

Word Frequency Effects in Speech Production: Retrieval of Syntactic Information and of Phonological Form

Jörg D. Jescheniak and Willem J. M. Levelt

In 7 experiments the authors investigated the locus of word frequency effects in speech production. Experiment 1 demonstrated a frequency effect in picture naming that was robust over repetitions. Experiments 2, 3, and 7 excluded contributions from object identification and initiation of articulation. Experiments 4 and 5 investigated whether the effect arises in accessing the syntactic word (*lemma*) by using a grammatical gender decision task. Although a frequency effect was found, it dissipated under repeated access to a word's gender. Experiment 6 tested whether the robust frequency effect arises in accessing the phonological form (*lexeme*) by having Ss translate words that produced homophones. Low-frequent homophones behaved like high-frequent controls, inheriting the accessing speed of their high-frequent homophone twins. Because homophones share the lexeme, not the lemma, this suggests a lexeme-level origin of the robust effect.

The word frequency effect in speech production was discovered by Oldfield and Wingfield (1965). In a picture-naming task, they found that pictures with low-frequency (LF) names (such as *syringe*) took longer to name than pictures with high-frequency (HF) names (such as *basket*). Wingfield (1968) established this effect as a genuinely lexical one. The effect was not due to differential speeds of object recognition but to naming itself.

In this research, we consider the word frequency effect in light of recent models of lexical access in speech production. In particular, we consider whether the frequency effect arises early, in selecting the semantically appropriate lexical item, or late, in the retrieval of the item's form information.

The article is organized as follows. We begin by sketching a model of the production lexicon. Against the background of this model, we then derive hypotheses about the potential loci of frequency effects from data in the literature on picture naming, hesitations in spontaneous speech, and speech errors. Finally, we turn to the core of this article, an experimental investigation of these hypotheses. In seven experiments, we analyze whether the frequency effect is due to either conceptualization or articulation (it is neither), whether it arises at the early, so-called *lemma level* (there is a frequency effect there, but it turns out not to be the frequency effect), or whether it originates at the late *word form*, or *lexeme level* (it does).

A Model of the Production Lexicon

According to current models, the production of spoken language involves three major levels of processing (Bock, 1982; Dell, 1986; Garrett, 1975, 1976, 1980; Kempen & Huijbers, 1983; Levelt, 1989). First, there are the processes of *conceptualization* that guide the generation of so-called preverbal messages specifying the concepts to be verbally expressed. Second, there are the processes of *formulation*, which involve the mapping of preverbal messages onto linguistic form. These formulation processes can be broken down into the processes of *grammatical encoding*, the selection of semantically appropriate lexical items and the generation of a syntactic frame or surface form, and *phonological encoding*, the computation of the phonetic form of the intended utterance, an articulatory plan that can be buffered and executed. Third, there are the processes of *articulation*, involving the retrieval of the phonetic plan, as well as the initiation and execution of articulation, with overt speech as a physical consequence.

Lexicalization is the set of processes in formulation that govern the retrieval of lexical entries from the mental lexicon. It proceeds in two steps, corresponding to grammatical and phonological encoding, respectively: *lemma selection* and *word form*, or *lexeme retrieval* (e.g., Butterworth, 1980a, 1989; Dell, 1986; Fromkin, 1971; Garrett, 1975, 1976, 1988; Kempen & Huijbers, 1983; Levelt, 1983, 1989). During lemma selection, a semantically appropriate item is selected; it is syntactically but not yet phonologically specified. During word form retrieval, the lexical item's phonological form is accessed. Evidence for the two-step theory of lexical access comes from speech error analyses (e.g., Garrett, 1988), investigations of the tip-of-the-tongue state (e.g., R. Brown & McNeill, 1966; for a review see A. S. Brown, 1991), analyses of prelexical hesitations (Butterworth, 1980b; Garrett, 1982), and experimental studies (Kempen & Huijbers, 1983).

Although the distinction between lemma selection and word form retrieval is now widely accepted, existing models diverge on the issue of whether these two processes interact. In a pioneering article, Dell and Reich (1981) provided statistical

Jörg D. Jescheniak and Willem J. M. Levelt, Max-Planck-Institut für Psycholinguistik, Nijmegen, The Netherlands.

We wish to thank Ger Desserjer, Hans Franssen, and Johan Weustink for invaluable technical assistance and Monika Baumann, Jean Fox Tree, Antje Meyer, Paul Meyer, Thomas Pechmann, Ardi Roelofs, Herbert Schriefers, Linda Wheeldon and four reviewers for helpful comments and discussions.

Correspondence concerning this article should be addressed to Jörg D. Jescheniak, who is now at Institut für Psychologie, FB12, Freie Universität Berlin, Habelschwerdter Allee 45, D-14195 Berlin, Germany. Electronic mail may be sent to joergj@zedat.fu-berlin.de.

evidence for the existence of a lexical bias effect in the generation of phonological errors (trouble in correctly retrieving a phonological form tends to lead to existing words) as well as a mixed-error bias (error words exhibit both semantic and phonological resemblance to the target word). From this, they concluded that the two processes are unlikely to be completely independent. These observations have been the major motivation for interactive models of lexical access, most notably Dell's (1986), which conceived lexical selection and word form encoding as fully interactive processes. Levelt et al. (1991a) argued that such models must also have specific implications for the time course of lexical access in normal (nonerroneous) speech processing. Their experimental results did not support the interactive view. In particular, there was no trace of lemma activation at the later, phonologically active stage; this, however, had to be expected on the interactive account. Rather, the data supported a modular version of the two-step theory, which states that accessing lemmas strictly precedes accessing lexemes, without feedback. In addition, the activation of a lexeme appeared to be contingent on the selection of the corresponding lemma, not just its activation. This argued against noninteractive models of the cascading type, in which there is unconditional unidirectional flow of activation. Further evidence for the modular theory was obtained by Schriefers, Meyer, and Levelt (1990). These following articles provide extensive and productive discussions in the literature: Dell & O'Seaghda (1991), Dell & O'Seaghda (1992), Harley (1993), Levelt (1992), and Levelt et al. (1991b). At present, it makes sense to explore all models concurrently.

The lexical production model advanced in our research is of the modular two-step type and largely follows that of Roelofs (1992). The lexicon is conceived of as an activation spreading network that specifies the semantic, syntactic, and phonological properties of words. Figure 1 shows a somewhat simplified fragment of the network.¹ The network contains three layers of nodes that are connected by arcs. Note that the arcs are directionally labeled for their type only; this does not necessarily coincide with the flow of information.

The top layer displays the conceptual stratum. Strictly speaking, this stratum is not part of the lexicon itself. Instead, it represents the propositional semantic system, as involved in cognitive activities such as reasoning. In language production it provides the conceptual input that guides the selection of lexical elements. Each concept for which there is a word in the language (i.e., each lexical concept) corresponds to a node at this level. The word's meaning is represented by the set of labeled links to other conceptual nodes. The meaning of "horse," for instance, is represented by the conceptual node *HORSE* and its network of conceptual connections. One of them, depicted in Figure 1, is the *isa* link to *ANIMAL*, which specifies a superordinate category of *HORSE*. So far, the model follows Collins and Loftus (1975).

In contrast to the Collins and Loftus (1975) model, the conceptual layer first feeds into a layer of abstract, phonologically unspecified lexical entries or lemmas. The lemma stratum is the first truly lexical layer. In Roelofs's (1992) treatment, lexical selection is defined as selection of a lemma. A lemma's role is to mediate between conceptual, syntactic, and phonologi-

cal lexical information. Within its stratum, a lemma connects to nodes that represent the word's syntactic properties, such as its syntactic category (cat), its gender (gen), or its subcategorization features. In the given example, the lemma *paard* (Dutch for *horse*) is specified as a noun with neuter gender.

Finally, each lemma node projects onto its associated lexeme node in the third stratum of the model. The lemma *paard*, for instance, connects to the lexeme /pa:rd/. At this level, the word's abstract phonological form is represented, in particular its segmental content and its metrical properties (syllabicity and accent structure). For reasons of simplicity, Figure 1 represents only the segmental information.

There are several reasons for distinguishing between lemmas and lexemes (see Kempen & Huijbers, 1983, who introduced these terms, and Levelt, 1989, for a fuller treatment). Most people have experienced the occasional tip-of-the-tongue state. In this state, one accesses the lemma but cannot retrieve the lexeme. This problem is a basic symptom of anomic aphasia. French, German, or Dutch anomic patients can access the lemma when the lexeme is inaccessible. The patient usually produces the noun's correct, gender-marked article, even when access to the noun's phonology is blocked (see Henaff Gonon, Bruckert, & Michel, 1989, for evidence from a French patient).

Within the lemma stratum, abstract syntactic nodes represent a lexical entry's syntactic category and gender. The model assumes that all nouns project onto the same noun node and that all nouns of one gender class (masculine, feminine, or neuter) are linked to a single node representing that gender. The model also incorporates a property of the Dutch gender system that we exploited in the experiments reported in this article. In Dutch, nouns of masculine and feminine gender require *de* as the singular definite article. Thus, the syntactic nodes representing feminine and masculine gender are linked to the same article lemma *de*. Nouns of neuter gender, on the other hand, are linked to the article lemma *het*. Dutch speakers refer to these nouns as *de* words and *het* words; we adopt these terms in this article. There needs to be two nodes representing masculine and feminine gender because some anaphors, such as personal pronouns, take different forms depending on whether they refer to a noun of masculine gender, for example, *hij[he]* or one of feminine gender, for example, *zij [she]*.

Usually a lemma is activated from its corresponding concept. However, the process is more complicated for so-called function words, such as articles. The activation of their lemmas depends in part or in full on the activation of other lemmas. Accessing an article lemma requires first accessing the corresponding noun's lemma. According to our model, access to the gender marking article lemma is mediated by the abstract gender node. See Levelt (1989, p. 238) and Bock and Levelt (in press) for a fuller treatment of this "indirect election" of function words.

¹ We do not, for example, consider how morphological complex words are stored and retrieved (see Levelt, 1989).

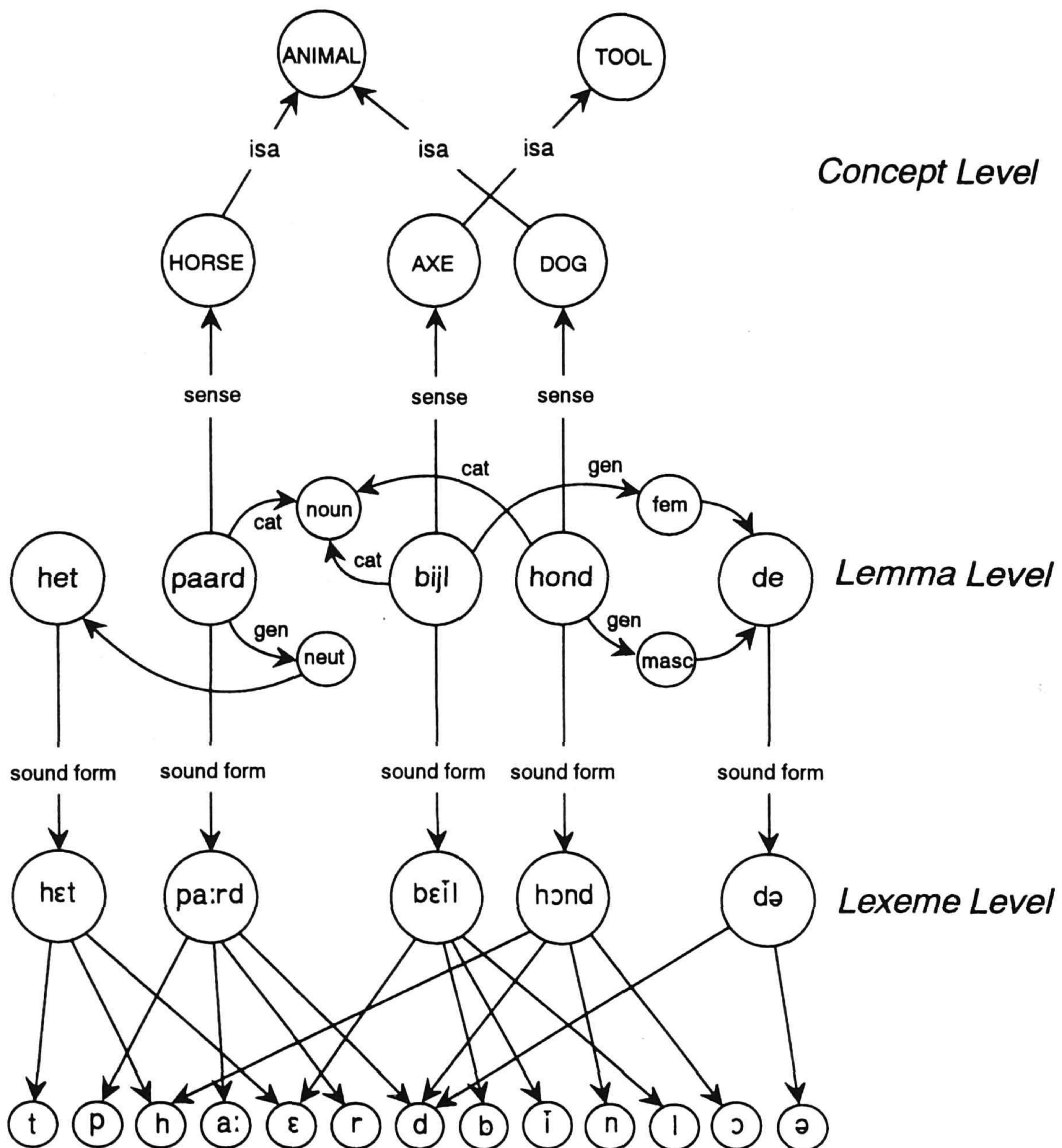


Figure 1. Fragment of the lexical production network. Cat = category; gen = gender; isa = is an instance of; fem = feminine; neut = neuter; masc = masculine.

Evidence for the Locus of Frequency Effects

Since Oldfield and Wingfield's (1965) discovery, the word frequency effect in production has figured in studies of picture naming, of prelexical hesitation, and of speech errors. These studies provide hypotheses about the locus of the effect.

Picture Naming

As previously mentioned, one important finding was that the word frequency effect appears to be lexical in origin. The effect

disappears if the objects have to be recognized but not named (cf. Bartram, 1976; Jescheniak, 1993a; Wingfield, 1967, 1968). Kroll and Potter (1984), however, did find a (24-ms) frequency effect in an object decision experiment. Although the word frequency effect in naming has repeatedly been confirmed (Humphreys, Riddoch, & Quinlan, 1988; Huttenlocher & Kubicek, 1983), these studies did not add to a further refinement of the locus discussion. The models that were used distinguished only between a concept and a word, and the frequency effect was usually adduced to the word level.

However, our model does distinguish between a word's lemma and its lexeme. Naming studies have, until recently, not acknowledged that distinction. The distinction, however, originated from work on prelexical hesitations and speech errors, and they do provide some initial but inconclusive indications about the origin of the frequency effect.

Prelexical Hesitations

MacKay and Osgood (1959) found for their sample of spontaneous speech that 47% of the speech pauses did not coincide with grammatical junctures; they were genuine hesitations. These pauses, whether filled (with *er* or the like) or unfilled, were found to precede open-class words more often than closed-class words (see also Martin & Strange, 1968). Garrett (1982), on the basis of normal and aphasic word-finding problems, argued that

in the normal course of events, prelexical hesitations be viewed as form based—i.e. the hesitation arises not out of a search for a lexical item which satisfies conceptual constraints, but rather out of processes which retrieve items from the form-based inventory. (p. 65)

In terms of our model, prelexical hesitations are caused in accessing lexemes, not lemmas. The question then arises whether prelexical hesitations depend on the frequency of the word they precede. If so, there would be evidence for a lexeme-level locus of the word frequency effect; the closed-class words in the MacKay and Osgood study were, after all, HF items. The hypothesis that prelexical hesitation is word frequency dependent was tested by Beattie and Butterworth (1979). They carefully distinguished between a word's frequency and its predictability in context and found that predictability, not frequency, correlated with hesitation probability. In a study of precolor word hesitations in pattern descriptions, however, Levelt (1983) did find a correlation between prelexical filled pauses and color word frequency, but almost no predictability effect. Hence, the hunting ground is still open; it may be the case that prelexical hesitation is both form and frequency related. This would testify to a lexeme-level locus of the word frequency effect. The delay in the retrieval of an LF lexeme may lead to a noticeable disruption of the utterance.

Speech Errors

Speech errors have long been the primary data of production research, and it has been repeatedly shown that LF words are more susceptible to errors than HF words (Dell, 1988, 1990; Stemberger, 1984; Stemberger & McWhinney, 1986). With respect to the frequency effect in speech errors, the same question can be asked as before: Is the observed frequency effect a result of accessing lemmas or of accessing lexemes? Until recently, the only relevant evidence here was the differential frequency sensitivity of meaning-based substitution errors (such as *belt* substituting for *collar*) as compared with form-based substitutions (such as *freeze* for *phrase*). Whereas the former are assumed to reflect a failure of lemma selection, the latter are probably due to trouble in word form retrieval (Garrett, 1980). A suggestive pattern emerges from comparisons of the intruding word's frequency and the target

word's frequency in these two types of word substitution. In meaning-based substitution errors, the intruding word is as likely to be lower in frequency than the intended word as it is to be higher in frequency. The situation, however, is much different for form-based substitution errors. With this type of error, the intruding word usually substitutes for a lower frequency word (del Viso, Igoa, & Garcia-Albea, 1991; Hotopf, 1980; Kelly, 1986). Such a pattern suggests that frequency is coded only at the lexeme level. However, this conclusion must be tentative. The effects reported are small, as are some of the error samples.² Clearly, further evidence is needed.

Dell (1990), to our knowledge, was the first to study the issue experimentally. In a study of experimentally elicited speech errors, he replicated the frequency effect that was observed earlier: An LF target is more susceptible to phonological error than an HF target. This effect was, moreover, independent of the target word's syntactic category (open or closed class). What is important for our purposes is Dell's finding of a homophone effect. LF words with an HF homophone (such as *wee* with homophone *we*) are as little prone to induced phonological errors as are their high-frequent twins. That is, an LF homophone profits from sharing its word form with an HF word. Although this seems to indicate that frequency is coded at the lexeme level, not at the lemma level (homophones share their lexeme but not their lemmas), Dell argued for a lemma-level explanation within his interactive model. In a computer simulation including lemma, lexeme, and segment nodes, he manipulated lemma resting-level activations according to the frequency of the words. In case of homophonic words, two lemmas project onto the same lexeme. When an LF homophone lemma is active, it spreads activation to its lexeme. During the next step, the latter feeds activation back to the homophonic twin lemma. Subsequently, both lemmas will send activation to their joint lexeme and down to its constituent segments. In this way, the LF homophone inherits the activation power of its HF twin. Of course, the model produces such a pattern only because it allows for feedback between levels. However, this assumption has been challenged. Whereas an interactive model can freely locate frequency effects at the lemma or the lexeme level, this is not so for a modular model. There, homophone data will be decisive for the one or the other level.

The three kinds of data reviewed are insufficient in locating the frequency effect with confidence. The most reliable data are the latency data from object recognition studies. They make a conceptual origin of the effect unlikely, although this has to be carefully checked in view of the results of Kroll and Potter (1984). It still needs to be shown that the effect is not due to the initiation of articulation. If the effect is indeed lexical, then new data are needed to distinguish between a lemma-level and a lexeme-level interpretation of the effect. We chose to use picture naming as our basic research paradigm. In addition, we used recognition, gender decision, and

² Also, this pattern is not always obtained. As Merrill Garrett (personal communication, June 1992) pointed out to us, no such differential effect of frequency is apparent in the Massachusetts Institute of Technology corpus.

translation tasks as needed. In all cases, response latency was the dependent variable.

Experiment 1: Picture Naming

The main purpose of Experiment 1 was to reliably replicate the word frequency effect in picture naming. In particular, we wanted to make sure that an effect could be obtained while controlling for word length and morphological complexity across the frequency contrast. A second purpose of the experiment was to assess the robustness of the effect over repetitions. Is the word frequency effect ephemeral, that is, dissipating with repeated use of the word? Or is it structural, insensitive to repeated processing of an individual item? Because the experimental task involved visual processing and object identification, both stages of lexical access, as well as articulation, any frequency effect obtained may be due to contributions from any of these levels of processing.

Method

Participants

Twelve native speakers of Dutch, recruited from the subject pool of the Max Planck Institute for Psycholinguistics, served as participants in the experiment. They received Dfl 8.50 (approximately \$4.50) for their participation. Participants were screened for normal or corrected-to-normal vision. None of them participated in more than one of the experiments reported here.

Materials

Forty-eight pictures, one half with an LF and one half with an HF name were sampled from a picture database at the Max Planck Institute. The pictures were line drawings of simple objects that were digitized by using a Hewlett-Packard scan program and were refined with the drawing program Designer. The following selection criteria were applied. (a) Only pictures with morphologically simple names and high name agreement were selected. Name agreement was checked in an informal pretest. (b) LF names were considered to be those with token frequencies of less than 12 in 1 million, and HF names were considered to be those with token frequencies of more than 60 in 1 million according to the lemma-based frequency counts in the Celex database.³ (c) LF and HF names were groupwise matched for word length as assessed by the number of syllables and segments (which constrained the set of potential items to mono- and bisyllabic words). (d) Also, although a perfect match of word onset was not possible, we were careful to ensure that no systematic differences in onset segments between the two groups of items occurred; word-initial consonants were equally distributed over the conditions. This is important when collecting voice-key data (see Pechmann, Reetz, & Zerbst, 1989). (e) All experimental items were *de* words; *het* words were used as fillers. Not only are *het* words in Dutch much harder to control for word length and morphological complexity than *de* words (there are many more *de* words), but comparability with later experiments required this restriction (to obtain a consistent response to the experimental items in gender decision).

Mean token lemma frequencies per million for LF and HF words were 6.0 and 150.7, respectively (word length in segments 3.7 each, in syllables 1.1 each; see Appendix A for a complete list of items).

The 48 experimental items were intermixed with 48 filler items (with names belonging to the *het* class). The names of the filler items covered a wide frequency range. Also, the fillers were selected in a way that no

obvious asymmetries in semantic domains would arise between items of different gender classes. Again, this was important for later experiments that involved gender decision on the same experimental items.

Finally, there were 10 practice pictures, one half with a *de* name and one half with a *het* name; they shared the word length and frequency characteristics of the experimental items.

Each of the 96 test items and each of the 10 practice items were to be presented three times. This way, each participant would receive a total of 318 trials, in which all practice trials preceded all test items. Four pseudorandomized trial sequences were constructed that had to be the same as in Experiment 4; each sequence was to be given to a different group of participants. This led to the following constraints on the randomization procedure. (a) No presentation of an experimental item was preceded by the presentation of a phonologically, semantically, or associatively related item. (b) No more than 5 items of the same gender class were presented in adjacent trials. (c) Repeated presentations of an individual experimental item were separated by at least 20 trials. Other than that, the sequences were random.

Procedure

Each participant was tested individually in a session lasting about 45 min. The participant was comfortably seated in a dimly lit, sound-attenuated booth. All visual stimuli were displayed centered on a high-resolution NEC Multisync II CRT as light-gray line drawings on a black background. Display size of the picture stimuli was approximately 80 mm × 80 mm. Viewing distance was about 60 cm. A Hermac AT computer controlled the display of all visual information and the on-line collection of the data. Participants responded into a Sennheiser microphone, and speech-onset latencies were measured with a voice key connected to the computer. On the experimenter's CRT, trial information and reaction times were displayed. The experimenter monitored the participant's responses through headphones and scored them for correctness. All sessions were taped with a Sony DTC 55 ES DAT-recorder. On one channel, a participant's vocal responses were recorded, and on the other channel, markers were set at the onset of the stimulus picture and the triggering of the voice key. These recordings were later consulted by the experimenter when in doubt about the correctness of a participant's response.

Each trial started with a visual warning signal (*) presented on the screen for 200 ms. Following a pause of 600 ms, the target picture was displayed. The whole picture appeared instantly on the screen. The timer was started simultaneously with the picture onset. Response latencies were measured to the closest millisecond. The display duration of the target picture was contingent on the participant's response. It disappeared as soon as a vocal response was initiated. However, if no response was registered within 2,000 ms, the picture disappeared anyway, and 1,500 ms later the next trial began.

³ Celex is the lexical database developed for English, Dutch, and German at the Center for Lexical Information at the University of Nijmegen. It is now available on CD-ROM through the Linguistic Data Consortium. Its Dutch version is based on 42,380,000 word tokens. All frequency counts reported here were taken from a lemma-based lexicon and refer to the token frequency in a million words. In this context, a lemma signifies the abstract representation that underlies an inflectional paradigm. So, for example, the lemma *close* represents not only the word form *close*, but also *closes*, *closed*, and *closing*. Thus, the lemma frequency of *close* corresponds to the sum frequency of all listed word forms. We also calculated frequency counts on the basis of a word form lexicon. For the particular item sets tested here, these counts do not differ in any systematic way from the lemma frequency counts.

At the beginning of the experiment, the participants studied a written instruction booklet that emphasized both the speed and accuracy of their responses. To reduce the proportion of deviant naming responses, we asked our participants to go through a booklet showing the stimuli. Next to each picture, the noun was printed that was used most frequently to name that object in the pretest. Participants were asked to use only the listed names. After having randomly assigned a participant to one of the four trial sequences, the experiment began with the series of practice trials. Following a short pause, the first half of the experimental trials were presented. After a second pause, the remaining trials were presented.

Results and Discussion

Observations were discarded from the analyses whenever any of the following conditions held: (a) A picture had been named other than expected; (b) a nonspeech sound preceded the utterance of the picture name, triggering the voice key; (c) a dysfluency occurred or an utterance was repaired; and (d) a speech onset latency exceeded 2,000 ms or deviated from a participant's and an item's mean by more than two standard deviations. These data points were replaced by estimates following the procedure as described by Winer (1971). Also, estimates were computed whenever the voice key failed to trigger but the participant gave the correct response. However, in the latter case, no error was coded. On the basis of these criteria, a total of 92 observations (5.3%) were marked as incorrect.

Averaged reaction times and errors were submitted to analyses of variance (ANOVAs). Statistical analyses involved two fixed within-subject variables: frequency (low vs. high) and repetition (1 through 3). To allow generalizations over both subject and item populations (Clark, 1973), we computed separate analyses treating both subjects and items as random variables. In the by-subject analysis, each data point was based on 24 observations; in the by-item analysis it was based on 12 observations.

Figure 2 displays average speech onset latencies. Overall, naming latencies for pictures with LF names were 62 ms slower than naming latencies for pictures with HF names—LF: 711 ms, HF: 649 ms; $F_1(1, 11) = 82.02, p < .001, MS_e = 832$; $F_2(1, 46) = 12.19, p < .01, MS_e = 11,214$. Also, there was a highly reliable repetition effect, with responses becoming faster with repeated presentation—Repetitions 1 through 3: 714 ms, 673 ms, 654 ms, respectively; $F_1(2, 22) = 74.96, p < .001, MS_e = 302$; $F_2(2, 92) = 29.88, p < .001, MS_e = 1,517$. Figure 2 also shows that the frequency effect was invariant over repetitions. The absence of any interaction between the frequency and repetition variables was substantiated in the statistical analysis ($F_{s} < 1$). The analysis of error rates revealed only a reliable effect of frequency—LF: 7.2%, HF: 3.5%; $F_1(1, 11) = 5.37, p = .04, MS_e = 2.65$; $F_2(1, 46) = 8.02, p = .01, MS_e = 0.89$.

The experimental results are clear-cut. First, when target words were controlled for length and morphological complexity, pictures with HF names were named faster and more accurately. The highly reliable frequency effect of 62 ms provided an estimate of the effect against which any contribution of nonlexical sources, such as object identification or articulation, could be compared. Second, the frequency effect was invariant over repetitions. This might be surprising, given

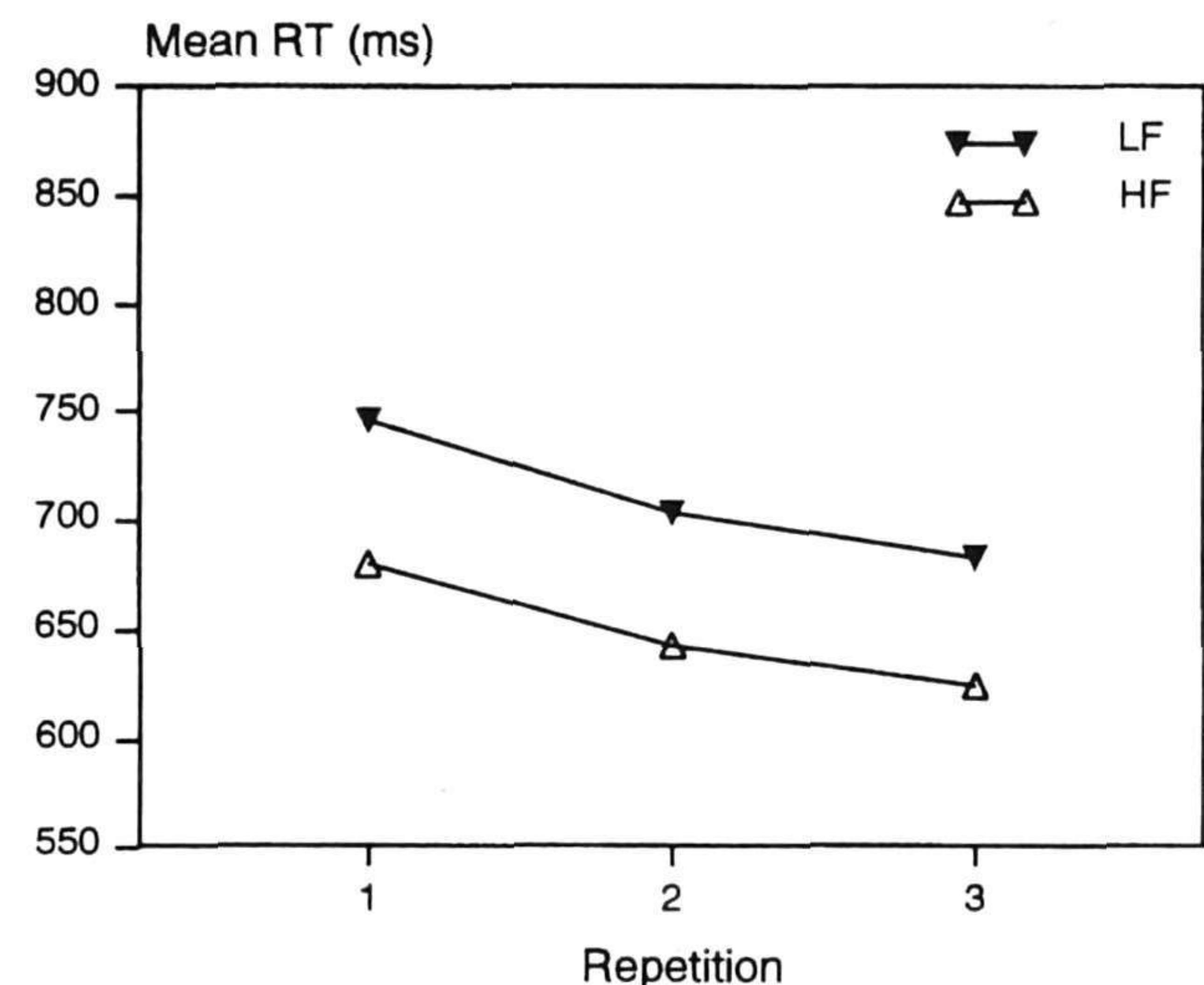


Figure 2. Naming latencies from Experiment 1. RT = reaction time; LF = low frequency; HF = high frequency.

that objective frequency counts could also be taken to reflect the recency of usage for a given speaker. Under this premise, one might expect the effect to decrease with repeated processing of the materials. However, the data show that this was not the case, at least not with three repetitions of an item. Clearly, whatever process gives rise to the frequency effect, it is not easily influenced by repetition, even if the preexposure of the pictures and their names might have had some effect. Third, a reliable frequency effect could be obtained within the chosen frequency ranges. The selected frequency ranges (LF < 12 tokens per million and HF > 60 tokens per million) are somewhat different from those usually reported in the literature. In most (word naming) studies, the selection of LF words is restricted to token frequencies below or equal to one per million. Our experiment showed, however, that such an extreme contrast is not essential when investigating frequency effects in conceptually driven word production.

Before we can assert that the obtained frequency effect is indeed lexical, we need to rule out two alternative accounts: (a) The frequency effect reflects differences in object recognition latencies and (b) it reflects differences in the speed of initiating articulation.

We address the former hypothesis first. It might be the case that less time is needed for the perceptual processing and categorization of objects with HF names. This issue is not a new one. As Wingfield (1968, p. 226) put it, "the speed with which an object is identified might be expected to relate in some systematic manner to its a priori probability in the environment." In other words, objects with HF names might simply be more common objects, which are more frequently encountered and therefore more easily identified.

The empirical evidence so far is not completely consistent. Wingfield (1967, 1968), Huttenlocher and Kubicek (1983), and Jescheniak (1993a) found no effect of frequency on object recognition, but Kroll and Potter (1984) did. It is not clear what caused these conflicting results, but some data reported by Bartram (1973, 1974) suggest that surface characteristics of

the picture stimuli might be important. However, all studies agree that object recognition alone cannot account for the full effect obtained in naming.

Experiment 2 was designed to assess the contribution of perceptual and conceptual processes to the effect observed in Experiment 1. If there is no frequency effect in a task that involves identification of the depicted object but not retrieval of its name, then the frequency effect has to be attributed to processes taking place during or after access to the mental lexicon.

Experiment 2: Object Recognition

We used a picture recognition task to test whether object identification latencies would relate to the object name's frequency. Participants saw a word immediately followed by a picture. Their task was to decide whether the word denoted the object in the picture and to push a *yes* or *no* button correspondingly. For example, after reading the word *krab* (*crab*), they had to respond *yes* if the picture showed a crab and *no* if the picture showed a cake. We measured the push-button latencies.

Method

Participants

Twelve paid participants recruited from the Max Planck Institute subject pool participated in the experiment.

Materials

The 48 experimental items used in Experiment 1 were intermixed with a new set of 48 filler items. In contrast to Experiment 1, each item was presented only once. Trials involving experimental items were always *yes* trials, and trials involving filler items were always *no* trials. Therefore, there was an equal number of *yes* responses and of *no* responses. We added an additional 20 practice items, one half of them to be presented in *yes* trials and one half of them in *no* trials. Four different pseudorandomized sequences were constructed with the constraint that no experimental item was preceded by a phonologically, semantically, or associatively related item.

Procedure

Each participant was tested individually in a session lasting about 15 min. All visual stimuli were presented centered on the screen. The words were displayed in lowercase Times Roman 35-point typeface. Individual characters were separated by blanks. Two push buttons were used, one for the *yes* response and one for the *no* response. The *yes* button was always assigned to the participant's dominant hand.

During a trial, the participant first saw a word for 1,000 ms. After a pause of 200 ms, the target picture was displayed for 2,000 ms. The timer was started at picture onset. It stopped when the participant pushed a button. If no response occurred within 2,000 ms, the response was coded as missing. After a period of 1,500 ms, the next trial began. At the beginning of Experiment 2, participants were tested on the set of 20 practice items. After a short pause, the 96 test items were presented.

Results and Discussion

Mean reaction times and error rates were submitted to ANOVAs with both participants and items as random variables. All responses longer than 2,000 ms and latencies deviating from a participant's and an item's mean by more than two standard deviations were treated as errors and replaced by estimates. This way, a total of 28 data points, or 4.9% of the data, were marked as incorrect. The main data analyses were carried out on the 48 experimental items, that is, the items requiring a *yes* response. They involved the fixed within-subject variable of frequency (low vs. high).

Object recognition latencies for pictures with LF and HF names were nearly identical. The 6-ms advantage of HF items was not significant (LF: 442 ms, HF: 436 ms; $F_s < 1$). Error rates were identical for both conditions, namely 4.9% each.

These results suggest that there is no word frequency effect in object recognition. However, the absence of an effect could be due to identity priming. A word may prime the recognition of a picture it denotes; this may obliterate any frequency effect in object recognition. If this priming hypothesis is correct, a frequency effect should be obtained when the word does not match the picture's name. We tested this by analyzing the reaction times for the filler items, that is, the items that had appeared in the *no* trials. Because the fillers had been drawn from the whole continuum of the frequency distribution rather than its extremes, we correlated reaction time with frequency instead of comparing means. The analyses revealed that contrary to the priming hypothesis, neither reaction times, $r(48) = .07, p = .65$, nor error rates, $r(48) = -.23, p = .11$, correlated with \log_{10} frequency. Therefore, there was no effect of name frequency on the speed and accuracy of object recognition, at least for the picture materials used here. In other words, a substantial frequency effect can be obtained in naming (as in Experiment 1) for pictures that do not show any frequency effect in recognition (as in Experiment 2). This frequency effect in naming must be due to lexicalization or articulation. This is the effect we analyzed in the following experiments. Still, given the Kroll and Potter (1984) findings, a genuine object frequency effect in object recognition may yet exist. That is not, however, the topic of our article. In Experiment 3 we focused on articulation, another possible source of our frequency effect.

Experiment 3: Articulation

Balota and Chumbley (1985) showed that word frequency can modulate the speed of articulatory programming and articulation to some extent. High-frequent articulatory programs for words may be compiled and executed faster than low-frequent ones. The main support for this view comes from the persistence of a frequency effect in a delayed naming task. In this task, participants see a word and prepare its pronunciation. After a variable delay, a cue signals them to initiate the vocal response. It is assumed that response preparation will proceed as far as it can. If the cue delay is long enough, the word will have been recognized and the articulatory motor program assembled and stored in a buffer. On recognizing the cue, the participant can retrieve and execute this motor

program. Any remaining effect of frequency, then, has to be attributed to the stage of response execution (however, see McRae, Jared, & Seidenberg, 1990; Monsell, Doyle, & Haggard, 1989; Savage, Bradley, & Forster, 1990). In Experiment 3 we investigated the contribution of articulatory processes to the effect obtained in Experiment 1, by using a delayed naming task. To ensure that our participants fully prepared the utterance at cue onset, we used only long cue delays, ranging from 1,000 to 1,600 ms (see Savage et al., 1990, for relevant arguments). In Experiment 3, the picture names rather than the pictures themselves were used to elicit responses. We did this to keep our experimental procedure comparable to the standard procedure reported in the literature (cf. Balota & Chumbley, 1985; Forster & Chambers, 1973). However, we see no principled reason to expect qualitatively different results with picture stimuli.

Method

Participants

Twelve paid participants recruited from the Max Planck Institute subject pool participated in the experiment.

Materials

The experimental words were the 48 names of the pictures from the preceding experiments. Additionally, twice as many filler words were selected, reflecting frequency, morphological complexity, and word length of the experimental words.

Procedure

Participants were tested individually in sessions lasting about 15 min each. They were explicitly told that cue onset would be variable across trials and could not be predicted. At the beginning of each trial, the target word was displayed in lowercase Times Roman 35-point typeface for 1,000 ms centered on the computer screen. Individual characters were separated by blanks. After variable delays of 1,000 ms, 1,300 ms, or 1,600 ms, a cue (!) indicated participants to initiate the utterance of the word. The cue remained visible until the response was initiated, and 1,500 ms later the next word was displayed. However, if no response was given within 2,000 ms after cue onset, the cue disappeared anyway, and 1,500 ms later the next trial began. The experimental words were always followed by a delay of 1,000 ms and an equal number of fillers by delays of 1,300 ms and 1,600 ms. Cue delay in itself was not an object of theoretical interest. Rather, variable delays were used only to prevent participants from anticipating cue onset.

Results and Discussion

Averaged reaction times were submitted to ANOVAs involving only the fixed variable frequency. Observations were discarded from the analyses if any of the following conditions held: (a) The response latency exceeded 1,500 ms; (b) a nonspeech sound had preceded the target word, triggering the voice key; (c) a dysfluency or repair occurred; or (d) the utterance was initiated before the response cue appeared. Finally, the data were corrected for outliers. On the basis of these criteria, 15 observations (2.6%) were coded as errors. Only one observation was categorized as being of the last

condition (d). Keeping the cue delay variable obviously prevented participants from anticipating the cue onset.

Mean speech onset latencies were nearly identical for LF and HF words (375 ms and 368 ms, respectively) and did not differ statistically, $F_1(1, 11) = 1.64, p = .23, MS_e = 192; F_2 < 1$. Mean error rates for HF words exceeded those for LF words (3.5% and 1.7%), but this difference was not reliable either, $F_1(1, 11) = 2.10, p = .18, MS_e = 0.50; F_2(1, 46) = 1.74, p = .19, MS_e = 0.30$.

When the target words were named with a delay, no reliable effect of frequency was obtained, and whatever there was went in opposite directions for latency and error proneness. We take this as evidence that articulatory processes do not notably contribute to the frequency effect as obtained in Experiment 1. Our finding contrasts with what Balota and Chumbley (1985) have observed, and we return to this divergence in the General Discussion.

Together with the results from Experiment 2, the data suggest that the word frequency effect observed in picture naming (Experiment 1) is indeed a purely lexical effect. Neither processes of conceptual identification nor processes of articulatory initiation are substantially involved. The next question to be addressed is at what point in the lexicalization process the frequency effect arises.

Experiment 4: Gender Decision

According to our modular two-step model of lexical access, lexicalization proceeds in two steps. First, a lemma is selected and in a subsequent step, the associated word form is retrieved. In the model, a lexical item's syntactic properties are stored at the lemma level (see Figure 1). The syntactic properties include the item's grammatical gender. According to the model, activation spreads immediately from an activated lemma to its gender node, and this process is independent of lexeme activation. Therefore, activation of a word's grammatical gender can be used to trace the first step of the lexicalization process: lemma selection. If accessing a word's lemma is frequency dependent, gender activation should be frequency dependent as well. In Experiment 4, we presented the same pictures as in Experiments 1–3. This time, however, participants did not name the pictures; rather, they were instructed to decide on the names' grammatical gender. In particular, we asked our participants to decide on the singular definite article that the object's name takes: Is the name a *de* word or a *het* word? In Dutch, masculine and feminine words are *de* words, whereas neuter words are *het* words. The participants pushed a *de* or a *het* button according to their decision. It should be noted that gender is not phonologically encoded in Dutch. Although morphological encoding of gender does occur in Dutch (e.g., all diminutives are neuter), we chose materials in which gender was not derivable from word form properties.

In Experiment 4 we also addressed a more trivial account of the frequency effect observed in naming. Participants may be slower on less frequent items because they are less confident about the object names. Because normative name agreement indices (comparable to the indices provided by Snodgrass & Vanderwart, 1980) were not available for the present item set, such an account cannot be rejected a priori. However, it clearly

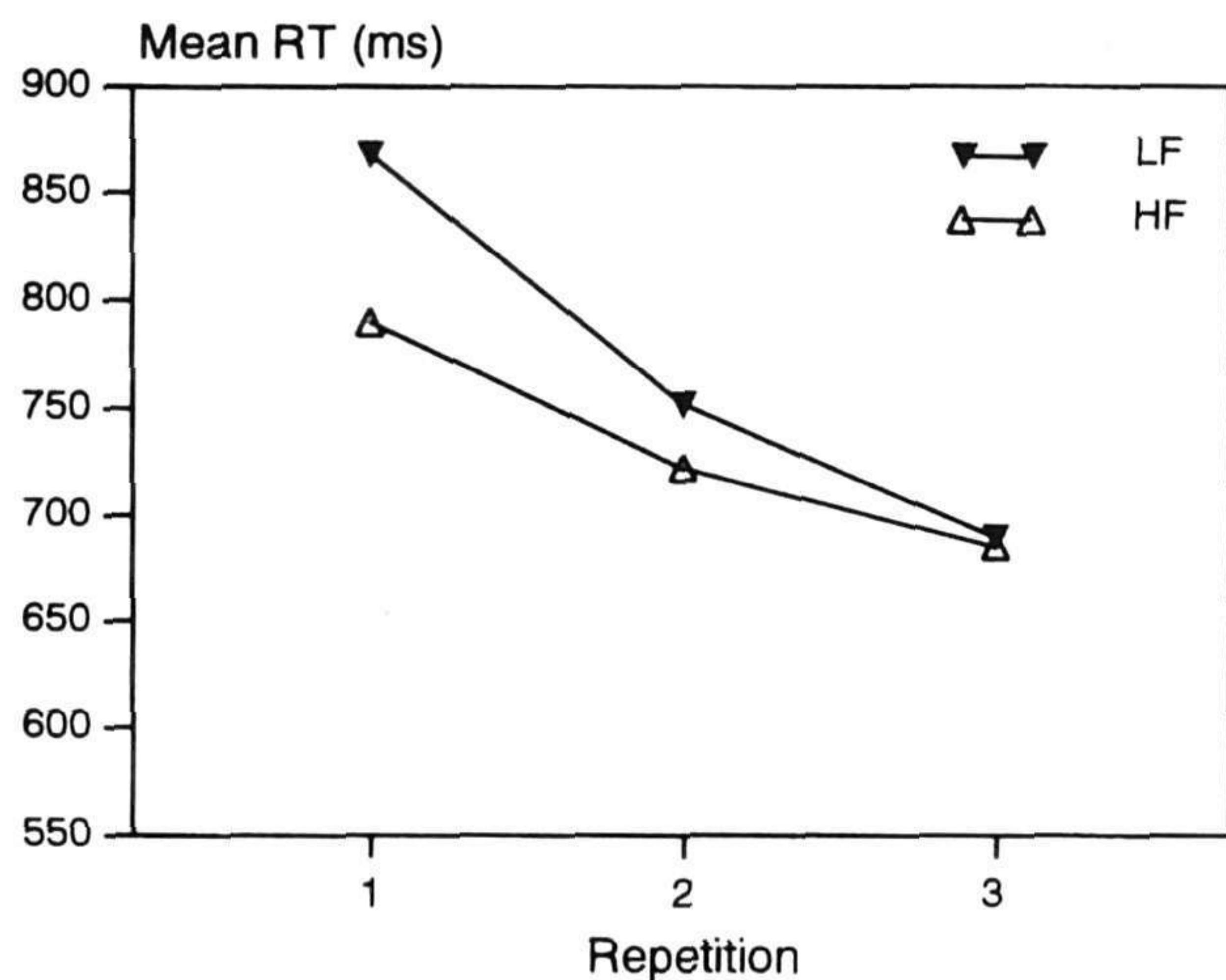


Figure 3. Gender decision latencies from Experiment 4. RT = reaction time; LF = low frequency; HF = high frequency.

attributes the effect we observed in naming to the process of lexical selection, that is, selection of the appropriate lemma. It converges with the lemma activation threshold account in that it predicts the same effect in gender decision as in naming.

Method

Participants

Twelve paid participants recruited from the Max Planck Institute subject pool participated in the experiment.

Materials

The picture stimuli were the same as in Experiment 1. Because there were 48 *de* items (the experimental items, one half of them with a LF name and one half of them with a HF name) and 48 *het* items (the filler items), each being presented three times, the probability of *de* and *het* responses was .50 each.

Procedure

The basic procedure, including the randomization of trials, the structure, and the timing of each trial, was identical to Experiment 1. The only difference was the type of response required. Instead of naming the picture, participants decided on the picture name's gender by pressing either a button labeled *de* or a button labeled *het*, whichever was appropriate. Before the experiment started, participants previewed the pictures. They then received the written instructions revealing details of the experimental task. Participants were informed that each response would be required equally often.⁴ The *de* button was always assigned to a participant's dominant hand. The design of the experiment did not allow us to check which picture names the participants actually used during the experiment. Therefore, immediately after the experiment, we gave them a second booklet containing the experimental pictures and asked them to write the object's name next to each picture. If they recorded a name other than expected, we excluded the corresponding observations from the statistical analyses.

Results and Discussion

All incorrect push-button responses, and latencies exceeding 2,000 ms or deviating from a participant's and an item's mean by more than two standard deviations, were treated as errors and replaced by estimates. The same was done for those items for which participants had used a name other than expected. This way, a total of 84 observations (4.9%) were discarded. The statistical analyses were carried out on the 48 experimental items, that is, the items requiring a *de* response. The analyses involved the fixed within-subject variables frequency (low vs. high) and repetition (1 through 3).

Average gender decision latencies are presented in Figure 3. Overall, the gender decision was performed more rapidly to HF items than to LF items. The 36-ms difference was significant in the by-subject analysis but only marginally significant in the by-item analysis—LF: 769 ms, HF: 733 ms; $F_1(1, 11) = 10.22, p < .01, MS_e = 2,419; F_2(1, 46) = 3.72, p = .06, MS_e = 13,227$. The variable repetition yielded an effect as well. Repeated presentations of an experimental item resulted in faster reaction times—Repetitions 1 through 3: 829 ms, 737 ms, 688 ms, respectively; $F_1(2, 22) = 138.61, p < .001, MS_e = 887; F_2(2, 92) = 76.69, p < .001, MS_e = 3,211$. Most interesting is the interaction between the two variables: $F_1(2, 22) = 10.76, p < .001, MS_e = 780; F_2(2, 92) = 5.19, p < .01, MS_e = 3,211$, reflecting a diminishing frequency effect over repetitions. This was further analyzed by means of Newman-Keuls paired-comparison tests (with $p = .05$). In the by-subject analysis, a reliable effect of frequency was obtained for Repetitions 1 and 2, but not for Repetition 3. In the by-item analysis, a significant effect was obtained for Repetition 1 only, but none for Repetitions 2 and 3. This suggests that the effect for Repetition 2 in the by-subject analysis is carried by a few items only and does not extend to the whole sample of words. The error rates showed a similar pattern. However, only the variable repetition reached significance—Repetitions 1 through 3: 8.0%, 4.2%, 2.5%, respectively; $F_1(2, 22) = 7.02, p < .01, MS_e = 1.59; F_2(2, 92) = 11.47, p < .001, MS_e = 0.49$.

Clearly, the frequency of the picture's name affected the decision on its grammatical gender. This held for the first presentation of an item only, when taking both the subject-based and the item-based analysis into account. Whereas initially the effect size was comparable to the one obtained in the naming experiment—77 ms in gender decision and 62 ms in naming—it rapidly decreased and entirely disappeared when the items were presented for the third time. This sharply contrasts with the naming results of Experiment 1. In naming, the word frequency effect was robust over repetitions; it was just as pronounced on the third naming trial as it had been on the first.

How can we account for this divergence? The most obvious explanation is that it reflects the participants' accommodation to the task. During their initial gender decisions they might silently generate the pictures' names in full noun phrases (i.e., article plus noun) and then monitor for the article in their internal speech. This will produce a word frequency effect as in

⁴ This was done to eliminate a bias toward the *de* response, which is likely to exist given that the distribution of the two gender classes in Dutch is not symmetrical.

Experiment 1 because the words and their forms are fully accessed. On later trials they would become more efficient, deriving the gender information without accessing the word forms. If frequency is indeed encoded at only the word form level, these later trials would no longer show a frequency effect.

To check for this, we familiarized 12 new participants to the gender decision task by means of a set of training items. Afterward, we presented them with the experimental items. If unfamiliarity with the task caused the initial effect of frequency in gender decision, then no frequency effect should be obtained on the latter items. Contrary to that prediction, though, the effect still appeared in gender decision, and it was of about the same size as before. This told us that repeated processing of the same item is essential for the frequency effect to dissipate in gender decision. In Experiments 5a and 5b, we investigated whether it is previous access to the same item or, more specifically, to its gender that makes the frequency effect vanish.

Experiments 5a and 5b: Naming and Gender Decision

In Experiments 5a and 5b, we did not familiarize the participants with the gender decision task but with the test items, by means of a naming task. After having named each item twice, the participants were (unexpectedly) given the gender decision task on the same items, also twice. The question is whether having recently accessed the item in naming suffices to eliminate the frequency effect in gender decision or whether something more specific must occur, namely accessing the item's gender information. The former was tested in Experiment 5a in which the naming task was as before: The participants produced the appropriate noun. The latter was tested in Experiment 5b in which the participants produced the full noun phrase, that is, article plus noun (producing the article involves accessing the item's gender information). In all other respects, the experiments were identical and are jointly discussed.

Given the results of Experiment 1, we expected to find a pronounced frequency effect in the naming data, and it should be invariant over the two repetitions, when participants produce the object name only, as in Experiment 5a. The matter is more complicated in Experiment 5b in which the full noun phrase is produced. If access to the article is fast, participants may initiate their response as soon as the article is available, without waiting for the noun's form to be available. If form access is the locus of the frequency effect, this strategy will lead to a diminished frequency effect in the naming trials. The crucial question for Experiments 5a and 5b, however, concerns the gender decision latencies. Will the frequency effect disappear in one or both of the experiments?

Method

Participants

Twenty-four paid participants were recruited from the Max Planck Institute subject pool. One half of them participated in Experiment 5a and one half of them in Experiment 5b.

Materials

The same pictures as in Experiment 1 were used.

Procedure

The complete experiment lasted about 55 min. Naming and gender decision trials were blocked. Participants always started with the naming task. The procedure for this task was identical to the one used in Experiment 1. In Experiment 5a, participants were instructed to give a simple naming response; in Experiment 5b, participants were instructed to give a full noun phrase naming response consisting of the name plus its definite article. They were told that a second experiment would follow immediately after the naming experiment. However, they were not informed about the nature of this second experiment. Also, they were unaware that the same picture stimuli were going to be used in the second experiment. The procedure for the subsequent gender decision task was identical to the one used in Experiment 4.

Each picture was presented four times, twice in the naming section and twice in the gender decision section. The experiments began with 10 practice items repeated three times each. Each participant received a total of 444 trials consisting of 48 experimental items and 48 filler items, each of which was presented four times, and 60 practice items. The two parts of the experiment were separated by a pause. There were also short pauses after the practice trials and after half of the naming and gender decision trials.

Results

Experiment 5a

The naming and gender decision data were analyzed in separate analyses. All responses longer than 2,000 ms or deviating from a participant's and an item's mean by more than two standard deviations were treated as errors and replaced by estimates. For the naming data, the same was done whenever a participant had used an unexpected name. If a participant had consistently used a different name during naming, the gender decision data for the respective item were also coded as errors. On the basis of these criteria, 116 observations, (5.0% of the data) were coded as errors, 58 observations (5.0%) in the naming part and 58 (5.0%) in the gender decision part. Figure 4 displays average reaction times. Naming and gender decision data are reported separately.

Naming. Overall, pictures with HF names were named 72 ms faster than pictures with LF names—HF: 697 ms, LF: 769 ms; $F_1(1, 11) = 96.65, p < .001, MS_e = 647; F_2(1, 46) = 13.81, p < .001, MS_e = 9,096$. Also, reaction times decreased from the first to the second presentation from 766 ms to 700 ms, $F_1(1, 11) = 101.99, p < .001, MS_e = 513; F_2(1, 46) = 67.62, p < .001, MS_e = 1,547$. As in Experiment 1, the two variables did not interact ($F_{s} < 1$); the frequency effect in naming was robust. Analyses of the error rates revealed a marginally significant repetition effect only—Repetitions 1 and 2: 6.3% and 3.8%, respectively; $F_1(1, 11) = 6.49, p = .03, MS_e = 0.63; F_2(1, 46) = 3.53, p = .07, MS_e = 0.58$.

Gender decision. Overall, gender decision was 34 ms faster for pictures with HF names than for those with LF names. This effect was significant in the by-subject analysis—HF: 736, LF: 770; $F_1(1, 11) = 23.23, p < .001, MS_e = 631$, but marginal in the by-item analysis, $F_2(1, 46) = 3.38, p = .07, MS_e = 8,656$. Reaction times decreased by 64 ms from the first to the second

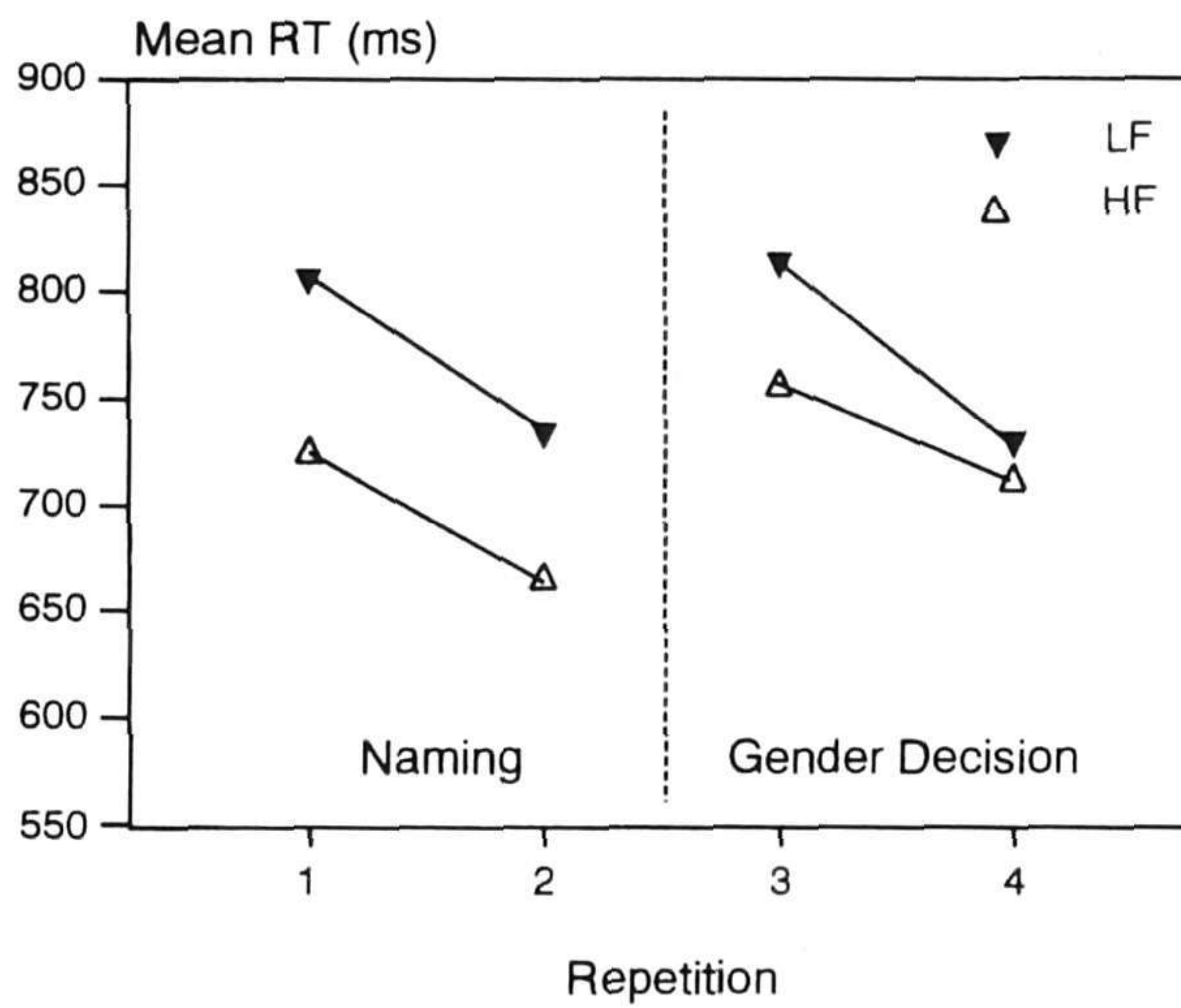


Figure 4. Naming and gender decision latencies from Experiment 5a. RT = reaction time; LF = low frequency; HF = high frequency.

presentation of an item, yielding a highly reliable repetition effect—Repetitions 1 and 2: 785 ms and 721 ms, respectively; $F_1(1, 11) = 73.62, p < .001, MS_e = 672; F_2(1, 46) = 71.43, p < .001, MS_e = 1,384$. More important, however, was a reliable interaction between frequency and repetition: Whereas the frequency effect amounted to 54 ms at the first presentation, it shrank to 15 ms at the second presentation, $F_1(1, 11) = 9.30, p = .01, MS_e = 484; F_2(1, 46) = 6.64, p = .01, MS_e = 1,384$. Newman-Keuls planned pairwise comparisons (with $p = .05$) showed that a reliable frequency effect was obtained at only the first presentation. This held for both the by-subject analysis and the by-item analysis. In the analysis of error rates, no significant effects were obtained.

Experiment 5b

The raw data were handled in the same way as in Experiments 5a. After applying the same criteria, a total of 79 observations (6.9%) of the naming data and 57 observations (4.9%) of the gender decision data were marked as incorrect. Mean reaction times are presented in Figure 5. As before, naming and gender decision data are reported separately.

Naming. Again, a strong frequency effect of 53 ms was obtained—LF: 773 ms, HF: 720 ms; $F_1(1, 11) = 26.12, p < .001, MS_e = 1,309; F_2(1, 46) = 6.16, p = .02, MS_e = 11,128$. Also, naming latencies decreased with repetition—Repetitions 1 and 2: 769 ms and 724 ms, respectively; $F_1(1, 11) = 27.07, p < .001, MS_e = 899; F_2(1, 46) = 30.26, p < .001, MS_e = 1,612$. There was no interaction between these two variables; although the frequency effect was numerically larger for the first than for the second presentation (59 ms vs. 47 ms, respectively), this was far from being significant, $F_1(1, 11) = 1.11, p = .31, MS_e = 383; F_2 < 1$. The analysis of the error rates revealed no significant effects.

Gender decision. Overall, reaction times for pictures with LF and HF names were nearly identical (LF: 721 ms, HF: 722 ms; $F_s < 1$). The variable repetition yielded a highly reliable effect—Repetitions 1 and 2: 748 ms and 695 ms, respectively;

$F_1(1, 11) = 28.75, p < .001, MS_e = 1,195; F_2(1, 46) = 46.52, p < .001, MS_e = 1,474$. The Frequency \times Repetition interaction was just significant in the by-subject analysis, $F_1(1, 11) = 4.68, p = .05, MS_e = 319$, but not at all in the by-item analysis, $F_2(1, 46) = 2.03, p = .16, MS_e = 1,474$. Actually, the effect in the by-subject analysis reflected a crossover interaction, and Newman-Keuls planned pairwise comparisons (with $p = .05$) revealed that the variable frequency had no effect at either of the two levels of the repetition variable. Analysis of the error rates revealed an effect of repetition only—Repetitions 1 and 2: 7.0% and 3.0%, respectively; $F_1(1, 11) = 6.14, p = .03, MS_e = 1.79; F_2(1, 46) = 9.57, p < .01, MS_e = 0.58$.

Discussion

As predicted, we replicated the finding from Experiment 1. We found a pronounced effect of word frequency in naming, and the effect was not modulated by repetition. This was true for both Experiments 5a and 5b. Although the frequency effect slightly decreased from Experiment 5a to 5b, it was still a substantial 53 ms in the latter case. Initiating an article plus noun noun phrase is strongly noun frequency dependent; we return to this issue in the General Discussion.

The crucial issue was the effect of naming on the subsequent gender decision. Here, the patterns for the two experiments clearly diverge. In Experiment 5a, after noun naming, there was an initial frequency effect, but the effect disappeared at the second presentation of the pictures in gender decision. This replicates the finding of Experiment 4. In Experiment 5b, however, after article plus noun naming, there was no frequency effect in gender decision.

Apparently, the dissipation of the frequency effect in gender decision is not merely due to repeatedly accessing the item. Instead, it is essential that the item's (lemma linked) gender information is repeatedly accessed. That was the case in Experiment 4 (repeated gender decision) and in Experiment 5b (repeated access to the gender-dependent article), but not in Experiment 5a.

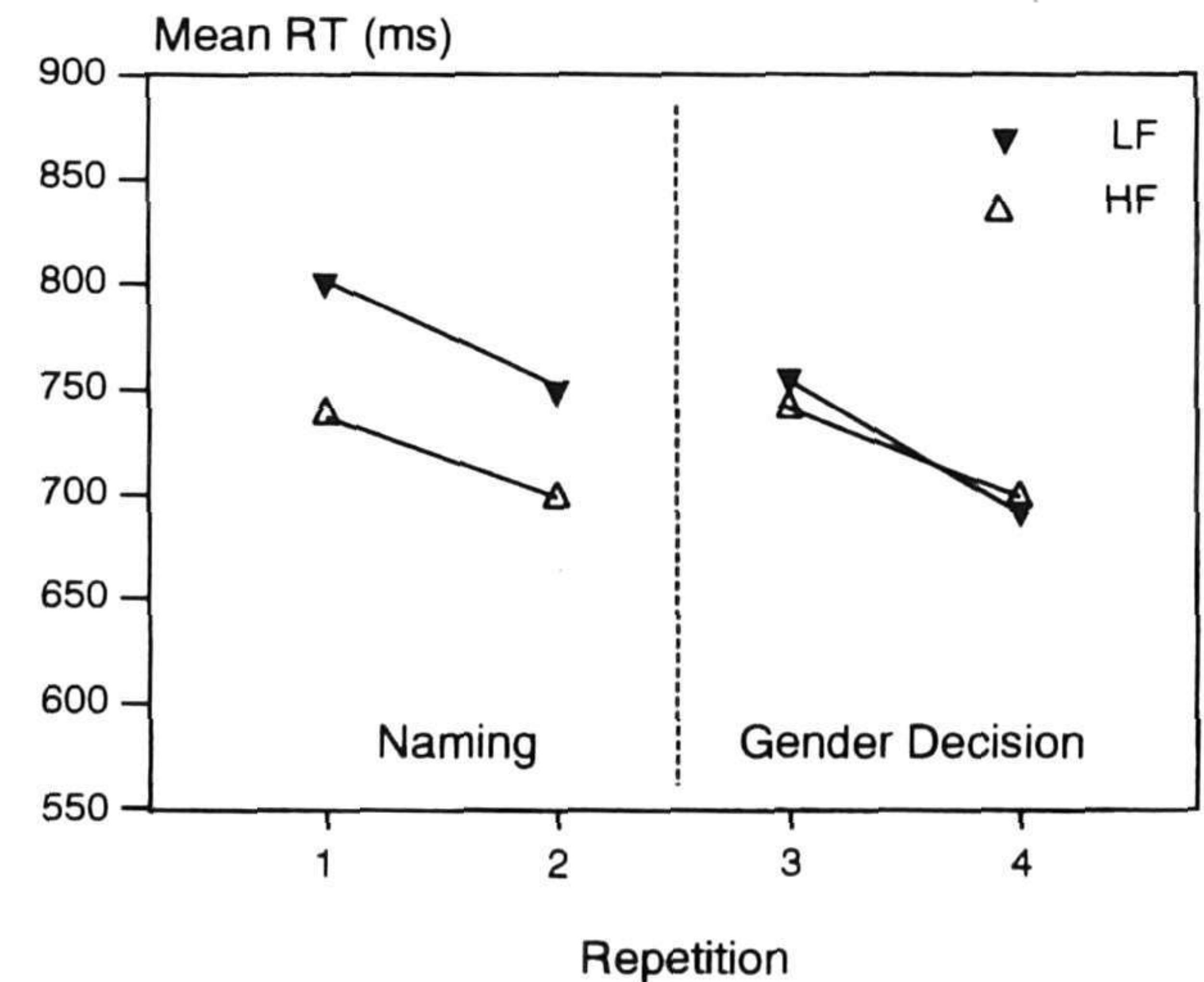


Figure 5. Naming and gender decision latencies from Experiment 5b. RT = reaction time; LF = low frequency; HF = high frequency.

What can we conclude from these findings and those in the previous three experiments with respect to the locus of the word frequency effect? First, we did not observe a robust frequency effect in gender decision, that is, an effect that is preserved over repetitions of an item. This contrasts with the frequency effect in naming, which is immutable in character. Second, the ephemeral effect in gender decision is probably not derived from the robust effect in naming. The gender decision effect cannot be explained by a strategy of internal naming on the participant's part, which lessens over trials. It is more likely that we have discovered a different frequency-dependent effect altogether. This is further confirmed by the finding that the effect in gender decision disappears after an article plus noun naming task on the same items. At the same time, the effect does not disappear within that naming task: The frequency effect was as large on the second trial in Experiments 5a and 5b as it was on the first. This testifies to a different origin of the two effects.

What, then, could be the cause of the ephemeral effect? We suggest that it is a recency effect. The connection strength between a lemma and its gender node (see Figure 1) increases every time the word's gender information is used and decays slowly thereafter. On average, selection of gender information is more recent for HF words than for LF words. Therefore, accessing gender information is faster for HF words than for LF words. When recency of use is equated between HF and LF words, as was the case for the later trials in Experiment 4 and in Experiment 5b, frequency dependency should diminish or disappear. We return to this explanation in the General Discussion.

This explanation needs further empirical exploration, but this is beyond the objectives of our research. Our aim here was to determine the locus of the robust word frequency effect. The results of Experiments 1–5 do not support the hypothesis that word frequency is encoded in the activation thresholds of lemmas. Also, they rule out the possibility that the participants' uncertainty about the names of the LF items, that is, the target lemmas, is the source of the effect. Given the model in Figure 1, we are left with two major possibilities to account for the robust effect: (a) Word frequency is encoded in the lemma-to-lexeme connection strengths and (b) word frequency is encoded in the lexeme activation thresholds. Experiment 6 was designed to directly contrast the out-of-lemma and the lexeme hypotheses.

Experiment 6: Production of Homophones

In Figure 6, the model representations of a pair of homophones and a nonhomophonic word are compared. Homophones, such as *we* and *wee*, have different lemmas but share their lexeme, that is, their spoken word form information. For a nonhomophone, the frequency of its lemma equals the frequency of its lexeme (ignoring cases in which a lemma is selected but not phonologically encoded; see the General Discussion). In Figure 6, $F_{\text{lem}3} = F_{\text{lex}3}$. For a homophone, however, the lexeme frequency is the sum of the two lemma frequencies; in Figure 6, $F_{\text{lex}1} = F_{\text{lem}1} + F_{\text{lem}2}$. The out-of-lemma hypothesis predicts that the speed of accessing the common lexeme from lemma_1 is a function of $F_{\text{lem}1}$, whereas accessing it from lemma_2 is determined by $F_{\text{lem}2}$. This means

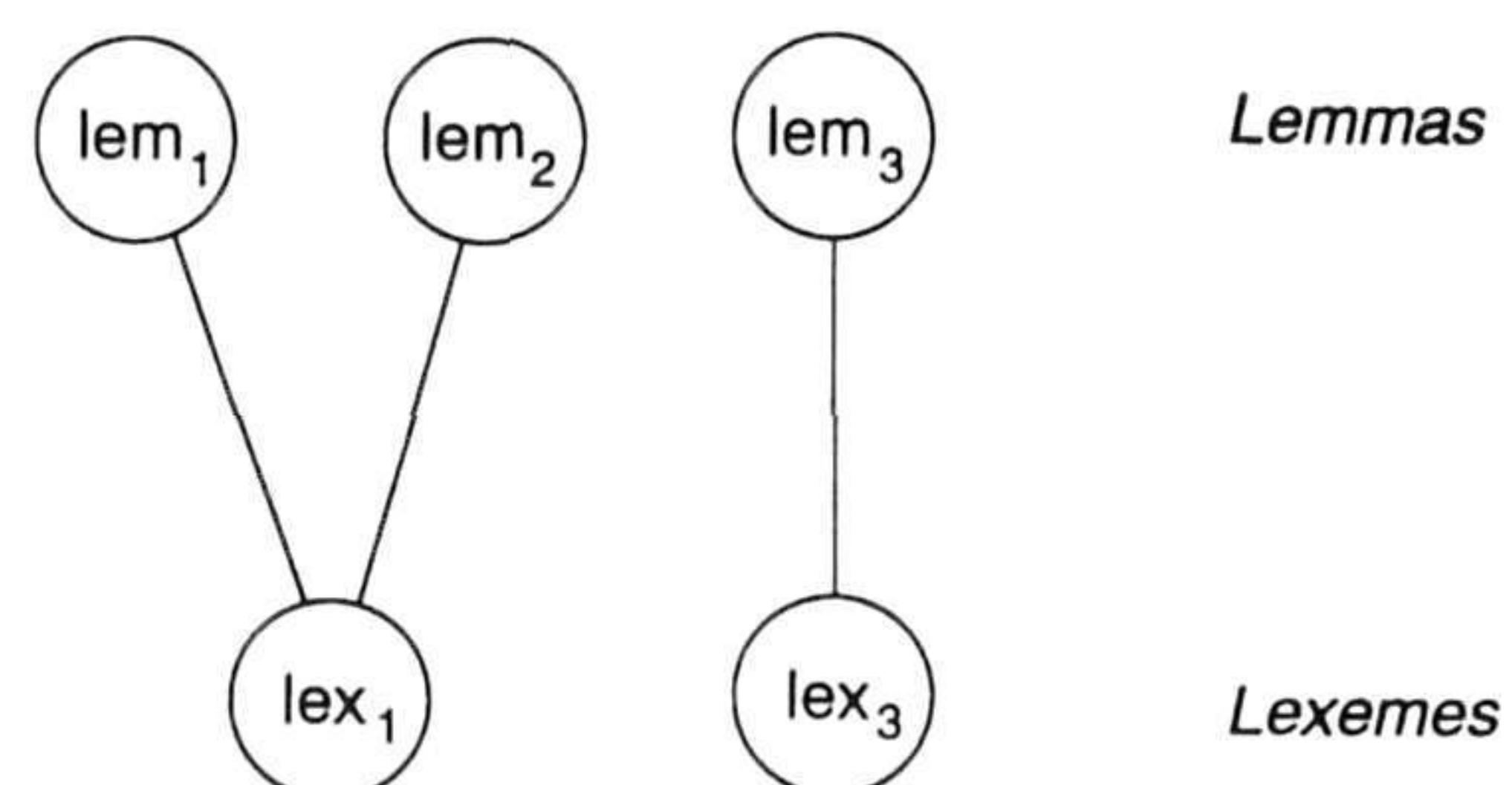


Figure 6. Model representation of homophones (left) and nonhomophones (right).

that accessing the lower frequency homophone of a pair will be slower than accessing the higher frequency one. According to the lexeme hypothesis, however, accessing speed is determined by the frequency of the shared lemma, that is, by $F_{\text{lem}1} + F_{\text{lem}2}$. This means that the two homophones will have the same naming latency and that it should be of the same order of magnitude as the latency of a nonhomophone with frequency $F_{\text{lem}1} + F_{\text{lem}2}$. Although we already excluded a lemma-threshold explanation of the word frequency effect, that hypothesis predicts the same homophone result as the out-of-lemma hypothesis. Hence, both of these explanations can be rejected if homophone latencies turn out to be determined by the sum frequency of the two homophonic words.

How can this be tested experimentally? We needed a task that necessitated access to both the lemma and the lexeme. A general difficulty is that homophones that are not related in meaning are relatively rare. A second difficulty is that picture naming tasks are impractical; most homophones are difficult or impossible to picture. For example, how would *we* and *wee* be pictured? Instead of a picture-naming task, we used an English–Dutch translation paradigm. Most participants in the Max Planck Institute subject pool have adequate to good knowledge of English. This makes it possible to give them the following simple task: to produce the Dutch translation equivalent for a visually presented English word. We can then measure the translation latency. This task involves accessing the Dutch lexeme; the Dutch word form information has to be retrieved to produce the word. How certain, though, can we be that this task also involves lemma access?

Theories of bilingual lexical representation diverge on the issue of whether corresponding words from different languages are directly connected through associative links within the bilingual lexicon (word association hypothesis), or whether this connection is established only through a conceptual representation (concept mediation hypothesis; e.g., Kirsner, Smith, Lockhart, King, & Jain, 1984; Potter, So, von Eckardt, & Feldman, 1984). An important variable determining the nature of the link between translation equivalents seems to be a speaker's expertise. Although Potter et al. (1984) argued that concept mediation holds for all bilingual speakers, regardless of second-language proficiency, Kroll and Curley (1988) provided some evidence compatible with the view that multiple translation routes exist. Whereas second-language learners may use a word association route in early phases, concept mediation dominates as they become more proficient. Both accounts, however, converge on the assumption that a proficient bilingual's translation is conceptually mediated. Because

participants tested in the following experiment would have some substantial second-language experience (at least 6 years of experience in speaking and reading English), it appears likely that they would rely on a conceptual route. For them the task would involve conceptually mediated lemma access and thus meet our requirements. Even if our participants were to respond through word association, our requirement could be met as long as the intralexical associative links were between lemmas and not between lexemes. Links between lexemes are unlikely. Translation equivalents do not only differ in form (ignoring cognates) but also often in syntactic properties, such as gender (e.g., *moon* is masculine in German [*der Mond*] but feminine in French [*la lune*]). According to our model, gender and other syntactic information is accessible only at the lemma level. Therefore, adequate translation of such terms in syntactic context will, as a rule, require lemma-level access. Bilingual competence is precisely the ability to acknowledge a word's syntax in using it. Hence, there will be lemma access on both accounts of the bilingual lexicon, at least for speakers with some substantial second-language experience.

However, there is a more serious problem with the translation task. Its first step consists of recognizing the visually presented English word, and this will be reflected in the translation latency. In other words, translation latency is not a pure measure of accessing the Dutch target word. To control for this additional variable, we used the same English words in a semantic decision task. Participants were instructed to give a positive push-button response if the English probe word denoted an animate entity (animal or human). They were to give a negative response to all other words. Each participant would perform the two tasks in different sessions: (a) produce the Dutch translation equivalent for the English probe and (b) carry out the semantic decision on the English probe. The difference in reaction time between these two tasks, then, was a measure for the time it took to produce the Dutch target word while controlling for the contribution of probe recognition on the overall reaction time.

Although the semantic decision latency could account for any contribution of probe word processing, we still needed to ensure that the English probe words were familiar to the population of participants tested in the experiment. Therefore, we pretested our experimental words for familiarity. In this experiment we contrasted three sets of experimental words. The first set, which we refer to as the homophone condition, included low-lemma-frequency words with high-lemma-frequency homophones. One example is the Dutch word *bos*. Its high-frequent reading is "forest" and its low-frequent reading is "bunch" (the words also differ in gender). The homophone condition included *bos* [bunch] and many pairs that differed in syntactic category. We used only homophone pairs that were also homographic. We also used only homophone pairs in which the sum lemma frequency exceeded the lemma frequency of the LF member by both more than 30 occurrences in 1 million and by at least a factor of 2. This criterion created a large contrast between the sum lemma frequency of the homophones and the LF controls.

The second set was composed of lemma frequency matched words that did not have a homophone (LF controls). Each word in this set was selected to form a frequency match for an

LF homophone in the first set. For *bos* [bunch], the LF control word was *hok* [kennel].

The third set consisted of high-lemma-frequency words (HF controls). Each of these words had a lemma frequency that matched the sum lemma frequency of the two members of a homophone pair. For *bos* this was the word *hoek* [corner]. The lemma frequency of *hoek* [corner] is roughly the sum of the lemma frequencies of *bos* [bunch] and *bos* [forest].

It should be noted that the HF homophones were neither mentioned nor probed for production. In our example, the English word *forest* did not appear in the experiment.

LF and HF controls were always of the same syntactic class as the homophones they matched (noun or adjective). Also, all experimental words were nonanimate so as to obtain a consistent response in the semantic decision task. Across the three sets we also controlled word length, in terms of both number of syllables and number of segments, and the proportion of concrete and abstract words.

We needed these three sets of words to test our two alternative hypotheses. With the out-of-lemma hypothesis we expected that the homophone condition would produce a pattern similar to the LF condition. With the lexeme hypothesis we expected that the homophone condition would produce a pattern that was indistinguishable from the HF condition (which we refer to as homophone effect). With both hypotheses we expected that the HF control condition would yield faster reactions than the LF control condition (which we refer to as frequency effect). This effect is essential because it ascertains that the experimental paradigm is sensitive to word frequency.

In addition to the three experimental sets, we had a fourth set of *animate fillers*. For each item in the three sets of experimental words, we selected an animate control word. These items were roughly controlled for lemma frequency and probe word length to avoid any biases in the semantic decision task.

We selected 11 items for each experimental condition, plus an additional set of 33 fillers. They are listed in Appendix B. Mean lemma frequencies for the different conditions were 9.5 (homophones), 9.4 (LF controls), and 352.3 (HF controls). Sum frequency of the members of the homophone pairs was 355.1.

We pretested the selected set of probe words by collecting familiarity judgments. We did this to evaluate whether any of the selected English probe words were unfamiliar to the population of participants tested here and to assess whether there was a systematic difference in familiarity across experimental conditions. We found that most of the words were highly familiar to the participants. Although HF control probe words were slightly more familiar than the rest, no difference between LF control and homophone probe words was observed (for details see Jescheniak, 1993b).

Method

Participants

Twenty participants were recruited from the Max Planck Institute subject pool. For participation in both experimental sessions they received a total of dfi 17.00 (approximately \$9.00). All participants had at least 6 years of experience in reading and speaking English. When

signing up for the experiment, they were informed that they would perform a task on English words and that good knowledge of that language was required. The data of one additional participant were discarded from the analyses because he was unable to perform the translation task with reasonable accuracy (more than 60% errors in one of the experimental conditions), as were the data from an additional participant who produced a high number of unreliable voice-key measures caused by extreme variations in loudness.

Materials

The selected 66 items were split into two blocks. One block contained 6 homophones, the corresponding 6 LF control words and 6 HF control words, and 18 animate filler words. The second block contained the remaining words. Additionally, for each block two fillers (one animate and one inanimate) were selected for presentation at the beginning of the respective block. Within a block, each item was presented three times (except the two block-initial fillers), resulting in a total of 110 and 92 trials, respectively. The trials were pseudorandomized with the constraints that (a) no more than 5 items requiring the same semantic decision would be presented in adjacent trials, (b) no homophone or control item would be preceded by a phonologically or semantically related probe or response, and (c) repeated presentations of any experimental item were separated by at least 10 intervening trials. For each block, two versions were created. Block version was completely crossed with block sequence, resulting in four different versions of the test materials. An equal number of participants were randomly assigned to each version, but each participant received the same version in both sessions. This way, the difference scores between translation and semantic decision latencies were based on observations in truly homologous positions across both parts of the experiment. Finally, an additional block consisting of 30 different items, one half animate and one half inanimate, was presented as a practice block.

Procedure

Participants were tested individually in two sessions lasting about 35 min each. The sessions were separated by an average of one week. In the first session, participants carried out the translation task. In the second session, they performed the semantic decision task. The apparatus was the same as in the preceding experiments. During the semantic decision session, they indicated their decision by pressing either a button labeled *yes* or a button labeled *no*. For each individual participant, the *yes* button was assigned to her or his dominant hand.

A trial was structured as follows. First, a warning signal (*) was presented for 200 ms. After a pause of 400 ms, the probe word was displayed. The probe onset and timer started simultaneously. Probe word display time was dependent on the participant's response: As soon as a response was initiated, the probe disappeared and about

Table 1
Mean Translation Latencies (in Milliseconds) and Error Rates (in Percentages) by Condition and Repetition for Experiment 6

Repetition	Condition					
	Homophones		LF controls		HF controls	
	M	%	M	%	M	%
1	861	8.2	963	11.4	827	6.4
2	775	5.5	869	6.8	745	3.6
3	752	5.0	831	6.8	724	5.9
Total	796	6.2	888	8.3	765	5.3

Note. LF = low frequency; HF = high frequency.

Table 2

Mean Semantic Decision Latencies (in Milliseconds) and Error Rates (in Percentages) by Condition and Repetition for Experiment 6

Repetition	Condition					
	Homophones		LF controls		HF controls	
	M	%	M	%	M	%
1	583	6.8	599	10.9	559	3.6
2	543	1.8	545	3.6	532	4.1
3	537	4.1	539	6.4	523	2.7
Total	554	4.2	561	7.0	538	3.5

Note. LF = low frequency; HF = high frequency.

2,100 ms later the next trial started. If no response occurred within 2,000 ms, a time-out code was recorded and about 2,100 ms later the next trial began.

Before the experiment started, participants studied written instructions. At the beginning of each block, they also briefly studied a booklet containing all probe words and their translation equivalents. Then the actual experiment began with the practice block followed by the two experimental blocks.

Results and Discussion

All statistical analyses were based on difference scores. These were obtained by subtracting each participant's semantic decision latency from that participant's translation latency for each item at each level of the repetition variable. Before computing these difference scores, we treated the raw data in the following way. First, we discarded observations and coded an error whenever any of the following conditions held. Translation data: (a) There was no response, (b) the response had been initiated after the 2,000-ms time-out interval, (c) a translation was other than expected, (d) a nonspeech sound produced before the onset of the word had erroneously triggered the voice key, or (e) an utterance had been repaired. Semantic decision data: (a) There was a wrong response or (b) the response latency had exceeded 2,000 ms. In all of these cases, observations were excluded in pairs. That is, if applying these criteria led to the exclusion of a translation response, the corresponding semantic decision response was also discarded and vice versa. In a next step, both translation and semantic decision data were replaced by Winer's (1971) procedure. Only then, in a final step, the difference scores were computed from the so-treated translation and semantic decision latencies. Although the incidence of erroneous translation data and erroneous semantic decision data was only 6.6% and 5.0%, respectively, the joint condition on difference scores left us with 11.2% missing data.

Tables 1 and 2 display mean translation latencies, mean semantic decision latencies, and the respective error rates broken down by condition and repetition. These data are listed for the sake of completeness; as mentioned, the statistical analyses were based on difference scores.

Averaged difference scores were submitted to ANOVAs, treating both participants and items as random variables. In the by-subject analysis, each data point was based on 11 observations, and in the by-item analysis, each data point was

based on 20 observations. The analyses involved the 2 three-level within-subject variables: condition (homophone vs. LF control vs. HF control) and repetition (1 through 3).

Figure 7 displays the mean difference scores. As can be seen, there was a clear effect of condition. The largest difference scores were found for the LF controls (327 ms) and the smallest difference scores for the HF controls (227 ms). The difference scores for the homophones were positioned in between (242 ms) the largest and the smallest scores. This pattern resulted in a reliable condition effect, $F_1(2, 38) = 77.26, p < .001, MS_e = 2,260; F_2(2, 30) = 5.40, p = .01, MS_e = 17,758$. Also, difference scores decreased with repeated presentation—Repetitions 1 through 3, respectively: 304 ms, 256 ms, 236 ms; $F_1(2, 38) = 65.98, p < .001, MS_e = 1,096; F_2(2, 60) = 30.37, p < .001, MS_e = 1,311$. As expected, the condition effect was largely unaffected by repetition ($F_s < 1$). To further analyze the condition effect, we computed Newman-Keuls paired-comparison tests (with $p < .05$). Both the by-subject analysis and the by-item analysis revealed that the HF control and LF control conditions differed significantly, as did the homophone and LF control conditions. Most important, there was no reliable difference between the homophone and the HF control conditions. The analysis of error rates yielded only a reliable repetition effect—Repetitions 1 through 3, respectively: 15.1%, 8.5%, 10.2%; $F_1(2, 38) = 9.68, p < .001, MS_e = 0.90; F_2(2, 60) = 9.51, p < .001, MS_e = 1.67$. Although for the LF control condition slightly higher error rates than for the other conditions were obtained (homophone: 10.3%; LF control: 14.8%; HF control: 8.6%), the condition effect was significant in only the by-subject analysis, $F_1(2, 38) = 5.41, p = .01, MS_e = 1.39$, but not in the by-item analysis, $F_2(2, 30) = 1.74, p = .19, MS_e = 7.85$.

It should be noted that these results did not depend in any significant way on the decision to analyze difference scores. Essentially the same pattern was obtained when analyzing translation latencies. In other words, both the homophone effect and the frequency effect surfaced in the translation data but were virtually nonexistent in the semantic decision data.

The results can be summarized in three points. First, there was a strong effect of word frequency. Thus, the experimental procedure was sensitive to the variable in focus here. Second, the production of a word did profit from an HF homophone. Actually, the homophones behaved like the HF controls, thus supporting the lexeme hypothesis and contradicting the prediction of the out-of-lemma hypothesis. Third, both the homophone effect and the frequency effect were robust over repetitions, which was the same pattern we observed in all earlier naming data.

We have argued that our translation task involves lemma access. However, we cannot exclude the possibility that our participants might have occasionally translated through a shallow lexeme-to-lexeme route. Still, this does not pose a fundamental problem for our interpretation of the data. Any such occasion should have reduced higher (concept or lemma) level contributions to the frequency and homophone effects. Still, we obtained substantial and robust effects. This further testifies to their lexemic origin and to the lexemic locus of the robust frequency effect obtained in Experiment 1. There is also some evidence, however, that our participants could not have relied on lexeme-to-lexeme associations exclusively. The evidence stems from semantically motivated translation errors. Although semantically motivated misselections occurred at a low rate (as such errors do in spontaneous speech), they did exist. Examples of such semantic substitution errors are the following: *parfum* [perfume] instead of *geur* [odor], *steen* [stone] instead of *kei* [boulder], *mus* [sparrow] instead of *mug* [mosquito], and *slang* [snake] instead of *slak* [snail]. It is important to point out that the error words were not restricted to the words that occurred as targets (as might be predicted under the lexeme-to-lexeme mapping hypothesis); some of the error words were neither targets in the same block nor targets in any of the preceding blocks. Semantically motivated word substitution errors have their source in a derailed lemma selection process (cf. Garrett, 1988). Although they can be accounted for naturally from lemma-mediated translation, it is

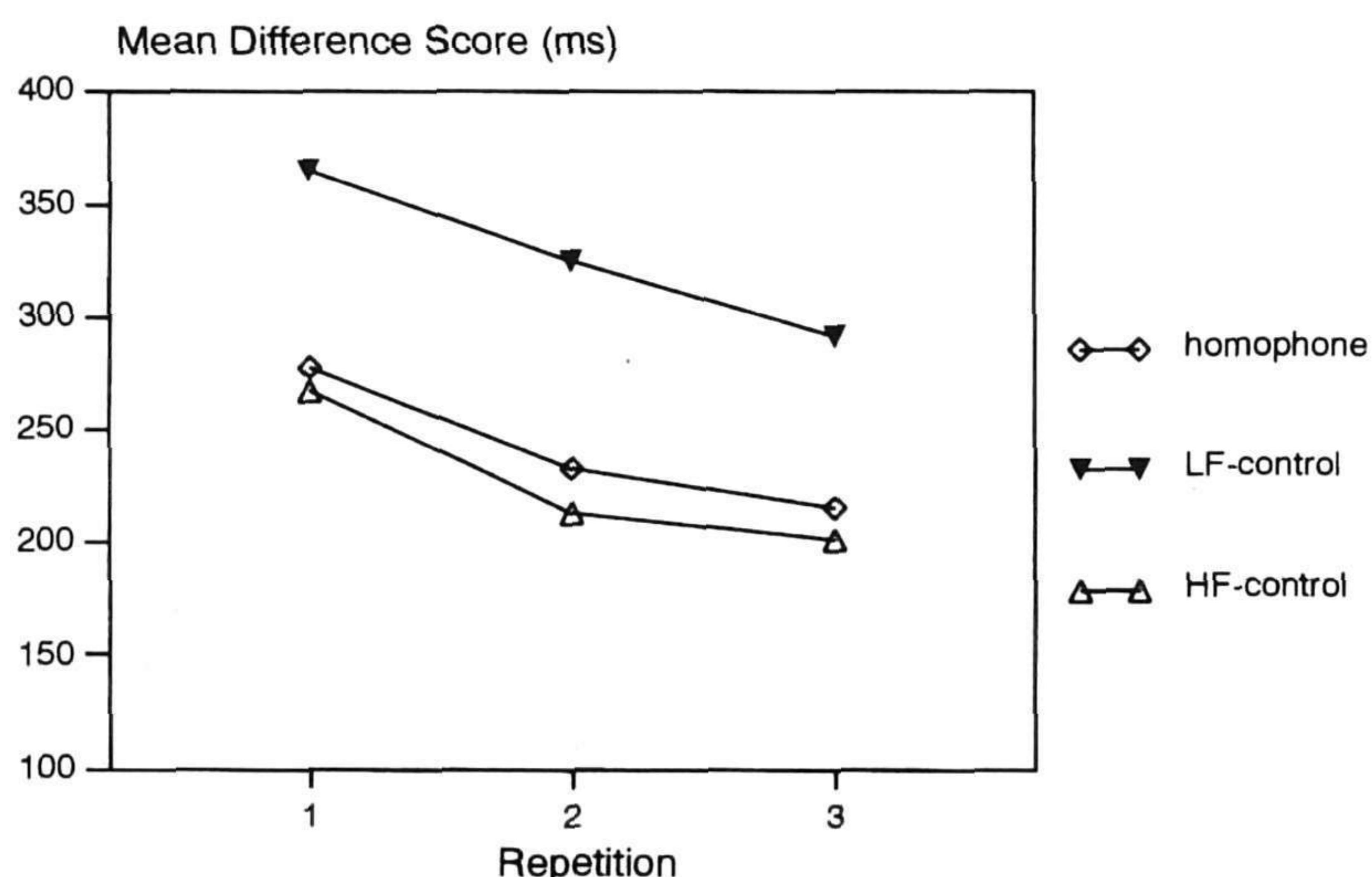


Figure 7. Difference scores from Experiment 6. LF = low frequency; HF = high frequency.

not obvious how they would arise under plain lexeme-to-lexeme mapping.

The pattern of results supports the hypothesis that word frequency is encoded as a lexeme threshold activation. Moreover, the finding is fully consistent with the model assumption, depicted in Figure 6: that homophone lemmas project onto the same lexeme node.

Dell (1990) has reported a related finding. He observed that an LF word is less susceptible to experimentally induced phonological errors if it has an HF homophone. Dell explained this effect as a lemma-level effect in an interactive two-stage model. A simulation study showed the feasibility of a lemma account. However, Dell did not rule out a lexeme-level account, and this account is, in our opinion, the more natural and simpler one. Given our data, it is the more natural one because lexeme-level frequencies, not lemma-level frequencies, determine response latencies. It is also simpler because the account does not need a feedback assumption. Both Dell's and our findings find a straightforward explanation in frequency-dependent lexeme thresholds.

To complete the argument for this conclusion, we ruled out a potential contribution of articulatory processes to our findings, as we did in Experiment 3 for the picture-naming findings. Our final experiment involved a delayed naming task on the materials of the homophone experiment. Experiment 7 was a control for material-dependent articulatory frequency effects (cf. Monsell et al., 1989; Savage et al., 1990).

Experiment 7: Articulation

Method

Participants

Twenty paid participants were recruited from the Max Planck Institute subject pool.

Materials

The words tested in this experiment were the 33 target words from Experiment 6. Additionally, twice as many filler words were used.

Procedure

Participants were tested individually in sessions lasting about 15 min each. The procedure was the same as in Experiment 3. Experimental words were always followed by a delay of 1,000 ms and an equal number of fillers by delays of 1,300 ms and 1,600 ms.

Results and Discussion

The raw data were treated as in Experiment 3. Only one response was categorized as initiated prior to cue onset.

Neither speech onset latencies nor error rates revealed any difference between the three experimental conditions. Mean speech onset latencies for the homophone, the LF control, and the HF control conditions were 366 ms, 360 ms, and 358 ms, respectively, $F_1(2, 38) = 1.39, p = .26, MS_e = 221; F_2 < 1$. Error rates for the homophone, the LF control, and the HF control conditions amounted to 1.8%, 2.7%, and 2.3%, respectively, yielding no significant condition effect here ($F_s < 1$).

The absence of any effect in delayed naming, then, indicated that both the frequency effect and the homophone effect observed in Experiment 6 were truly lexical effects, unpolluted by contributions from articulatory processes.

General Discussion

The experiments reported in this article investigated the locus of the word frequency effect in speech production. Experiment 1 established the effect in a picture-naming task in which the form complexity of the target words was controlled. In addition, the experiment showed that the frequency effect is resistant to repetition of the items; hence, we called this the robust word frequency effect. Experiment 2, involving an object recognition task, excluded object identification as the source of the robust effect. Experiment 3, involving delayed naming, excluded the initiation of articulation as a possible source. Together, these experiments showed that the effect is genuinely lexical in nature. A few remarks on the reliability of this conclusion are indicated in the proceeding paragraphs.

In our object identification experiment we used the same method as Wingfield (1968) in his classical study and obtained the same results. In addition, we excluded a priming explanation by showing the absence of a frequency effect in the *no* responses. That finding agrees with Huttenlocher and Kubicek (1983), who showed that (picture-picture) priming in picture naming was fully additive to the word frequency effect. In other words, even if our (and Wingfield's) method would have caused priming, this would not have reduced, let alone annihilated, the frequency effect. Therefore, it is safe to say that our Experiment 1 results are not due to object recognition. As discussed, this does not exclude the possibility that a genuine object frequency effect might be demonstrable in an object recognition experiment. However, that is a different topic.

What about potential articulatory contributions to the word frequency effect? We did not obtain a reliable frequency effect in delayed naming in Experiment 3 or in Experiment 7, and this is in seeming contrast to what Balota and Chumbley (1985) reported. By using cue delays between 150 and 1,400 ms, Balota and Chumbley observed a substantial frequency effect of 44 ms (at the 150-ms delay) that dropped to 18 ms at the longest delay. In their Experiment 1, in which cue delays had been blocked, significant effects were still obtained at the longest delays. In their Experiment 2, in which cue delays varied from trial to trial, no reliable effect was obtained at delays longer than 900 ms. Our experiments differ from the Balota and Chumbley research in two important aspects: one concerning properties of the materials and one concerning the range of cue delays used.

Although the frequency contrast used by Balota and Chumbley (1985) is comparable to the one in the materials for our Experiments 1–5, our words were considerably shorter, which makes a direct comparison difficult. Our items were almost exclusively monosyllabic words, whereas in Balota and Chumbley's research the majority of items were multisyllabic. A more appropriate comparison is an unpublished study by Balota and Shields (1988) in which only monosyllabic words were tested. They also used a frequency contrast comparable to the one we used. Cue delays ranged from 100 ms to 1,300 ms.

For these materials, Balota and Shields found a small but significant effect of 6 ms, which did not interact with cue delay. Although significant, this effect was substantially smaller than the effect reported in the original (Balota and Chumbley) research and was of the same order of magnitude as our (insignificant) 7-ms effect in Experiment 3.

A notable difference between Balota and Chumbley's (1985) and Balota and Shield's (1988) research and our Experiments 3 and 7 concerns the range of cue delays tested. There is evidence that this is a critical variable (cf. Monsell et al., 1989; Savage et al., 1990). In an experiment using Balota and Chumbley's (1985) original materials, Savage et al. (1990, Experiment 3) failed to replicate the effect when delays were restricted to the range from 800 to 1,200 ms. Only after having included shorter delays (150, 400, and 650 ms) in filler trials, a marginally significant 14-ms frequency effect at the 900-ms delay was observed, and none at longer delays (Savage et al., 1990, Experiment 4). It seems that the frequency effect in delayed naming crucially depends on the presence of short delays. However, as Savage et al. pointed out, any effect at short delays is irrelevant for the issue at hand. At those delays participants might still be engaged in preparing the utterance, or the early cue might interrupt the response preparation. Only effects at long delays can provide information about the contribution of articulatory processes. Neither the original research by Balota and Chumbley nor the research by Savage et al. yielded an effect of frequency at delays over 1,000 ms. Likewise, Forster and Chambers (1973) observed a nonsignificant 4-ms difference with delays of 2,000 ms. Our Experiments 3 and 7 exclusively used long (1,000 to 1,600 ms) delays. This might explain why we did not observe a reliable effect. However, as Savage et al. (1991) argued, there are good reasons for the claim that any genuinely articulatory contribution should surface at these longer delays. The delays in our Experiments 3 and 7 were long, but we did not find such an effect. This further supports our conclusion that the word frequency effect obtained in Experiment 1 is lexical in nature.

Our subsequent experiments tested whether the locus of the frequency effect is to be sought in the activation threshold of lemmas, in the lemma–lexeme connection strength, or in the activation threshold of lexemes. Experiments 4, 5a, and 5b involved a gender decision task. In our theoretical model, gender is a property of lemmas; the gender node receives its activation from the lemma node. If, as Dell (1990) suggested, the locus of the robust word frequency effect is the lemma threshold activation, we should find a robust frequency effect in gender decision. That did not happen. We did find an effect, but it disappeared after the first trial for an item. We called this the *ephemeral effect*. It cannot be accounted for by a frequency-dependent lemma threshold. Rather, we interpreted it as a recency effect (and we return to that interpretation shortly).

This left us with two further loci to explore: the connection from lemma to lexeme (the out-of-lemma hypothesis) and the lexeme threshold (the lexeme hypothesis). These loci were contrasted in Experiment 6, in which we measured the speed of producing homophones. The out-of-lemma hypothesis (and the lemma threshold hypothesis, for that matter) predicts that the speed is determined by lemma frequencies. The lexeme hypothesis predicts that the speed is determined by lexeme

frequencies. The translation experiment decisively supports the latter view. Experiment 7 excluded an explanation of this result in terms of initiation of articulation.

Our tentative explanation for the ephemeral frequency effect was that the lemma-to-gender connection strength is facilitated on use, after which it slowly decays. However, there is an alternative account that needs some consideration. During the first encounter of a picture, participants retrieve lemma and gender and respond on the basis of this information. At the same time, an episodic memory trace is established. During subsequent encounters of that particular picture, the retrieval of the episodic memory competes with the lexical gender retrieval. Episodic memory retrieval may be faster for LF words than for HF words, and it may even be faster for both; this reduces the difference in performance on HF and LF words. According to this account, the disappearing frequency effect in gender decision does not reflect changes in the accessibility of a word's lexical representation but a task-specific artifact of episodic memory, recalling a previous decision on the same item. Gender decision would not be the "royal road" to the lemma but just some other kind of task, in which an episodic memory trace established during an earlier response speeds up the current response.

Some theorists have indeed proposed episodic accounts of repetition priming effects in lexical tasks (cf. Feustel, Shiffrin, & Salasoo, 1983; Jacoby, 1983). The basic idea is that on discovering the resemblance between the present stimulus and a previous stimulus–response pairing, participants based their response on the memory for the earlier response and not on lexical information. Such an account is plausible but by no means necessarily correct in general. For example, Mitchell and Brown (1988) found that repetition priming of picture naming was not affected by delay, whereas episodic recognition was affected. Also, priming occurred regardless of successful recognition. Both findings are unexpected under the episodic memory account and are fully compatible with a lexical processing account. Graf and Mandler (1984) provided some evidence that although participants can use episodic memory traces, they do not do so spontaneously.

There are, however, conditions in which a contribution of episodic memory to a participant's performance in lexical tasks would be expected. If the task is relatively complicated and the trace of a previous response is readily available (i.e., if the lags between repeated encounters are short and only relatively few items are involved), episodic memory may take over (see Wheeldon & Monsell, 1992, for a detailed discussion of episodic retrieval accounts). It is, however, unlikely that these conditions were met in our experiments. First, the experiments involved about 100 items (including filler items and practice items), which makes it hard to imagine that participants would easily remember the response associated with a particular item. Second, repeated presentations of an individual item were separated by at least 20 intervening items (and often substantially more), creating long lags. Third, performance in gender decision, although more demanding than naming, was still relatively efficient and fast. Gender decision responses on the first presentation in Experiment 4 were well under 1 s, even for the LF items. Retrieval of an episodic trace would have to be much faster to beat the lexically mediated response retrieval. These considerations make an episodic memory ac-

count of our findings unlikely. It is, however, practically ruled out by a recent additional finding. At the Max Planck Institute, van Berkum (personal communication, June 1993) ran a replication of our Experiments 5a and 5b by using exactly the same procedure and materials, with just one slight modification. He asked participants to produce a noun phrase of the form adjective plus noun in the naming trials. This utterance format requires a gender marking suffix on the adjective. For example, participants had to produce *klein-e_{masc or fem} hond* [small dog] but *klein huis* [small house]. Hence, to generate the correct form of the noun phrase, participants had to retrieve the picture name's gender. van Berkum found a reliable effect of frequency in noun phrase naming, but did not find any effect in subsequent gender decision (*de* vs. *het*), just like in our Experiment 5b. This was true even though the decision had to be made on a different gender marking morpheme, and gender had surfaced in a subtle form in the naming trials. Episodic memory cannot have been involved because the *de* and *het* in gender decision had not been part of the naming response. This finding further corroborates our assumption that the diminished difference in performance on items of different frequency indeed reflects changes in the accessibility of a word's gender. The lemma-to-gender connection strength is facilitated on use and slowly decays thereafter.

What could be the functional sense of such a mechanism? It may play a role in the normal production of spoken discourse. A new entity is usually introduced with a full indefinite nominal phrase (e.g., I saw a big horse). In many languages, the indefinite article, the adjective, or both reflect the noun's gender. However, maintaining reference to the same entity is typically done by anaphoric means (e.g., it crossed the road). In Dutch and many other languages, pronominal anaphors are gender marked. This means that the speaker has to reaccess the gender information of a recently accessed word to produce the appropriate anaphor. However, it is not necessary for the speaker to reactivate the original word form because there is no reuse of that word form. Hence, it suffices for the speaker to access the lemma only and through it the gender information. The function of the recency effect would then be to facilitate anaphoric reference to recently introduced discourse entities, therefore contributing to the fluency of the utterance. It is obvious that this interpretation needs further exploration. If it is correct, however, it implies that lemmas and their lexemes may have different accessing frequencies, even for nonhomophones.

Why did we observe the robust word frequency effect in the *Naming* section of Experiment 5b (i.e., when speakers produced full noun phrases)? According to one account, we should not have obtained it, at least not for an item's second naming trial. On that trial, the lemma-to-gender link is in a state of facilitation. The speaker can easily reaccess the gender information and produce the appropriate article; this is a frequency-independent process. While that process (including the articulation of the article) is going on, the speaker accesses the head word's lexeme. Although the speed of the latter process is frequency dependent, it will not show up in the voice-key data. This is because the speaker could initiate the response as soon as the article was available.

The data tell us otherwise, however. We did find a robust frequency effect in the naming part of Experiment 5b. Clearly,

the production of the article is not initiated before the form of the head noun has become available. This makes sense. The speaker does not produce two independent words (article and noun) but one *phonological word* (see Levelt, 1989, p. 373 ff.), a blended pattern of the two word shapes. The phonological word is the domain of syllabification. It can, in fact, happen that syllabification straddles the lexical boundary between article and noun (as in Dutch *het eten* → *hə-te-tə* [the meal]; see Levelt, 1993, for a further discussion of the process of syllabification during speech production). This means, however, that the article cannot be adequately produced without having retrieved the head noun's word form; the earlier word has to wait for the retrieval of the later word. Recently, Schriefers (1992) presented empirical evidence for such dependencies in the production of noun phrases.

A further issue concerning the robust frequency effect is the precise interpretation of its lexemic locus. We explored the possibility of lexical density. Goldinger and van Summers (1989) found that the density of a word's lexical environment has some affect on the word's acoustic-phonetic realization. Is naming latency also related to lexical density, that is, to the number of words that are phonologically similar to the target word? Can this account for the frequency effect? We performed various analyses on the lexical environment of the words used in our research. Measures of lexical density were derived on the basis of both the N definition (cf. Coltheart, Davelaar, Jonasson, & Besner, 1977) and the cohort definition (Marslen-Wilson & Welsh, 1978). Subsequent stepwise regression analyses revealed that none of the measures contributed to naming latency when object recognition latency and \log_{10} frequency were taken into account. Also, correlations between lexical density and word frequency were virtually nonexistent (for details see Jescheniak, 1993b). This suggests, then, that properties of a word form's lexical environment cannot account for the effect in naming. This further supports the conclusion that frequency-dependent accessibility in speech production is an inherent property of the lexical item itself. Our research supports the notion that this property is the item's lexeme activation threshold.

References

- Balota, D. A., & Chumbley, J. I. (1985). The locus of word-frequency effects in the pronunciation task: Lexical access and/or production? *Journal of Memory and Language*, 24, 89–106.
- Balota, D. A., & Shields, L. (1988, November). *The locus of word frequency effects in pronunciation*. Paper presented at the meeting of the Psychonomic Society, Chicago.
- Bartram, D. J. (1973). Effects of familiarity and practice on naming pictures of objects. *Memory & Cognition*, 1, 101–105.
- Bartram, D. J. (1974). The role of visual and semantic codes in object naming. *Cognitive Psychology*, 6, 325–356.
- Bartram, D. J. (1976). Levels of coding in picture-picture comparison tasks. *Memory & Cognition*, 4, 593–602.
- Beattie, G., & Butterworth, B. (1979). Contextual probability and word frequency as determinants of pauses in spontaneous speech. *Language and Speech*, 22, 201–211.
- Bock, J. K. (1982). Toward a cognitive psychology of syntax. *Psychological Review*, 89, 1–47.
- Bock, J. K., & Levelt, W. J. M. (in press). Language production: Grammatical encoding. In M. A. Gernsbacher (Ed.), *Handbook of psycholinguistics*. San Diego, CA: Academic Press.

- Brown, A. S. (1991). A review of the tip-of-the-tongue experience. *Psychological Bulletin, 109*, 204–223.
- Brown, R., & McNeill, D. (1966). The “tip-of-the-tongue” phenomenon. *Journal of Verbal Learning and Verbal Behavior, 5*, 325–337.
- Butterworth, B. (1980a). Some constraints on models of language production. In B. Butterworth (Ed.), *Language production: Vol. 1. Speech and talk* (pp. 423–459). San Diego, CA: Academic Press.
- Butterworth, B. (1980b). Evidences from pauses in speech. In B. Butterworth (Ed.), *Language production: Vol. 1. Speech and talk* (pp. 155–176). San Diego, CA: Academic Press.
- Butterworth, B. (1989). Lexical access in speech production. In W. Marslen-Wilson (Ed.), *Lexical representation and process* (pp. 108–135). Cambridge, MA: MIT Press.
- Clark, H. (1973). The language-as-fixed-effect fallacy: A critique of language statistics in psychological research. *Journal of Verbal Learning and Verbal Behavior, 12*, 335–359.
- Collins, A. M., & Loftus, E. F. (1975). A spreading-activation theory of semantic processing. *Psychological Review, 82*, 407–428.
- Coltheart, M., Davelaar, E., Jonasson, J. T., & Besner, D. (1977). Access to the internal lexicon. In S. Dornic (Ed.), *Attention and performance VI* (pp. 535–555). Hillsdale, NJ: Erlbaum.
- Dell, G. S. (1986). A spreading activation model of retrieval in sentence production. *Psychological Review, 93*, 283–321.
- Dell, G. S. (1988). The retrieval of phonological forms in production: Tests of predictions from a connectionist model. *Journal of Memory and Language, 27*, 124–142.
- Dell, G. S. (1990). Effects of frequency and vocabulary type on phonological speech errors. *Language and Cognitive Processes, 5*, 313–349.
- Dell, G. S., & O’Seaghdha, P. G. (1991). Mediated and convergent lexical priming in language production: A comment on Levelt et al. (1991). *Psychological Review, 98*, 604–614.
- Dell, G. S., & O’Seaghdha, P. G. (1992). Stages of lexical access in language production. *Cognition, 42*, 287–314.
- Dell, G. S., & Reich, P. A. (1981). Stages in sentence production: An analysis of speech error data. *Journal of Verbal Learning and Verbal Behavior, 20*, 611–629.
- del Viso, S., Igoa, J. M., & Garcia-Albea, J. E. (1991). On the autonomy of phonological encoding: Evidence from slips of the tongue in Spanish. *Journal of Psycholinguistic Research, 20*, 161–185.
- Feustel, T. C., Shiffrin, R. M., & Salasoo, A. (1983). Episodic and lexical contributions to the repetition effect in word identification. *Journal of Experimental Psychology: General, 112*, 309–346.
- Forster, K. I., & Chambers, S. M. (1973). Lexical access and naming time. *Journal of Verbal Learning and Verbal Behavior, 12*, 627–635.
- Fromkin, V. A. (1971). The non-anomalous nature of anomalous utterances. *Language, 47*, 27–52.
- Garrett, M. F. (1975). The analysis of sentence production. In G. Bower (Ed.), *Psychology of learning and motivation: Vol. 9* (pp. 133–177). San Diego, CA: Academic Press.
- Garrett, M. F. (1976). Syntactic processing in sentence production. In E. Walker & R. Wales (Eds.), *New approaches to language mechanisms* (pp. 231–256). Amsterdam: North-Holland.
- Garrett, M. F. (1980). Levels of processing in sentence production. In B. Butterworth (Ed.), *Language production: Vol. 1. Speech and talk* (pp. 177–220). San Diego, CA: Academic Press.
- Garrett, M. F. (1982). Production of speech: Observations from normal and pathological language use. In A. W. Ellis (Ed.), *Normality and pathology in cognitive functions* (pp. 19–76). San Diego, CA: Academic Press.
- Garrett, M. F. (1988). Processes in language production. In F. J. Newmeyer (Ed.), *The Cambridge Survey of Linguistics: Vol. 3. Language: Psychological and biological aspects* (pp. 69–96). Cambridge, MA: Harvard University Press.
- Golding, S. D., & van Summers, W. (1989). *Lexical neighborhood in speech production: A first report*. Research on Speech Production (Progress Report No. 15). Bloomington: Indiana University, Department of Psychology.
- Graf, P., & Mandler, G. (1984). Activation makes words more accessible, but not necessarily more retrievable. *Journal of Verbal Learning and Verbal Behavior, 23*, 553–568.
- Harley, T. A. (1993). Phonological activation of semantic competitors during lexical access in speech production. *Language and Cognitive Processes, 8*, 291–310.
- Henaff Gonon, M., Bruckert, R., & Michel, F. (1989). Lexicalization in an anomic patient. *Neuropsychologia, 27*, 391–407.
- Hotopf, W. H. N. (1980). Semantic similarity as a factor in whole-word slips of the tongue. In V. A. Fromkin (Ed.), *Errors in linguistic performance* (pp. 97–109). San Diego, CA: Academic Press.
- Humphreys, G. W., Riddoch, M. J., & Quinlan, P. T. (1988). Cascade processes in picture identification. *Cognitive Neuropsychology, 5*, 67–103.
- Huttenlocher, J., & Kubicek, L. F. (1983). The source of relatedness effects on naming latency. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 9*, 486–496.
- Jacoby, L. L. (1983). Perceptual enhancement: Persistent effects of an experience. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 9*, 21–38.
- Jescheniak, J. D. (1993a). *Semantic interference in sequential picture naming*. Unpublished manuscript.
- Jescheniak, J. D. (1993b). *Word frequency effects in speech production*. Unpublished doctoral dissertation, University of Nijmegen, Nijmegen, The Netherlands.
- Kelly, M. H. (1986). *On the selection of linguistic options*. Unpublished doctoral dissertation. Cornell University, Ithaca, NY.
- Kempen, G., & Huijbers, P. (1983). The lexicalization process in sentence production and naming: Indirect election of words. *Cognition, 14*, 185–209.
- Kirsner, K., Smith, M. C., Lockhart, R. S., King, M. L., & Jain, M. (1984). The bilingual lexicon: Language-specific units in an integrated network. *Journal of Verbal Learning and Verbal Behavior, 23*, 519–539.
- Kroll, J. F., & Curley, J. (1988). Lexical memory in novice bilinguals: The role of concepts in retrieving second language words. In M. Gruneberg, P. Morris, & R. Sykes (Eds.), *Practical aspects of memory: Vol. 2* (pp. 389–395). New York: Wiley.
- Kroll, J. F., & Potter, M. C. (1984). Recognizing words, pictures, and concepts: A comparison of lexical, object, and reality decisions. *Journal of Verbal Learning and Verbal Behavior, 23*, 39–66.
- Levelt, W. J. M. (1983). Monitoring and self-repair in speech. *Cognition, 14*, 41–104.
- Levelt, W. J. M. (1989). *Speaking: From intention to articulation*. Cambridge, MA: MIT Press.
- Levelt, W. J. M. (1992). Accessing words in speech production: Stages, processes and representations. *Cognition, 42*, 1–22.
- Levelt, W. J. M. (1993). Timing in speech production: With special reference to word form encoding. *Annals of the New York Academy of Sciences, 682*, 283–295.
- Levelt, W. J. M., Schriefers, H., Vorberg, D., Meyer, A. S., Pechmann, T., & Havinga, J. (1991a). The time course of lexical access in speech production: A study of picture naming. *Psychological Review, 98*, 122–142.
- Levelt, W. J. M., Schriefers, H., Vorberg, D., Meyer, A. S., Pechmann, T., & Havinga, J. (1991b). Normal and deviant lexical processing: Reply to Dell and O’Seaghdha (1991). *Psychological Review, 98*, 615–618.
- Maclay, H., & Osgood, C. E. (1959). Hesitation phenomena in spontaneous English speech. *Word, 15*, 19–44.
- Marslen-Wilson, W. D., & Welsh, A. (1978). Processing interactions and lexical access during word recognition in continuous speech. *Cognitive Psychology, 10*, 29–63.

- Martin, J. G., & Strange, W. (1968). The perception of hesitation in spontaneous speech. *Perception & Psychophysics*, 3, 427-432.
- McRae, K., Jared, D., & Seidenberg, M. S. (1990). On the roles of frequency and lexical access in naming. *Journal of Memory and Language*, 29, 687-715.
- Mitchell, D. B., & Brown, A. S. (1988). Persistent repetition in picture naming and its dissociation from recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 14, 213-222.
- Monsell, S., Doyle, M. C., & Haggard, P. N. (1989). Effects of frequency on visual word recognition tasks: Where are they? *Journal of Experimental Psychology: General*, 118, 43-71.
- Oldfield, R. C., & Wingfield, A. (1965). Response latencies in naming objects. *Quarterly Journal of Experimental Psychology*, 17, 273-281.
- Pechmann, T., Reetz, H., & Zerbst, D. (1989). Kritik einer Meßmethode: Zur Ungenauigkeit von voice-key Messungen [A critical assessment of the reliability of voice-key measurements]. *Sprache & Kognition*, 8, 65-71.
- Potter, M. C., So, K.-F., von Eckardt, B., & Feldman, L. B. (1984). Lexical and conceptual representation in beginning and proficient bilinguals. *Journal of Verbal Learning and Verbal Behavior*, 23, 23-38.
- Roelofs, A. (1992). A spreading-activation theory of lemma retrieval in speaking. *Cognition*, 42, 107-142.
- Savage, G. R., Bradley, D. C., & Forster, K. I. (1990). Word frequency and the pronunciation task: The contribution of articulatory fluency. *Language and Cognitive Processes*, 5, 203-236.
- Schriefers, H. (1992). Lexical access in the production of noun phrases. *Cognition*, 45, 33-54.
- Schriefers, H., Meyer, A. S., & Levelt, W. J. M. (1990). Exploring the time course of lexical access in production: Picture-word interference studies. *Journal of Memory and Language*, 29, 86-102.
- Snodgrass, J. G., & Vanderwart, M. (1980). A standardized set of 260 pictures: Norms for name agreement, image agreement, familiarity, and visual complexity. *Journal of Experimental Psychology: Human Learning and Memory*, 6, 174-215.
- Stemberger, J. P. (1984). Structural errors in normal and agrammatic speech. *Cognitive Neuropsychology*, 1, 281-313.
- Stemberger, J. P., & McWhinney, B. (1986). Frequency and the lexical storage of regularly inflected forms. *Memory & Cognition*, 14, 17-26.
- Wheeldon, L. R., & Monsell, S. (1992). The locus of repetition priming of spoken word production. *Quarterly Journal of Experimental Psychology*, 44A, 723-761.
- Winer, (1971). *Statistical principles in experimental design*. New York: McGraw-Hill.
- Wingfield, A. (1967). Perceptual and response hierarchies in object identification. *Acta Psychologica*, 26, 216-226.
- Wingfield, A. (1968). Effects of frequency on identification and naming of objects. *American Journal of Psychology*, 81, 226-234.

Appendix A

List of Picture Names Used in Experiments 1-5

The approximate English translation is given in brackets.

LF Names

bezem [broom], bijl [axe], fluit [flute], hark [rake], harp [harp], kam [comb], kano [canoe], krab [crab], pauw [peacock], peer [pear], rups [caterpillar], schaar [scissors], slak [snail], sleet [sledge], snavel [beak], spin [spider], step [scooter], tang [tongs], tol [top], uil [owl], vaas [vase], worst [sausage], zaag [saw], zwaan [swan]

HF Names

arm [arm], auto [car], bank [sofa], bloem [flower], boom [tree], boot [boat], brief [letter], broek [trousers], deur [door], fles [bottle], hond [dog], kerk [church], mond [mouth], muur [wall], neus [nose], schoen [shoe], ster [star], stoel [chair], tafel [table], trap [steps], vinger [finger], vis [fish], voet [foot], zak [bag]

Appendix B

List of Dutch Target Words and the English Probe Words (in Brackets) by Experimental Condition for Experiment 6

Homophones

blik [tin], bos [bunch], dom [cathedral], echt [matrimony], haast [hurry], weer [weather], zeer [ache], steeds [urban], stof [dust], leer [leather], laag [stratum]

LF Controls

breuk [fracture], hok [kennel], pauk [kettledrum], erwt [pea], haard [fireplace], wieg [cradle], naad [seam], stipt [punctual], stoom [steam], waan [delusion], loon [wage]

HF Controls

beeld [image], hoek [corner], boos [angry], hoog [high], thans [now], daar [there], reeds [already], nooit [never], stoel [chair], geur [odor], bang [afraid]

Received March 12, 1993
Revision received September 24, 1993
Accepted October 4, 1993 ■