; Schriefers, de Ruiter, & Steigerwald, 1999; Schriefers, Hantsch, & Jescheniak, 2002). In line with this expectation, a noun frequency effect on the speech onset latencies was now obtained.

This result is important because in Experiments 1 and 2, no noun frequency effect on utterance latencies was observed. There was an adjective frequency effect, but the set of adjectives was small, and adjective frequency may have been confounded with uncontrolled variables. For the much larger sets of nouns (50 rather than 10 items), such an account is less likely. Because all utterances began with either de or *het*, and this variable was included in the baseline model, the frequency effect could be separated from any effects WYeeV93+ related to differences in the phonetic onsets of the utterances. In short, Experiment 3 provides good evidence that speech onset latencies for Dutch determiner-adjective-noun phrases depend on retrieval of the noun, which, in turn, depends on its frequencies of the nouns. It was important to establish the noun frequency effect, because the goal of the study was to determine whether speech onset latencies for phrases depended exclusively on the frequencies of the phrases or also on the frequencies of the component words.

The adjective frequency seen in Experiments 1 and 2 was not replicated. Experiment 1, testing the production of adjective-noun phrases, yielded an adjective but no noun frequency effect, whereas Experiment 2, testing determiner-adjective-noun phrases, yielded a noun but no adjective frequency effect. Thus, a frequency effect was seen for the word that had to be retrieved first. In Experiment 3, this was the noun because its gender determined the form of the utterance-initial determiner. This pattern suggests that the participants initiated their utterance as soon as the information minimally needed to do so was available. In Experiment 1, they prioritized the retrieval of the adjective and began to speak as soon as the base form of the adjective had been retrieved. By contrast, in Experiment 2, they prioritized the retrieval of the noun and began to speak as soon as its gender and the corresponding form of the determiner were available.

In addition to the noun frequency effect, a phrase frequency effect was observed. This finding replicates a central finding of Experiments 1 and 2, and of the experiments reported by Janssen and Barber (2012). Thus, as in Janssen and Barber's study, we found that speakers were faster to initiate high-frequency phrases than phrases lower in frequency, and this held for phrases with and without determiners.

General Discussion

Janssen and Barber (2012) reported two phrase production studies carried out with Spanish and French speakers, in which they found that the speech onset latencies for the phrases depended exclusively on the frequencies of the entire phrases and not on the frequencies of the individual words. This held for three phrase types: noun-adjective phrases, determiner-noun-adjective phrases, and noun-noun conjunctions. The authors suggested that word combinations might be stored as units in the mental lexicon. The goal of the present experiments was to determine whether parallel patterns of results would be obtained for Dutch adjective-noun and determiner-adjective-noun phrases. Their results were clear-cut. As in Janssen and Barber's study, speech onset latencies were predicted by the frequencies of the phrases. In addition, they were predicted by the frequencies of the adjectives in Experiment 1 and by the frequencies of the nouns in Experiment 2. We first discuss the word frequency effects and then turn to the phrase frequency effects.

Word Frequency Effects Reflect on Planning Strategies

The adjective frequency effect seen in Experiment 1 is consistent with the results of numerous earlier reports of frequency effects for single word production (e.g., Cuetos et al., 2010; Jescheniak & Levelt, 1994) and effects of the frequency of the first word of longer utterances (e.g., Alario et al., 2002; Griffin & Spieler, 2006). Opinions about the origin of word frequency effects on speech onset latencies are divided, but most of the recent evidence suggests that word frequency affects not only the speed of access to word forms but also, perhaps to a lesser extent, the speed of access to semantic and grammatical representations of words (e.g., Kittredge et al., 2008). Despite the remaining uncertainty about their origins, word frequency effects indicate that words are major processing units in speech production. This is a central claim of the classic "words and rules" view.

If, as the classic view holds, speakers plan phrases by retrieving and combining individual words, one might expect frequency effects for both words of adjective-noun phrases. Contrary to this expectation, the frequencies of the nouns did not predict the speech onset latencies in Experiments 1 and 2. To account for this finding, we proposed above that the participants retrieved those representations of the nouns that are most sensitive to frequency too late for them to have an effect on the speech onset latencies. To assess this hypothesis, we carried out Experiment 3, in which participants had to include the definite determiners in the noun phrases. As the forms of Dutch determiners depend on the grammatical gender of the noun, noun information now had to be retrieved before speech onset. Consequently, a noun frequency effect was now observed.

Given that Dutch definite determiners depend on the grammatical gender but not the sound form of the noun, one might want to conclude that the retrieval of the grammatical representations of words, their *lemmas*, is frequency sensitive. However, we cannot be sure that the speakers of Experiment 3 *only* retrieved the necessary grammatical information about the nouns before speech onset; maybe they retrieved the phonological forms of the nouns as well. Starreveld and La Heij (2004) reported a series of picture-word interference experiments in which target pictures were combined with phonologically related or unrelated distractors. Phonological facilitation was

obtained not only when participants produced bare nouns but also when the task was to produce only the definite determiner or to perform a gender-decision task for the object names (see also Navarrete, Basagni, Alario, & Costa, 2006). The latter two tasks could, in principle, be carried out without retrieval of phonological information, but the results suggest that phonological information was retrieved nonetheless. Bürki, Sadat, Dubarry, and Alario (2016) reported an Electroencephalogram (EEG) study using the picture-word interference paradigm, in which German speakers produced determiner-noun phrases in the presence of distractor words that were phonologically related or unrelated to the target and matched or mismatched in grammatical gender. In German, as in Dutch, the form of the definite determiner depends on the grammatical gender of the noun. For the speech onset latencies, the authors observed a gender interference effect and phonological facilitation. Importantly, the EEG analyses showed an earlier onset of the phonological effect compared with the gender congruency effect. These studies suggest that determiner selection may entail the activation of the word form of the noun. This conclusion is consistent with results of EEG studies by Strijkers and colleagues (Strijkers et al., 2010; Strijkers, Holcomb, & Costa, 2011) and an MEG study reported by Miozzo, Pulvermüller, and Hauk (2015), which showed early onsets (within 200 ms after picture onset) of word frequency effects in picture naming tasks. In sum, the noun frequency effect observed in Experiment 2 cannot be seen as conclusive evidence that lemma retrieval is frequency sensitive.

We proposed above that the likely reason why we saw only an adjective frequency effect in Experiments 1 and 2, and only a noun frequency effect in Experiment 3, was that the participants began to speak as soon as all information needed to initiate the first word of the utterance was available. For adjective-noun phrases, this was the stem of the adjective, and for determiner-adjective-noun phrases, it was the determiner, which could only be retrieved via the noun lemma. Whether the utterance format only affected when the participants initiated their utterances, or whether it also affected when they retrieved different words, remains to be determined. However, the latter view seems more plausible to us and is consistent with slot-and-insert models in the literature (e.g., Dell, Burger, & Svec, 1997), and with the view that in phrase production several concepts and the associated words can be activated simultaneously, albeit to different degrees (Jescheniak et al., 2003; Meyer, 1996). To outline the general idea, when participants see a display, the color and object concept and the associated words are activated in parallel. For adjective-noun phrases, they activate a two-slot structural frame specifying the order of the two-word categories. The structural frame initially sends additional activation to lemmas matching the selection requirements for the first slot. This is the case for the adjective, which is therefore selected first. Subsequently, additional activation is sent to units matching the selection requirements for the second slot. The noun lemma belongs to this category and is selected second. Upon lemma selection, the corresponding form representations and articulatory commands receive additional jolts of activation, so that the articulation of the adjective is initiated before the articulation of the noun.

When the definite determiner is included in the utterance, a slightly more complex scenario arises because the noun appears in the third position but its gender determines the form of the element in the first position of the frame. As a first approximation, the following scenario is plausible: At the conceptual level, color and object information are activated in parallel along with a representation of a concept of

definiteness, indicating that a definite determiner must be included in the utterance. Activation from the noun lemma and the definiteness concept converges on the lemma of one of the determiners, either *de* or *het*. A three-slot structural frame is activated, which specifies the serial positions of the word categories, and first sends additional activation to the determiner, then to the adjective, and, finally, to the noun. Consequently, the three lemmas are selected in that order and articulation can be initiated as soon as the form of the determiner has been retrieved. This scenario assumes that the selection of the determiner hinges on the retrieval of the noun lemma. Alternatively, in line with Starreveld and La Heij's (2004) proposal, the determiner form may receive activation from the form of the noun and the definiteness concept. The ordering of the words within the phrase would then occur at the form level. In both scenarios, the task, to express or not to express definiteness, is implemented in the number of units to retrieve (adjective and bare noun vs. adjective, noun, and determiner) and the choice of a matching structural frame that governs the serial order and timing of the selection of the components of the phrase.

We assume that the participants in our experiments began to speak as soon as sufficient information (the beginning of a color adjective or a determiner) was available. This yielded an adjective frequency effect in Experiment 1 and a noun frequency effect in Experiment 2. Our results are consistent with finding by Griffin and Spieler (2006), who asked participants to describe sets of objects in utterances such as "the clock and the television are above the needle" and found an effect of the frequency of the first noun but not of the second or third noun on the phrase onset latencies. By contrast, other studies found effects of the frequencies of words in noninitial positions: Alario et al. (2002) reported two picture description experiments showing additive effects of adjective and noun frequencies on the speech onset latencies for English determiner-adjective noun phrases. Janssen and Caramazza (2011) found a noun frequency effect for the onset latencies of English adjective-noun phrases elicited by presenting participants with words printed in different colors (e.g., the word *comb* printed in red to elicit the phrase *red comb*). Finally, Schnur (2011) asked participants to describe scenes in sentences such as "he opens the gate" and found an effect of the frequency of the postverbal noun on the speech onset latencies. In all of these studies, participants apparently used a broader planning span, including several words, than in the current study, such that effects of the frequencies of noninitial words on the utterance latencies could be seen.

Differences in planning span may be related to properties of the materials or design of the studies (e.g., Konopka & Meyer, 2014) or to speaker characteristics (e.g., Michel Lange & Laganaro, 2014; Schriefers & Teruel, 1999). Differences in the statistical analyses may also account for some of the differences in the observed results. For instance, Janssen and Barber (2012) pointed out that earlier studies did not control for effects of phrase frequency, so that some of the reported word frequency effects may actually be phrase frequency effects. In fact, if we do not include phrase frequency in the models of the log-RT for Experiments 1 or 2, we obtain additive effects of noun and adjective effects, as observed in the study by Alario and colleagues (2002). Modeling the log-RT of Experiment 3 in the same way only yields a noun frequency effect, perhaps because the retrieval of the adjective is faster than the retrieval of the determiner via the noun lemma. For the present purposes, it is important to note that all of the observed patterns of results are compatible with the view that speakers create utterances by combining individual words.

One may ask why no word frequency effects were observed in Janssen and Barber's (2012) study. The first content word in the Spanish and French phrases they investigated was a noun and, given the evidence presented here and in earlier studies, one would expect the noun frequencies to predict the speech onset latencies. One possibility is insufficient power to detect a small effect. We used the same line drawings and colors as Janssen and Barber, but we created 460 color-object combinations for Experiment 1 and 468 for Experiment 2, whereas Janssen and Barber used 100 items in their Experiment 1 and 112 items in Experiment 2. In an unpublished experiment using a stimulus set of similar size, we did not find any significant frequency effects. If these designs were underpowered with respect to both word and phrase frequency effects, which (if any) of the effects are observed would be determined by chance. Moreover, the word and phrase frequencies were correlated in Janssen and Barber's study, as in ours, and the statistical treatment of the data chosen by Janssen and Barber may have worked against discovering both word and phrase frequency. An important task for further phrase production research is to assess the effects of word and phrase frequencies in materials when these frequencies are uncorrelated, as was, for instance, the case in a memory study by Jacobs et al. (2016).

Accounting for Phrase Frequency Effects

In the present study, as in Janssen and Barber's (2012), efforts were made, both in designing the materials and in analyzing the data, to control for visual and conceptual influences on the speech onset latencies, and it is likely that the phrase frequency effect has a linguistic basis. Existing accounts of frequency effects for multiword sequences focus on the storage of word sequences or on advantages arising during the encoding or decoding processes of sequences of words. Though our data do not allow us to discriminate between these views, we briefly discuss each option (for further discussions, see also Jacobs et al., 2016; Siyanova-Chanturia & Martinez, 2015).

Storage accounts entail that the mental lexicon encodes not only individual words but also representations of word sequences. One view is that only very high frequent word sequences, such as *don't know*, are stored (Goldberg, 2006; Tremblay et al., 2011); another view is that less frequent word sequences have stored representations as well (Dąbrowska, 2014; Ramscar, Yarlett, Dye, Denny, & Thorpe, 2010; for discussion, see also Arnon & Snider, 2010; Jacobs et al., 2016). As the phrases used in the present study were generally low in frequency, compared with phrases such as *don't know*, only the latter view applies. The storage view entails that each time a phrase is encountered, a memory trace of this word combination is created and maintained in memory. Models differ in their views of the organization of these trace, whether they are, for instance, sets of exemplars, or whether there are in addition more abstract representations capturing commonalities across instances (e.g., Dąbrowska, 2014; Ramscar et al., 2010).

Stored representations of phrases could exist at one or several levels of representation, in analogy to proposals made for the representation of compounds. For instance, Levelt et al. (1999) proposed that frequent compounds are represented as single units at the lemma level and in terms of individual morphemes at word-form level (see also Bien, Levelt, & Baayen, 2005; Jacobs & Dell, 2014). Other authors have proposed stored representations of compounds at the level of word forms (Caramazza, 1997; Janssen, Bi, & Caramazza, 2008).

How well these views apply to the representation of phrases remains to be seen. Given the evidence that word frequency effects have multiple origins, we would expect the same to hold for phrase frequency effects. Representation of phrases at several levels offers a natural (though not particularly parsimonious) account for the fact that phrase frequency appears to affect a wide range of processes, including both utterance planning and articulation (e.g., Arnon & Cohen Priva, 2014; Aylett & Turk, 2006; Bannard & Matthews, 2008). Furthermore, given that we find both phrase and word frequency effects, it seems most likely that any stored unitary representations of phrases coexist with representations of phrases in terms of their constituent words. For instance, when a speaker plans the phrase *brown shoe*, a stored "brown_shoe" representation may be activated along with the word representations *brown* and *shoe*. The speed of initiating the phrase may then depend on the combined effects of the activation of these representations, with the articulatory commands required to produce the first syllable of the utterance receiving activation from both the "brown shoe" unit and the "brown" unit (see also Bell et al., 2009; Bybee, 2002, 2006; Hendrix, Bolger, & Baayen, 2017; Jacobs et al., 2016).

Other accounts of frequency effects for multiword sequences highlight an obvious shortcoming of the storage position, namely, the size and complexity of the lexicon (e.g., Baayen, Hendrix, & Ramscar, 2013). They focus on computational advantages arising during the combination of words (e.g., Tremblay et al., 2011). Frequent phrases may be faster to plan than less frequent ones because the selection of the specific combination of words or their ordering becomes easier with practice. Which mechanisms might underlie such practice effects remains to be seen. One option is that lemma selection is facilitated by word-to-word associations within frequent phrases. For instance, the lemma *shoe* may be faster to select together with the lemma *brown* than with the lemma *green* because of stronger mutual associations between *brown* and *shoe* than between *green* and *shoe*. We cannot exclude that associations contributed to the phrase frequency effects in our study, though the associations between the adjectives and nouns were certainly not strong. The association norms for Dutch provided by De Deyne and Storms (2008) include all adjectives and all but one of the nouns we used. The association strength from all adjectives to all nouns was zero, that is, none of the nouns were given as a response to any of the adjectives. Backwards associations, with the color term given in response to the noun, were listed for three items, with association strengths below 10%. Of course, in evaluating these findings, one needs to keep in mind that providing verbal associations is a metalinguistic task, and that it may not be sensitive to small differences in the association strengths between words.

Another way in which phrase frequency may affect processing is through changes in the "vertical" links between units, that is between the conceptual representation of a phrase and the corresponding linguistic units. For instance, in the spirit of naive discrimination theory (Baayen et al., 2013; Hendrix et al., 2017), one might speculate that the word *brown* and its phonemes are often experienced together with the concept *brown shoe*, so that the word and phonemes are tightly linked to the concept, more so than the word *green* and its phonemes are linked to the concept *green shoe*, and are therefore faster to retrieve. To reiterate, the mechanisms through which experience with a specific phrase facilitates generating the phrase at a later occasion remain to be uncovered.

In this discussion, as in much of the literature, frequency effects are directly linked to psychological processes occurring during the comprehension or production of speech. It is assumed that the frequency of a word or phrase in a corpus of text corresponds to the frequency of usage by a group of speakers and that frequent use facilitates comprehension or production. However, it has long been known that depending on one's sample, word frequency can be strongly correlated with other lexical variables (e.g., length, neighborhood density, or syllable structure), age of acquisition, and properties of the concepts referred to (e.g., their imageability or concreteness). The same is likely to be true for phrases; for instance, Arnon, McCauley, and Christiansen (2017) recently discussed age of acquisition effects for phrases. Though efforts were made here and in Janssen and Barber's (2012) study to separate phrase frequency from other influences, it is still possible that the observed phrase frequency effects were WYeeV93+ caused by combinations of other variables that were not captured. In short, although phrase frequency effects appear to be quite real, their bases remain to be determined. Doing so is important both for theories of the mental lexicon and theories of phrase and sentence generation.

Conclusions

In sum, we replicate the phrase frequency effects on utterance onset latencies demonstrated by Janssen and Barber (2012), but we demonstrate that phrase frequency effects co-exist with word frequency effects. Thus, speakers are affected by their experience with specific words as well as their experience with specific word sequences. For models of lexical knowledge, this implies that there is knowledge of phrases and of their components. In addition, we showed that a word frequency effect is obtained for the word that needs to be accessed first in the utterance. Depending on the structure of the utterance, this can be the first content word or a downstream word that governs the form of a determiner in utterance-initial position. Whether a frequency effect is observed only for this word or also for the following words probably depends on the speakers' choice of planning strategy.

References

Alario, F. X., Costa, A., & Caramazza, A. (2002). Frequency effects in noun phrase production: Implications for models of lexical access. *Language and Cognitive Processes, 17,* 299–319. http://dx.doi.org/10.1080/01690960143000236

Almeida, J., Knobel, M., Finkbeiner, M., & Caramazza, A. (2007). The locus of the frequency effect in picture naming: When recognizing is not enough. *Psychonomic Bulletin & Review, 14,* 1177–1182. http://dx.doi.org/10.3758/BF03193109

Antón-Méndez, I., Schütze, C. T., Champion, M. K., & Gollan, T. H. (2012). What the tip-of-the-tongue (TOT) says about homophone frequency inheritance. *Memory & Cognition, 40,* 802–811. http://dx.doi.org/10.3758/s13421-012-0189-1

Arnon, I., & Cohen Priva, U. (2013). More than words: The effect of multi-word frequency and constituency on phonetic duration. *Language and Speech, 56,* 349–371. http://dx.doi.org/10.1177/0023830913484891

Arnon, I., & Cohen Priva, U. (2014). Time and again: The changing effect of word and multiword frequency on phonetic duration. *The Mental Lexicon, 9,* 377–400.

Arnon, I., McCauley, S. M., & Christiansen, M. H. (2017). Digging up the building blocks of language: Age-of-acquisition effects for multiword phrases. *Journal of Memory and Language, 92,* 265–280. http://dx.doi.org/10.1016/j.jml.2016.07.004

Arnon, I., & Snider, N. (2010). More than words: Frequency effects for multi-word phrases. *Journal of Memory and Language, 62,* 67–82. http://dx.doi.org/10.1016/j.jml.2009.09.005

Aylett, M., & Turk, A. (2006). Language redundancy predicts syllabic duration and the spectral characteristics of vocalic syllable nuclei. *The Journal of the Acoustical Society of America, 119,* 3048–3058. http://dx.doi.org/10.1121/1.2188331

Baayen, R. H. (2013). languageR: Data sets and functions with "Analyzing linguistic data: A practical introduction to statistics." *R package on CRAN.* Retrieved from https://CRAN.R-project.org/package=languageR

Baayen, R. H., Hendrix, P., & Ramscar, M. (2013). Sidestepping the combinatorial explosion: An explanation of n-gram frequency effects based on naive discriminative learning. *Language and Speech, 56,* 329–347. http://dx.doi.org/10.1177/0023830913484896

Balota, D. A., & Chumbley, J. I. (1984). Are lexical decisions a good measure of lexical access? The role of word frequency in the neglected decision stage. *Journal of Experimental Psychology: Human Perception and Performance, 10,* 340–357. http://dx.doi.org/10.1037/0096-1523.10.3.340

Balota, D. A., & Chumbley, J. I. (1990). Where are the effects of frequency in visual word recognition tasks? Right where we said they were! Comment on Monsell, Doyle, and Haggard (1989). *Journal of Experimental Psychology: General, 119,* 231–237. http://dx.doi.org/10.1037/0096-3445.119.2.231

Balota, D. A., Pollatsek, A., & Rayner, K. (1985). The interaction of contextual constraints and parafoveal visual information in reading. *Cognitive Psychology, 17,* 364–390. http://dx.doi.org/10.1016/0010-0285(85)90013-1

Bannard, C., & Matthews, D. (2008). Stored word sequences in language learning: The effect of familiarity on children's repetition of four-word combinations. *Psychological Science, 19,* 241–248. http://dx.doi.org/10.1111/j.1467-9280.2008.02075.x

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2014). lme4: Linear mixed-effects models using Eigen and S4. *R package version, 1*(7), 1–23. Retrieved from http://CRAN.R-project.org/package=lme4

Bates, E., D'Amico, S., Jacobsen, T., Székely, A., Andonova, E., Devescovi, A., . . . Tzeng, O. (2003). Timed picture naming in seven languages. *Psychonomic Bulletin & Review, 10,* 344–380. http://dx.doi.org/10.3758/BF03196494

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. http://dx.doi.org/10.1016/j.cognition.2004.11.006

Belke, E., & Meyer, A. S. (2007). Single and multiple object naming in healthy ageing. *Language and Cognitive Processes, 22,* 1178–1211. http://dx.doi.org/10.1080/01690960701461541

Bell, A., Brenier, J. M., Gregory, M., Girand, C., & Jurafsky, D. (2009). Predictability effects on durations of content and function words in conversational English. *Journal of Memory and Language, 60,* 92–111. http://dx.doi.org/10.1016/j.jml.2008.06.003

Bien, H., Levelt, W. J. M., & Baayen, R. H. (2005). Frequency effects in compound production. *Proceedings of the National Academy of Sciences of the United States of America, 102,* 17876–17881. http://dx.doi.org/10.1073/pnas.0508431102

Boersma, P. (2001). Praat, a system for doing phonetics by computer. *Glot International, 5,* 341–345.

Broadbent, D. E. (1967). Word-frequency effect and response bias. *Psychological Review, 74,* 1–15. http://dx.doi.org/10.1037/h0024206

Brysbaert, M., Keuleers, E., & New, B. (2011). Assessing the usefulness of google books' word frequencies for psycholinguistic research on word processing. *Frontiers in Psychology, 2,* 27. http://dx.doi.org/10.3389/fpsyg.2011.00027

Bürki, A., Sadat, J., Dubarry, A. S., & Alario, F. X. (2016). Sequential processing during noun phrase production. *Cognition, 146,* 90–99. http://dx.doi.org/10.1016/j.cognition.2015.09.002

Bybee, J. (2002). Word frequency and context of use in the lexical diffusion of phonetically conditioned sound change. *Language Variation and Change, 14,* 261–290. http://dx.doi.org/10.1017/S0954394502143018

Bybee, J. (2006). From usage to grammar: The mind's response to repetition. *Language, 82,* 711–733. http://dx.doi.org/10.1353/lan.2006.0186

Bybee, J., & Scheibman, J. (1999). The effect of usage on degrees of constituency: The reduction of don't in English. *Linguistics: An Interdisciplinary Journal of the Language Sciences, 37,* 575–596. http://dx.doi.org/10.1515/ling.37.4.575

Caramazza, A. (1997). How many levels of processing are there in lexical access? *Cognitive Neuropsychology, 14,* 177–208. http://dx.doi.org/10.1080/026432997381664

Caramazza, A., Costa, A., Miozzo, M., & Bi, Y. (2001). The specific-word frequency effect: Implications for the representation of homophones in speech production. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 27,* 1430–1450. http://dx.doi.org/10.1037/0278-7393.27.6.1430

Cleland, A. A., Gaskell, M. G., Quinlan, P. T., & Tamminen, J. (2006). Frequency effects in spoken and visual word recognition: Evidence from dual-task methodologies. *Journal of Experimental Psychology: Human Perception and Performance, 32,* 104–119. http://dx.doi.org/10.1037/0096-1523.32.1.104

Conklin, K., & Schmitt, N. (2008). Formulaic sequences: Are they processed more quickly than nonformulaic language by native and nonnative speakers? *Applied Linguistics, 29,* 72–89. http://dx.doi.org/10.1093/applin/amm022

Connine, C. M., Mullennix, J., Shernoff, E., & Yelen, J. (1990). Word familiarity and frequency in visual and auditory word recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 16,* 1084–1096. http://dx.doi.org/10.1037/0278-7393.16.6.1084

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. http://dx.doi.org/10.1006/jmla.2001.2804

Costa, A., Sebastian-Galles, N., Miozzo, M., & Caramazza, A. (1999). The gender congruency effect: Evidence from Spanish and Catalan. *Language and Cognitive Processes, 14,* 381–391. http://dx.doi.org/10.1080/016909699386275

Cuetos, F., Bonin, P., Alameda, J. R., & Caramazza, A. (2010). The specific-word frequency effect in speech production: Evidence from Spanish and French. *The Quarterly Journal of Experimental Psychology, 63,* 750–771. http://dx.doi.org/10.1080/17470210903121663

Dąbrowska, E. (2014). Recycling utterances: A speaker's guide to sentence processing. *Cognitive Linguistics, 25,* 617–653. http://dx.doi.org/10.1515/cog-2014-0057

Dahan, D., Magnuson, J. S., & Tanenhaus, M. K. (2001). Time course of frequency effects in spoken-word recognition: Evidence from eye movements. *Cognitive Psychology, 42,* 317–367. http://dx.doi.org/10.1006/cogp.2001.0750

Damian, M. F. E. (2003). Articulatory duration in single-word speech production. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 29,* 416–431. http://dx.doi.org/10.1037/0278-7393.29.3.416

Damian, M. F., & Dumay, N. (2007). Time pressure and phonological advance planning in spoken production. *Journal of Memory and Language, 57,* 195–209. http://dx.doi.org/10.1016/j.jml.2006.11.001

Damian, M. F., & Dumay, N. (2009). Exploring phonological encoding through repeated segments. *Language and Cognitive Processes, 24,* 685–712. http://dx.doi.org/10.1080/01690960802351260

De Deyne, S., & Storms, G. (2008). Word associations: Norms for 1,424 Dutch words in a continuous task. *Behavior Research Methods, 40,* 198–205. http://dx.doi.org/10.3758/BRM.40.1.198

Dell, G. S. (1986). A spreading-activation theory of retrieval in sentence production. *Psychological Review, 93,* 283–321. http://dx.doi.org/10.1037/0033-295X.93.3.283

Dell, G. S. (1990). Effects of frequency and vocabulary type on phonological speech errors. *Language and Cognitive Processes, 5,* 313–349. http://dx.doi.org/10.1080/01690969008407066

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. http://dx.doi.org/10.1037/0033-295X.104.1.123

Dhooge, E., De Baene, W., & Hartsuiker, R. J. (2016). The mechanisms of determiner selection and its relation to lexical selection: An ERP study. *Journal of Memory and Language, 88,* 28–38. http://dx.doi.org/10.1016/j.jml.2015.12.004

Diependaele, K., Lemhöfer, K., & Brysbaert, M. (2013). The word frequency effect in first- and second-language word recognition: A lexical entrenchment account. *The Quarterly Journal of Experimental Psychology, 66,* 843–863. http://dx.doi.org/10.1080/17470218.2012.720994

Engbert, R., Nuthmann, A., Richter, E. M., & Kliegl, R. (2005). SWIFT: A dynamical model of saccade generation during reading. *Psychological Review, 112,* 777–813. http://dx.doi.org/10.1037/0033-295X.112.4.777

Field, A. (2009). *Discovering statistics using SPSS.* Thousand Oaks, CA: Sage publication.

Goldberg, A. (2006). *Constructions at work: The nature of generalization in language.* Oxford, UK: Oxford University Press.

Griffin, Z. M. (2001). Gaze durations during speech reflect word selection and phonological encoding. *Cognition, 82,* B1–B14. http://dx.doi.org/10.1016/S0010-0277(01)00138-X

Griffin, Z. M., & Bock, K. (1998). Constraint, word frequency, and the relationship between lexical processing levels in spoken word production. *Journal of Memory and Language, 38,* 313–338. http://dx.doi.org/10.1006/jmla.1997.2547

Griffin, Z. M., & Spieler, D. H. (2006). Observing the what and when of language production for different age groups by monitoring speakers' eye movements. *Brain and Language, 99,* 272–288. http://dx.doi.org/10.1016/j.bandl.2005.08.003

Hendrix, P., Bolger, P., & Baayen, H. (2017). Distinct ERP signatures of word frequency, phrase frequency, and prototypicality in speech production. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 43,* 128–149. http://dx.doi.org/10.1037/a0040332

Huettig, F. (2015). Four central questions about prediction in language processing. *Brain Research, 1626,* 118–135. http://dx.doi.org/10.1016/j.brainres.2015.02.014

Jackendoff, R. (2002). *Foundations of language.* Oxford, UK: Oxford University Press. http://dx.doi.org/10.1093/acprof:oso/9780198270126.001.0001

Jackendoff, R. (2011). What is the human language faculty? Two views. *Language, 87,* 586–624. http://dx.doi.org/10.1353/lan.2011.0063

Jacobs, C. L., & Dell, G. S. (2014). "hotdog," not "hot" "dog": The phonological planning of compound words. *Language, Cognition and Neuroscience, 29,* 512–523. http://dx.doi.org/10.1080/23273798.2014.892144

Jacobs, C. L., Dell, G. S., Benjamin, A. S., & Bannard, C. (2016). Part and whole linguistic experience affect recognition memory for multiword sequences. *Journal of Memory and Language, 87,* 38–58. http://dx.doi.org/10.1016/j.jml.2015.11.001

Janssen, N., & Barber, H. A. (2012). Phrase frequency effects in language production. *PLoS ONE, 7*(3), e33202. http://dx.doi.org/10.1371/journal.pone.0033202

Janssen, N., Bi, Y., & Caramazza, A. (2008). A tale of two frequencies: Determining the speed of lexical access for Mandarin Chinese and English compounds. *Language and Cognitive Processes, 23,* 1191–1223. http://dx.doi.org/10.1080/01690960802250900

Janssen, N., & Caramazza, A. (2011). Lexical selection in multi-word production. *Frontiers in Psychology, 2,* 81. http://dx.doi.org/10.3389/fpsyg.2011.00081

Jescheniak, J. D., & Levelt, W. J. M. (1994). Word frequency effects in speech production: Retrieval of syntactic information and phonological form. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 20,* 824–843. http://dx.doi.org/10.1037/0278-7393.20.4.824

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. http://dx.doi.org/10.1037/0096-1523.29.2.441

Jongman, S. R., Roelofs, A., & Meyer, A. S. (2015). Sustained attention in language production: An individual differences investigation. *The Quarterly Journal of Experimental Psychology, 68,* 710–730. http://dx.doi.org/10.1080/17470218.2014.964736

Kawamoto, A. H., Liu, Q., Lee, R. J., & Grebe, P. R. (2014). The segment as the minimal planning unit in speech production: Evidence based on absolute response latencies. *The Quarterly Journal of Experimental Psychology, 67,* 2340–2359. http://dx.doi.org/10.1080/17470218.2014.927892

Kittredge, A. K., Dell, G. S., Verkuilen, J., & Schwartz, M. F. (2008). Where is the effect of frequency in word production? Insights from aphasic picture-naming errors. *Cognitive Neuropsychology, 25,* 463–492. http://dx.doi.org/10.1080/02643290701674851

Konopka, A. E., & Meyer, A. S. (2014). Priming sentence planning. *Cognitive Psychology, 73,* 1–40. http://dx.doi.org/10.1016/j.cogpsych.2014.04.001

Korvorst, M., Roelofs, A., & Levelt, W. J. M. (2006). Incrementality in naming and reading complex numerals: Evidence from eyetracking. *The Quarterly Journal of Experimental Psychology: Human Experimental Psychology, 59,* 296–311. http://dx.doi.org/10.1080/17470210500151691

Kuperberg, G. R., & Jaeger, T. F. (2016). What do we mean by prediction in language comprehension? *Language, Cognition and Neuroscience, 31,* 32–59. http://dx.doi.org/10.1080/23273798.2015.1102299

Kuperman, V., & Van Dyke, J. A. (2013). Reassessing word frequency as a determinant of word recognition for skilled and unskilled readers. *Journal of Experimental Psychology: Human Perception and Performance, 39,* 802–823. http://dx.doi.org/10.1037/a0030859

Levelt, W. J. M. (1989). *Speaking: From intention to articulation.* Cambridge, MA: MIT Press.

Levelt, W. J. M., Roelofs, A., & Meyer, A. S. (1999). A theory of lexical access in speech production. *Behavioral and Brain Sciences, 22,* 1–38. http://dx.doi.org/10.1017/S0140525X99001776

Matthews, D., & Bannard, C. (2010). Children's production of unfamiliar word sequences is predicted by positional variability and latent classes in a large sample of child-directed speech. *Cognitive Science, 34,* 465–488. http://dx.doi.org/10.1111/j.1551-6709.2009.01091.x

McDonald, S. A., & Shillcock, R. C. (2003). Eye movements reveal the on-line computation of lexical probabilities during reading. *Psychological Science, 14,* 648–652. http://dx.doi.org/10.1046/j.0956-7976.2003.psci_1480.x

Mehta, R., & Zhu, R. J. (2009). Blue or red? Exploring the effect of color on cognitive task performances. *Science, 323,* 1226–1229. http://dx.doi.org/10.1126/science.1169144

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. http://dx.doi.org/10.1006/jmla.1996.0026

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. http://dx.doi.org/10.1016/S0749-596X(02)00509-0

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. http://dx.doi.org/10.1016/S0010-0277(98)00009-2

Meyer, A. S., Wheeldon, L., van der Meulen, F., & Konopka, A. (2012). Effects of speech rate and practice on the allocation of visual attention in multiple object naming. *Frontiers in Psychology, 3,* 39. http://dx.doi.org/10.3389/fpsyg.2012.00039

Michel Lange, V., & Laganaro, M. (2014). Inter-subject variability modulates phonological advance planning in the production of adjective-noun phrases. *Frontiers in Psychology, 5,* 43. http://dx.doi.org/10.3389/fpsyg.2014.00043

Miozzo, M., Pulvermüller, F., & Hauk, O. (2015). Early parallel activation of semantics and phonology in picture naming: Evidence from a multiple linear regression MEG study. *Cerebral Cortex, 25,* 3343–3355. http://dx.doi.org/10.1093/cercor/bhu137

Moors, A., De Houwer, J., Hermans, D., Wanmaker, S., van Schie, K., Van Harmelen, A. L., . . . Brysbaert, M. (2013). Norms of valence, arousal, dominance, and age of acquisition for 4,300 Dutch words. *Behavior Research Methods, 45,* 169–177. http://dx.doi.org/10.3758/s13428-012-0243-8

Munson, B. (2007). Lexical access, lexical representation, and vowel production. *Laboratory Phonology, 9,* 201–228.

Navarrete, E., Basagni, B., Alario, F.-X., & Costa, A. (2006). Does word frequency affect lexical selection in speech production? *The Quarterly Journal of Experimental Psychology, 59,* 1681–1690. http://dx.doi.org/10.1080/17470210600750558

Oldfield, R. C., & Wingfield, A. (1965). Response latencies in naming objects. *The Quarterly Journal of Experimental Psychology, 17,* 273–281. http://dx.doi.org/10.1080/17470216508416445

Partee, B. H. (1995). Lexical semantics and compositionality. In L. R. Gleitman & M. Liberman (Eds.), *Invitation to cognitive science* (pp. 311–360). Cambridge, MA: MIT Press.

Piai, V., Roelofs, A., & Maris, E. (2014). Oscillatory brain responses in spoken word production reflect lexical frequency and sentential constraint. *Neuropsychologia, 53,* 146–156. http://dx.doi.org/10.1016/j.neuropsychologia.2013.11.014

Pinker, S. (1998). *How the mind works*. London, UK: Penguin Press.

Plaut, D. C., & Kello, C. T. (1999). The emergence of phonology from the interplay of speech comprehension and production: A distributed connectionist approach. In B. MacWhinney (Ed.), *The emergence of language* (pp. 381–415). Mahwah, NJ: Erlbaum.

Pluymaekers, M., Ernestus, M., & Baayen, R. H. (2005). Articulatory planning is continuous and sensitive to informational redundancy. *Phonetica, 62,* 146–159. http://dx.doi.org/10.1159/000090095

Ramscar, M., Yarlett, D., Dye, M., Denny, K., & Thorpe, K. (2010). The effects of feature-label-order and their implications for symbolic learning. *Cognitive Science, 34,* 909–957. http://dx.doi.org/10.1111/j.1551-6709.2009.01092.x

Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin, 124,* 372–422. http://dx.doi.org/10.1037/0033-2909.124.3.372

R Core Team. (2015). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from https://www.R-project.org/

Reichle, E. D., Pollatsek, A., Fisher, D. L., & Rayner, K. (1998). Toward a model of eye movement control in reading. *Psychological Review, 105,* 125–157. http://dx.doi.org/10.1037/0033-295X.105.1.125

Roelofs, A. (1992). A spreading-activation theory of lemma retrieval in speaking. *Cognition, 42,* 107–142. http://dx.doi.org/10.1016/0010-0277(92)90041-F

Roelofs, A. (2008). Attention, gaze shifting, and dual-task interference from phonological encoding in spoken word planning. *Journal of Experimental Psychology: Human Perception and Performance, 34,* 1580–1598. http://dx.doi.org/10.1037/a0012476

Schnur, T. T. (2011). Phonological planning during sentence production: Beyond the verb. *Frontiers in Psychology, 2,* 319. http://dx.doi.org/10.3389/fpsyg.2011.00319

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. http://dx.doi.org/10.1037/0278-7393.25.3.702

Schriefers, H., Hantsch, A., & Jescheniak, J. D. (2002). Determiner selection in noun phrase production. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 28,* 941–950. http://dx.doi.org/10.1037/0278-7393.28.5.941

Schriefers, H., & Teruel, E. (1999). Phonological facilitation in the production of two-word utterances. *European Journal of Cognitive Psychology, 11,* 17–50. http://dx.doi.org/10.1080/713752301

Severens, E., Van Lommel, S., Ratinckx, E., & Hartsuiker, R. J. (2005). Timed picture naming norms for 590 pictures in Dutch. *Acta Psychologica, 119,* 159–187. http://dx.doi.org/10.1016/j.actpsy.2005.01.002

Shi, L. F. (2015). English word frequency and recognition in bilinguals: Inter-corpus comparison and error analysis. *International Journal of Audiology, 54,* 674–681. http://dx.doi.org/10.3109/14992027.2015.1030509

Siyanova-Chanturia, A., Conklin, K., & van Heuven, W. J. (2011). Seeing a phrase "time and again" matters: The role of phrasal frequency in the processing of multiword sequences. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 37,* 776–784. http://dx.doi.org/10.1037/a0022531

Siyanova-Chanturia, A., & Martinez, R. (2015). The idiom principle revisited. *Applied Linguistics, 36,* 549–569.

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. http://dx.doi.org/10.1037/0278-7393.6.2.174

Stadthagen-Gonzalez, H., Damian, M. F., Pérez, M. A., Bowers, J. S., & Marín, J. (2009). Name-picture verification as a control measure for object naming: A task analysis and norms for a large set of pictures. *The Quarterly Journal of Experimental Psychology, 62,* 1581–1597. http://dx.doi.org/10.1080/17470210802511139

Starreveld, P., & La Heij, W. (2004). Phonological facilitation of grammatical gender retrieval. *Language and Cognitive Processes, 19,* 677–711. http://dx.doi.org/10.1080/01690960444000061

Strijkers, K., Costa, A., & Thierry, G. (2010). Tracking lexical access in speech production: Electrophysiological correlates of word frequency and cognate effects. *Cerebral Cortex, 20,* 912–928. http://dx.doi.org/10.1093/cercor/bhp153

Strijkers, K., Holcomb, P. J., & Costa, A. (2011). Conscious intention to speak proactively facilitates lexical access during overt object naming. *Journal of Memory and Language, 65,* 345–362. http://dx.doi.org/10.1016/j.jml.2011.06.002

Tremblay, A., & Baayen, R. H. (2010). Holistic processing of regular four-word sequences: A behavioral and ERP study of the effects of structure, frequency, and probability on immediate free recall. In D. Wood (Ed.), *Perspectives on formulaic language, acquisition and communication* (pp. 151–173). London, UK: Continuum International.

Tremblay, A., Derwing, B., Libben, G., & Westbury, G. (2011). Processing advantages of lexical bundles: Evidence from self-paced reading and sentence recall tasks. *Language Learning, 61,* 569–613. http://dx.doi.org/10.1111/j.1467-9922.2010.00622.x

Troia, G. A., Roth, F. P., & Yeni-Komshian, G. H. (1996). Word frequency and age effects in normally developing children's phonological processing. *Journal of Speech & Hearing Research, 39,* 1099–1108. http://dx.doi.org/10.1044/jshr.3905.1099

Underwood, G., Schmitt, N., & Galpin, A. (2004). The eyes have it: An

eye-movement study into the processing of formulaic sequences. In N. Schmitt (Ed.), *Formulaic sequences: Acquisition, processing, and use* (pp. 153–172). Amsterdam, the Netherlands: John Benjamins Press. http://dx.doi.org/10.1075/lllt.9.09und

Valsecchi, M., Künstler, V., Saage, S., White, B. J., Mukherjee, J., & Gegenfurtner, K. R. (2013). Advantage in reading lexical bundles is reduced in non-native speakers. *Journal of Eye Movement Research, 6,* 1–16.

Appendix A

Names of Objects and Colors Presented in Experiment 1–3

Object Names With Neuter Gender

anker (anchor), bed (bed), blad (leaf), boek (book), bord (plate), bot (bone), glas (glass), hart (heart), horloge (watch), huis (house), oog (eye), potlood (pencil), raam (window), slot (lock), stoplicht (stoplight), stuur (steering wheel)

Object Names With Non-Neuter Gender

bel (bell), bezem (broom), borstel (brush), dokter (doctor), doos (box), fles (bottle), fontein (fountain), giraffe (giraffe), harp (harp), helicopter (helicopter), hond (dog), kangoeroe (kangaroo), ketting (chain), lepel (spoon), munt (coin), paperclip (paperclip), piano (piano), pijl (arrow), rugzak (backpack), sandwich (sandwich), schoen (shoe), sjaal (scarf), sok (sock), spiegel (mirror), spijker (nail), stoel (chair), taart (cake), telefoon (telephone), tol (spinning top), trommel (drum), voet (foot), vork (fork), wereldbol (globe), zaag (saw)

Color Names

blauw (blue), bruin (brown), geel (yellow), grijs (gray), groen (green), oranje (orange), paars (purple), rood (red), roze (pink), zwart (black)

Appendix B

AIC of the Mixed Effect Models

Table B1
AIC of the Models for Speech Error Rates, Speech Onset Latencies, Gaze Shift Latencies, and Manual Response Latencies in Experiment 1

Model	Error rate	Speech onset latencies	Gaze shift latencies	Manual response latencies
Baseline	4,582	−20,491	−19,558	−17,728
Log-adjective frequency	4,579	−20,655	−19,590	−17,755
Log-phrase frequency	4,587	−20,658	−19,591	−17,769
Log-noun frequency	4,589	−20,655	−19,577	−17,795

Note. AIC = Akaike information criterion.

(*Appendices continue*)

Table B2
AIC of the Models for Speech Error Rates and Speech Onset Latencies in Experiment 2

Model	Error rate	Speech onset latencies
Baseline	5,592	−21,089
Log-adjective frequency	5,578	−21,126
Log-phrase frequency	5,580	−21,134
Log-noun frequency	5,584	−20,915

Note. AIC = Akaike information criterion.

Table B3
AIC of the Models for Speech Error Rates and Speech Onset Latencies in Experiment 3

Model	Error rate	Speech onset latencies
Baseline	6,102	−25,238
Log-adjective frequency	6,105	−25,257
Log-phrase frequency	6,118	−25,255
Log-noun frequency	6,121	−25,246

Note. AIC = Akaike information criterion.

Appendix C

Results of Mixed-Effects Models for Error Rates in Experiment 1, 2, and 3

Experiment	Fixed effects	Beta	SE	z value
1	Intercept	−2.88	.153	−18.794***
	Trial	−.259	.072	−3.612***
	Log-RT color verification	.201	.038	5.302***
	Familiarity rating	−.013	.015	−.912
	Log-adjective frequency	−.047	.060	−.787
	Log-phrase frequency	−.012	.053	−.233
	Log-noun frequency	−.046	.045	−1.035
2	Intercept	−2.997	.113	−26.441***
	Trial	−.189	.063	−2.985**
	Log-RT of color verification	.223	.05	4.497***
	Log-adjective frequency	.043	.072	.593
	Log-phrase frequency	.021	.065	.327
	Log-noun frequency	−.104	.066	−1.56
3	Intercept	−2.257	.137	−16.460***
	Determiner	.135	.058	2.351*
	Trial	−.094	.06	−1.574
	Familiarity rating	−.029	.018	−1.567
	Log-noun frequency	−.107	.061	−1.768
	Log-phrase frequency	−.055	.054	−1.010
	Log-adjective frequency	.072	.062	1.154

Note. SE = standard error; RT = reaction time.
* $p < .01$.   ** $p < .001$.   *** $p < .00001$.

Appendix D

Results of Mixed-Effects Models for Speech Onset Latency, Gaze Shift Latency, and Manual Response Latency in Experiment 1

Dependent variable	Fixed effects	Beta	SE	t value
Speech onset latency	Intercept	3.020	.014	211.37
	Trial	−.013	.002	−5.30
	Log-RT of color verification	.002	.001	1.52
	Log-adjective frequency	−.017	.002	−8.08
	Log-phrase frequency	−.005	.002	−3.26
	Log-noun frequency	.000	.002	.08
Gaze shift latency	Intercept	3.212	.016	199.54
	Trial	−.009	.004	−2.33
	Log-RT of color verification	.005	.001	4.14
	Familiarity rating	−.002	.001	−3.29
	Length in phonemes	.003	.001	3.96
	Log-adjective frequency	−.006	.002	−3.81
	Log-phrase frequency	−.003	.002	−2.28
	Log-noun frequency	−.002	.002	−1.00
Manual response latency	Intercept	3.270	.018	185.99
	Trial	−.015	.005	−3.3
	Log-RT of color verification	.006	.001	3.81
	Familiarity rating	−.001	.000	−2.82
	Log-adjective frequency	−.005	.002	−2.73
	Log-phrase frequency	−.005	.002	−2.94
	Log-noun frequency	−.007	.002	−3.93

Note. SE = standard error; RT = reaction time.

Appendix E

Results of Mixed-Effects Models for Speech Onset Latency in Experiment 2

Fixed effects	Beta	SE	t value
Intercept	2.964	.011	268.5
Trial	−.008	.003	−2.32
Log-RT of color verification	.009	.002	4.82
Log-adjective frequency	−.007	.002	−3.2
Log-phrase frequency	−.006	.002	−2.82
Log-noun frequency	−.003	.002	−1.51

Note. SE = standard error; RT = reaction time.

Appendix F

Results of Mixed-Effects Models for Speech Onset Latency in Experiment 3

Fixed effects	Beta	SE	t value
Intercept	2.960	.010	285.55
Trial	−.008	.003	−2.49
Log-RT of color verification	.009	.002	5.08
Log-adjective frequency	−.007	.002	−3.44
Log-phrase frequency	−.005	.002	−2.61
Log-noun frequency	−.002	.002	−1.23

Note. SE = standard error; RT = reaction time.