ELSEVIER

Cognition 64 (1997) 115–152

COGNITION

Syntactic processes in speech production: the retrieval of grammatical gender

Jos J.A. van Berkum*

Max Planck Institute for Psycholinguistics, Postbus 310, 6500 AH Nijmegen, The Netherlands

Received 6 November 1996, final version 1 July 1997; accepted 1 July 1997

Abstract

Jescheniak and Levelt (Jescheniak, J.-D., Levelt, W.J.M. 1994. Journal of Experimental Psychology: Learning, Memory and Cognition 20 (4), 824–843) have suggested that the speed with which native speakers of a gender-marking language retrieve the grammatical gender of a noun from their mental lexicon may depend on the recency of earlier access to that same noun's gender, as the result of a mechanism that is dedicated to facilitate gender-marked anaphoric reference to recently introduced discourse entities. This hypothesis was tested in two picture naming experiments. Recent gender access did not facilitate the production of gender-marked adjective noun phrases (Experiment 1), nor that of gender-marked definite article noun phrases (Experiment 2), even though naming times for the latter utterances were sensitive to the gender of a written distractor word superimposed on the picture to be named. This last result replicates and extends earlier gender-specific picture-word interference results (Schriefers, H. 1993. Journal of Experimental Psychology: Learning, Memory, and Cognition 19 (4), 841–850), showing that one can selectively tap into the production of grammatical gender agreement during speaking. The findings are relevant to theories of speech production and the representation of grammatical gender for that process. © 1997 Elsevier Science B.V.

Keywords: Speech production; Grammatical gender agreement; Anaphoric reference; Lexical memory; Picture-word interference

1. Introduction

By and large, the words we use are those that convey the things we want to say. To express the act of walking, most speakers of English will use the verb 'walk', rather than, say 'talk'. In many languages, however, the *exact* form of, say, a verb, or an adjective, does not just depend on the concept to be expressed, but also frequently

* Tel.: +31 24 3521911; fax: +31 24 3521213; e-mail: berkum@mpi.nl

PII S0010-0277(97)00026-7

covaries with the nature of some nearby linguistic constituent. In English, for example, verbs agree with their subject in person, as in 'I walk' but 'he walk*s*', and in number, as in 'people talk' but 'the neighbor talk*s*'. In Russian, adjectives must agree with the nouns they modify, both in number, as in 'bol's-*oj* avtomobil', 'bol's-*ie* avtomobili' ('a large car', 'large cars'), and in gender, as in 'bol's-*oj* avtomobil', 'bol's-*oe* taksi' ('a $large_{masc}$ car_{masc}', 'a $large_{neu}$ $taxi_{neu}$'; Corbett, 1994). And in Swahili, virtually everything agrees with the gender of a nearby noun, as in '*ki*-kapu *ki*-kubwa *ki*-moja *ki*-lianguka' ('$basket_{gender\text{-}7}$ $large_{gender\text{-}7}$ $one_{gender\text{-}7}$ $fell_{gender\text{-}7}$', 'one large basket fell'; Corbett, 1991)

Such agreement phenomena have been estimated to occur in approximately 75% of the world's languages (Mallinson and Blake, 1981). But only fairly recently have psycholinguists begun to explore how the native speakers of all these languages actually produce agreement, as they speak (e.g., Lapointe and Dell, 1989; Bock and Miller, 1991; Bock and Eberhard, 1993; Nicol, 1995; Vigliocco et al., 1995). Much of this work has been conducted in English, and has therefore mainly explored how speakers produce number agreement between a subject and a verb. But *number* is just one of several features in which words can agree. In a large number of languages, words also agree in grammatical *gender* (Corbett, 1991).

In view of the similarities between the phenomena of gender and number agreement, it is not unlikely that speakers will to some extent use the same mechanisms to produce both as they speak. But there is an important difference in where the agreement information originally comes from. Number is typically a part of the conceptual message that the speaker is trying to convey, e.g., whether it is just one neighbor that talks, or several. Grammatical gender, however, is an *inherent property of a noun*, somehow encoded as part of the native speaker's knowledge of the nouns in his or her language. In the course of speaking, therefore, this information will have to be retrieved from the speaker's memory store for words, the mental lexicon. The aim of the studies reported here is to find out how speakers do this.

Like closely related German, the Dutch language requires its speakers to get their genders right. A few exceptions aside, every Dutch noun has one of two possible genders, which, amongst other things, makes a difference for the singular definite article. A *common* gender noun (historically either masculine or feminine) only goes with 'de', as in '*de* ster' ('the_{com} $star_{com}$'), and is for that reason often referred to as a *de-word*. A *neuter* gender noun, on the other hand, only goes with 'het', as in '*het* huis' ('the_{neu} $house_{neu}$'); it is therefore often called a *het-word*. Apart from the definite article, several other determiners also agree with their singular head noun in gender, as in '*deze* ster' versus '*dit* huis' ('$this_{com}$ star', '$this_{neu}$ house'). So does the relative pronoun, as in 'de ster *die*...' versus 'het huis *dat*...' ('the star $that_{com}$...', 'the house $that_{neu}$...'), and most adnominal adjectives, as in 'een groen*e* ster' versus 'een groen huis' ('a $green_{com}$ star', 'a $green_{neu}$ house').

The grammatical gender of a Dutch noun thus determines the exact form of a variety of nearby constituents, the so-called agreement targets (see van van Berkum, 1996 for a more extensive overview of the Dutch gender system). This indicates that, to get those forms right, a speaker of this language needs to retrieve gender information from the mental lexicon quite regularly, somewhere around once every 10 s at least, according to a corpus-based estimate[1]. And this is just for Dutch, a language with a relatively moderate degree of gender-marking limited to singular nouns; many languages have a much more extensive gender system (Corbett, 1991).

The frequency with which gender must be retrieved from the mental lexicon clearly imposes a considerable real-time demand on a speaker: not only must he or she frequently recover a noun's gender, but this must be done in time, often before the noun itself has been used, and early enough to have the appropriate word forms ready for uninterrupted, fluent speaking. Non-native speakers of gender languages will readily appreciate this demand. But most native speakers will hardly be aware that it exists at all; to them, gender agreement usually comes for free. How do these speakers retrieve grammatical gender information from memory as they speak, such that their fluency is preserved?

The answer to this need not be specific to gender in any interesting way. That is, the mechanisms that subserve the retrieval of a word's grammatical gender may just be those that also underlie retrieval of, say, a word's sound form, retrieval of the fact that a robin is a bird, or retrieval of some autobiographical memory. A more interesting possibility, though, is that elements of the architecture involved in gender retrieval are somehow tailored to the particulars of gender-marked language use, perhaps due to the evolution of language-specific machinery in the human species (Fodor, 1983; Bickerton, 1984; Pinker and Bloom, 1990; Pinker, 1994).

Jescheniak and Levelt (1994) have recently suggested that the mechanism for gender retrieval may indeed be shaped to meet a very specific demand of speech production. In one of their studies, they observed that speakers of Dutch were faster to retrieve the gender of a particular noun, like 'ster' ('star'), if they had, some 20 min before, used that noun in a gender-marked speech utterance, as in /*de* ster/, than if they had merely produced the noun itself, e.g., /ster/, before[2]. Furthermore, the benefit that speakers obtained from having retrieved a particular noun's gender a little while back was much larger if the word at hand had a relatively low average frequency of use in the language, e.g., 'slak' ('snail'), than if it had a relatively high frequency of use, e.g., 'ster' ('star'). This differential benefit of recent gender access actually eliminated a gender retrieval advantage for high-frequent words: Whereas speakers were faster to retrieve the gender of a high-frequent word like 'ster' than of a low-frequent word such as 'slak' if, in the experiment, they did so for the first time, the difference disappeared when the gender of both words had already been retrieved before.

Jescheniak and Levelt tentatively explained these results as a *gender recency effect*: the time it takes to retrieve a noun's gender from the mental lexicon would

[1]In a 30 000 word randomly drawn subsample of the *Eindhoven* corpus of spontaneous spoken Dutch (Uit den Boogaart, 1975), 1369 noun tokens occurred in the context of a gender-marked determiner, adjective, relative pronoun, or any combination of these. This suggests that gender must on average be retrieved every 22nd word *at least* (a combination of *N* gender-marked constituents all agreeing with some single noun token could also be interpreted to involve *N* gender retrievals rather than just one). Given an average speaking rate of 2–3 words/s (Levelt, 1989), this in turn suggests that native speakers of Dutch need to retrieve gender from their lexicon every 7–11 s at least.

[2]In the text, spoken responses will be denoted by slashes; linguistic stimuli as such will be in single quotes.

be determined by how long it has been since the last time the same noun's gender was retrieved. This not only explained why more recent access led to faster re-access, but also why equating low- and high-frequent words on recency of access to their gender information seemed to equate those words on the speed of later re-access to that information.

But why would such a recency effect exist? Using a recent network model of the mental lexicon (Levelt, 1989; Roelofs, 1992) as their theoretical framework, Jescheniak and Levelt suggested that the effect at hand might reflect a unique architectural property of the lexical memory that was used to represent the gender of a word: a recency-sensitive link from the representation of an individual word as a lexical-syntactic entity (often called a *lemma*; after Kempen and Huijbers, 1983) to a representation of the appropriate grammatical gender (e.g., a neuter gender node, cf. Jescheniak and Levelt, 1994, Fig. 1). Interestingly, they also speculated that this architectural property might exist because it would be of help in the normal production of spoken discourse, in particular when speakers need grammatical gender in order to refer back to something they have already talked about before.

> 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 .... 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.
> (Jescheniak and Levelt, 1994, p. 841)

Thus, speakers might benefit from, to use a computer systems analogy, a gender-dedicated 'disk caching mechanism', a memory device that would keep the grammatical gender information retrieved from long term storage in a temporarily much more accessible state, just in case the speaker would need to re-access that same noun's gender shortly afterwards.

The two experiments reported in this paper evaluate this hypothesis. Although in itself it involves a fairly constrained phenomenon, Jescheniak's and Levelt's suggestion has a potentially wider scope than one would at first suspect. First of all, much of the chronometric research in speech production has, for obvious good reasons, focused on the production of single words in isolation. By doing so, however, it has inevitably ignored an essential feature of the speaker's task, i.e., the coordination of words into phrases, sentences, and discourse. Gender agreement is a relatively tractable word coordination phenomenon, and knowledge about a dedicated mechanism for repeated gender retrieval could therefore help us to begin to understand what is involved in the production of more complex utterances (see, for example, Schriefers, 1993). Also, whereas there is a large body of research on speech errors and hesitations, we know very little about the factors that make a positive contribution to fluency (Bock, 1995). Evidence for a dedicated gender retrieval mechanism can help us understand how native speakers so easily generate speech at an average 2–3 words/s (Levelt, 1989).

Such evidence would also reveal a considerable degree of adaptation of memory architecture to the demands of speech production. This is non-trivial, because (as already hinted at above) there is a plausible alternative: gender recency effects could simply reflect the operation of a much more general mechanism, shaped by the demands of memory retrieval in a *variety* of cognitive domains (cf. Anderson, 1993; Newell and Rosenbloom, 1981) rather than by specific demands in the domain of gender-marked speech production[3].

In addition, Jescheniak's and Levelt's theoretical account includes a very explicit hypothesis about how native speakers represent their knowledge of grammatical gender in memory: the gender of a word would be represented by a (recency-sensitive) link from the lexical-syntactic representation of that word, its lemma, to a *generic* gender representation. In the lexical network model at hand, the lemma node of a neuter Dutch noun like 'huis' ('het huis', 'dit huis', 'een huis dat...'; 'the_{neu} house', 'this_{neu} house', 'a house that_{neu}...') would for example be linked to a single neuter gender node, which would in turn provide access to the range of specific agreement targets for this gender ('het', 'dit', 'dat', etc.). All other neuter noun lemmas would be connected to the same neuter gender node (each with their own uniquely recency-sensitive link, though). The lemma nodes of common gender nouns like 'ster' ('de ster', 'deze ster', 'een ster die...'; 'the_{com} star', 'this_{com} star', 'a star that_{com}...') would instead be linked to a single common gender node providing access to the specific agreement targets for that particular gender ('de', 'deze', 'die', etc.).

In view of the apparent lack of systematicity in why certain nouns have certain genders (Germans use 'der Löffel', 'die Gabel', and 'das Messer' for a spoon, fork, and knife, respectively), many language researchers are convinced that native speakers simply memorize, i.e. *store*, grammatical gender noun by noun, perhaps indeed along the lines proposed by Jescheniak and Levelt. But systematicity in gender assignment does exist, and it has led several leading gender theorists (Zubin and Köpcke, 1981; Corbett, 1991, 1994) to suggest that native speakers actually avoid explicit memorization, and rapidly *compute* the gender of a noun from its meaning or form instead. Although some intriguing findings with anomic patients appear to go against this view (especially Badecker et al., 1995; see also Henaff Gonon et al., 1989), there is as yet no substantial evidence from normal speakers that directly bears on the representation of grammatical gender in memory (but see Vigliocco et al., 1997). Evidence that gender is stored along the lines suggested by Jescheniak and Levelt would therefore be of considerable interest.

The aim of the two experiments reported below is to study how native speakers retrieve grammatical gender from the mental lexicon as they speak. Does recent access to a noun's gender during speech production facilitate later re-access to that same noun's gender, in a way that suggests dedicated retrieval mechanisms for the task of speaking? In the critical study (Jescheniak and Levelt, 1994; Experiment 5), recency of gender access was manipulated in a speech production task

[3]It should be said here that, although Jescheniak and Levelt tentatively committed themselves to a *dedicated* mechanism, they were silent on the issue of whether this mechanism would be part of our innate machinery for language or instead develop as a function of language experience.

by having participants name simple pictures, either with bare nouns, (e.g., /ster/, /huis/, /slak/, ...), or with definite article noun phrases (e.g., /de ster/, /het huis/, /de slak/, ...). But the speed with which these participants could subsequently (re)access the gender of these nouns was assessed, not in a second speech production task, but in a so-called *gender decision task*. Here, participants were shown the same pictures, but now asked to rapidly decide upon the grammatical gender of every picture name by pushing one of two response buttons labeled 'DE' and 'HET'. This binary decision task was assumed to tap into the initial (but not the later) processes involved in gender-marked speech production, i.e., identifying a concept, selecting the appropriate lemma to name it with, and retrieval of its grammatical gender.

Two further experiments with the same naming-plus-gender-decision paradigm have yielded conflicting results, however. The first of these studies (van van Berkum, 1996 Chp. 4, Experiment 1) extended the findings of Jescheniak and Levelt, showing that people not only made faster DE-HET gender decisions after having produced the nouns at hand with the same gender-marked definite article, but also if they had produced those nouns with a gender-marked *adjective* (e.g., /groene ster/, / groen huis/, /groene slak/, ...). This seemed to corroborate Jescheniak's and Levelt's assumption that (access to) the representation of gender is generic, i.e., independent of the particular agreement targets that realize it. Unfortunately, these results could not be replicated in a larger second study (van van Berkum, 1996; Chp. 4, Experiment 2), the results of which actually cast some doubt on the use of a binary gender decision task to indirectly assess gender retrieval for speech production. In line with the nature of the hypothesis at hand, it therefore seemed desirable to not only *manipulate* recency of gender access with speech production, but to also *assess the effect on later re-access* with a real speech production task. This was the approach of the following experiments.

2. Experiment 1

To study the retrieval of gender in speech production, Experiment 1 used a black-and-white/color picture naming task. In this task, the participant saw a series of pictures, some colored, some black-and-white (B and W; actually white-on-black), with the instruction to name a B and W picture with the bare noun, and a colored picture with the noun plus a color adjective. Unknown to the participant, the trial series contained a number of critical non-consecutive prime-target trial pairs. This could, for instance, be a car depicted in B and W on one trial, and a car depicted in red several trials later, in which case participants would produce the 'target' adjective noun phrase /rode auto/, 'red car', a little while after having produced the bare noun 'prime' /auto/, 'car'. In indefinite adjective noun phrases, most Dutch adjectives inflect for the gender of the noun they modify, e.g., 'rod*e* auto', 'rood huis' ('red_{com} car', 'red_{neu} house'). To produce the target response /rode auto/, participants thus need to retrieve the gender of 'auto'.

In the trial sequence just discussed, the production of the target utterance /rode auto/ merely follows the production of the prime utterance /auto/, and is therefore the *first* utterance with this noun that requires the retrieval of its gender in the experimental session. But if the prime utterance is /groen*e* auto/, 'green$_{com}$ car', i.e., triggered by a differently colored rather than a B and W picture of a car, the subsequent target utterance /rode auto/ will be the *second* utterance with this noun that requires the retrieval of its gender. Given that the most recent retrieval of this noun's gender is now only a few trials back, rather than at some undetermined but relatively much earlier moment in the participant's pre-experimental life, the gender recency account of Jescheniak and Levelt (1994) would predict faster gender retrieval for the target response.

Such faster retrieval is of course not necessarily reflected in the speed with which a target response like /rode auto/ is initiated. As gender is only needed for the adjective *suffix*, participants could in principle begin with /ro.../ at a point in time that is not systematically related to the moment at which the noun's gender becomes available. Work by Schriefers (1993) (see also Experiment 2) has shown, however, that the utterance onset response times of such gender-marked color adjective noun phrases are indeed sensitive to the speed with which the gender of a noun can be retrieved. The benefit of recent gender retrieval should therefore be visible in onset response times.

The gender recency account thus predicts that the speed with which participants can begin to utter a gender-marked noun phrase after recent use of the same noun in another gender-marked noun phrase will be faster than that after recent use of the same noun alone. The latter *noun-priming* condition controls for any facilitation effects due to recent use of the noun (and recent exposure of the associated picture) itself. The former *gender-priming* condition also necessarily involves recent noun use (and recent picture exposure), but it adds the effect, if any, of a recent gender retrieval for that particular noun.

The middle two rows of Table 1 exemplify these two conditions in a schematic way, for the case of three intervening trials. The number of intervening trials actually varied between two and six, so that, with the trial timing at hand, some 9–22 s separated verbal production of the target response from that of the prime response. Although this is a much shorter interval than the ~20 min used by Jescheniak and Levelt (1994) (Experiment 5), one would expect it to be a very, perhaps even more, appropriate time span within which to test their gender recency account: in natural spoken discourse, typical repeated reference (and the associated retrieval of gender, if needed) will most likely occur somewhere between a few seconds and a few minutes.

The results of an earlier experiment with similar methodology (van van Berkum, 1996, Chp. 4, Experiment 3) had pointed to a side-effect of such a repeated naming design: very recent use of a noun may increase its second, critical use by hundreds of milliseconds. In naming research, the effect of repetition is well-known, and its exact locus has been the subject of several studies (e.g., Wheeldon and Monsell, 1992). In the current design, however, a substantial noun repetition effect presents a problem: it might facilitate the second, critical naming response to such an extent that an additional gain from recent gender retrieval will be hard to obtain. If no such

Table 1
Example naming trial sequences for Experiment 1

Prime condition	Critical elements of example trial sequence						
		Prime trial			Pre-target trial	Target trial	
Null-priming	...	/trommel/	...	...	/blauw huis/	/rode auto/	...
Noun-priming	...	/*auto*/	...	...	/blauw huis/	/rode auto/	...
Gender-priming	...	/groen*e auto*/	...	...	/blauw huis/	/rode auto/	...
Phrase-priming	...	/*rode auto*/	...	...	/blauw huis/	/rode auto/	...

Only the intended prime utterance (left), the intended target utterance (right), and the last of in this case three intervening utterances (middle) are shown; prime response overlap with the target response is emphasized in italics.

benefit would be observed, one would thus not know whether this truly reflected the absence of a gender recency effect, or whether, given the response time range they were in, the participants simply could not get any faster.

Two control conditions were therefore added to the design (see Table 1, first and last row). In the *null-priming* condition, the target response, e.g., /rode auto/, followed a prime response with a different noun, e.g., /trommel/. Relative to this baseline, the *noun-priming* condition should provide an estimate of the effect of recent noun use (and picture exposure) in the current experiment, and should as such also serve as a basic check on the validity of the naming data yielded within this paradigm. The other control condition, *phrase-priming*, directly addressed the possibility of a floor effect in naming response times. In this condition, a target response such as /rode auto/ followed an identical prime response, i.e., also /rode auto/. If, under the conditions studied here, the participants' naming response times would already be at floor level due to very recent noun use, and such that recent gender use does not yield any *additional* facilitation, then one would not expect to see any substantial additional facilitation due to recent phrase use either. If, on the other hand, there is such substantial phrase recency facilitation, then it would be harder to argue that, should there be no gender recency effect, that this is due to some general floor effect in response times.

As can be seen in Table 1, this design effectively stacks a number of potential priming (or 'recency' or 'repetition') effects by having more and more elements of the target utterance also produced in the prime utterance: nothing in the null-priming condition, just the noun in the noun-priming condition, the noun and its gender in the gender-priming condition, and the noun, its gender, and the adjective (i.e., the whole phrase) in the phrase priming condition. To summarize the predictions for this design:

1. There should be a substantial *noun recency effect* in adjective noun naming responses, i.e., noun-primed responses should be considerably faster than null-primed responses.
2. If the gender recency account of Jescheniak and Levelt (1994) holds, there should be a *gender recency effect*, i.e., gender-primed responses should be faster than merely noun-primed responses, *unless participants already perform at ceiling.*

3. But if participants already perform at ceiling due to mere noun (and object) repetition, there should also be no *phrase recency effect*, i.e., phrase-primed responses should be no faster than gender- or noun-primed responses.

2.1. Method

2.1.1. Participants

Forty-eight native speakers of Dutch were recruited from the Max Planck Institute participant pool. All participants had normal color vision, and none of them had participated in any of the Jescheniak and Levelt (1994) experiments, which had also been conducted at the Max Planck Institute, or any of the related Experiments 1–3 reported in van Berkum (van van Berkum, 1996, Chp. 4). They were paid for their participation.

2.1.2. Materials

The experimental items were 48 line drawings of relatively simple objects, eight of which depicted low-frequent de-words (e.g., 'bijl', 'axe'), eight medium-frequent de-words (e.g., 'appel', 'apple'), eight high-frequent de-words (e.g., 'auto', 'car'), eight low-frequent het-words (e.g., 'anker', 'anchor'), eight medium-frequent het-words (e.g., 'hek', 'gate'), and eight high-frequent het-words (e.g., 'bed', 'bed'). Pictures were sampled from a database at the Max Planck Institute.

All picture names were monomorphemic words of one or two syllables, for which name agreement had been checked in an informal pretest. Every low-frequent (LF) picture name occurred less than 12 on a million times in a representative sample of written Dutch, and every high-frequent (HF) picture name had a word frequency of more than 60 on a million; these cut-off points were taken from Jescheniak and Levelt (1994). Medium-frequent (MF) picture names had a word frequency between 12 and 60 on a million. Word frequencies were derived from the 42 million word *INL* corpus of written Dutch, as recorded in the *CELEX* Dutch 'lemma lexicon' (Burnage, 1990). These 'lemma' frequencies, which collapse over singular and plural wordform occurrences, were checked against the average frequency of the singular wordform in the 600 000 words written *Eindhoven* corpus (Uit den Boogaart, 1975).

The six item sets (2 genders × 3 frequency classes) were further matched on average length in syllables and segments, as well as on average diphone frequency, phonological neighborhood density, and phonological neighborhood frequency, all variables that could have an independent effect on naming response times[4]. Table 2

[4]The *diphone frequency* of a word is the average of the log-transformed frequencies of its separate diphones in the *INL* text corpus. A higher diphone frequency indicates that the phoneme transitions within a word are, on average, more frequent ones, which means that the word's pronunciation will on the whole be a more common one. A word's *phonological neighborhood density* is the total number of words in the *CELEX* Dutch wordform lexicon that are of the same length as the one at hand and differ from it by a *single* phoneme, and its *phonological neighborhood frequency* is the sum of the log-transformed frequencies of those neighboring words in the INL text corpus (on a million word tokens).

Table 2
Average characteristics of the experimental items in Experiment 1

	de-words			het-words		
	LF	MF	HF	LF	MF	HF
Number of items	8	8	8	8	8	8
Lemma frequency	7.5	27.6	140.4	7.8	27.4	135.8
Log lemma frequency	1.91	3.25	4.88	1.85	3.23	4.79
Wordform frequency	1.9	9.4	47.4	2.5	7.3	50.8
Number of syllables	1.4	1.3	1.3	1.4	1.3	1.3
Number of phonemes	3.9	3.9	3.8	3.8	4.1	3.8
Diphone frequency	9.4	9.4	9.3	9.6	9.6	9.5
Neighborhood density	12.9	11.8	10.5	11.0	13.3	13.9
Neighborhood frequency	16.30	14.70	16.00	17.2	15.9	23.8

LF, Low-frequent; MF, medium-frequent; HF, high-frequent.

summarizes the properties of the experimental item sets; see Table A1 in Appendix A for a complete list of items.

The 48 experimental items featured on all 48 target trials, as well as on 12 prime trials in the noun-, gender-, and phrase-primed conditions each. For the null-primed condition, an additional 18 comparable pictures of simple objects were sampled from the above mentioned picture database. These control items all had monomorphemic names of one or two syllables, and of at least medium lemma frequency (12 or more on a million). Half of the control pictures had de-word names, half had het-word names, and the two sets were matched on the same criteria as used for the experimental item sets. The de-word control items were paired with the eight de-word experimental items in each frequency subset, such that the nouns of every pair had the same metrical structure, and such that there was no associative, semantic or (other) phonological relatedness between them. A similar procedure was used for het-word control items. Table A1 also shows the control items used.

In addition to the 48 experimental pictures and 18 control pictures, 16 further pictures were sampled to serve in the last trial immediately before a target trial. To increase the design's sensitivity, these 'pre-target' trials were controlled in several respects. First, every de-word target trial was preceded by a het-word pre-target trial, or vice versa (see Table 1), this to avoid any form of general (i.e., not *noun-specific*) gender priming or overt suffix priming. Second, pre-target pictures were always presented in color, so that the pre-target naming response would be of the same type as the target response (adjective noun naming). And, although the pre-target items occasionally shared the metrical structure of their target item, they were not associatively, semantically or (otherwise) phonologically related. All pre-target pictures had monomorphemic names of one or two syllables, and of at least medium lemma frequency (12 or more on a million). A final 26 filler (and practice) pictures, half of them with a de-word name, and half with a het-word name, were selected from the same database.

Pictures could be presented in B and W, red, green, yellow or blue. During the experiment, each of the 48 experimental items was presented once *as a target*. These target presentations were always in red or green, and therefore always required adjective noun responses, e.g., /rode auto/. For the *null-priming* condition, the control items of 12 experimental items were presented in B and W three to seven trials before the colored target occurrence of those experimental items (e.g., 'beker', ..., 'rode appel'). For the *noun-priming* condition, 12 other experimental items were presented in B and W three to seven trials before their colored occurrence as a target (e.g., 'bed', ..., 'rood bed'). For the *gender-priming* condition, 12 further experimental items were presented in an alternative color (either red or green) three to seven trials before their target occurrence (e.g., 'groene bijl', ..., 'rode bijl'). And for the *phrase-priming* condition, the 12 remaining experimental items were presented in the same color three to seven trials before their target occurrence (e.g., 'rode auto', ..., 'rode auto'). A participant would thus respond to 48 prime...target trial 'cycles' in all, 36 of which would repeat the same item, and 12 of which would not.

Sixteen different pseudo-random trial lists were created. Every list featured all 48 experimental pictures, 12 of them preceded by a phrase-prime picture, 12 by a gender-prime picture, 12 by a noun-prime picture, and the remaining 12 by a null-prime (control) picture. Randomizations were such that no more than six items of the same gender were presented in a row, and such that 12.5% of the trials were in red, 12.5% in green, 12.5% in yellow, 12.5% in blue, and 50% in B and W. To keep the total number of pictures manageable, experimental and control pictures were, after their use in critical trials, also used in filler trials. Across all 16 lists, every experimental item occurred in every priming condition in both red and green, and was matched by an experimental item of the alternative gender (but the same frequency class) occurring under the same conditions. In addition, the six item sets were matched, across lists, on the average list position of the four critical target presentations of an experimental item. Each of the resulting 16 randomizations contained 384 naming trials, divided into blocks of approximately 55 trials each.

2.1.3. Apparatus

Each participant was tested individually in a dimly lit, sound-attenuated booth. The picture stimuli were displayed centered on an EGA-driven NEC Multisync II computer monitor, as white or colored line drawings on a black background. Display size of the pictures was approximately 8 × 8 cm, and viewing distance was roughly 60 cm. A Hermac 386 computer with NESU experiment software (Nagengast and Baumann, 1994) controlled the presentation of the picture stimuli and the collection of response time data. Participants responded into a Sennheiser microphone, and speech onset latencies were measured by a voice key connected to the computer. Speech errors and problematic voice key responses were scored manually by the experimenter. Each session was also recorded on tape.

2.1.4. Procedure

Each participant was randomly assigned to one of the 16 experimental lists. Participants were first shown all pictures together with their names. During this preview block, each of 108 randomly ordered pictures was presented on screen for 2 s, after which the name appeared underneath the picture for at most another

4 s. Participants were asked to think of the name of the picture as soon as it was shown, to compare the name they had thought of to the intended name as soon as the latter was shown below the picture, and to pay some extra attention to the intended name if it had surprised them. No overt response was required. The participant could press a button to initiate the next preview trial as soon as he or she felt comfortable with the picture's intended name, or wait for the current trial to end by itself (at 6 s after picture onset). All pictures were presented in B and W.

For the main experiment, participants were asked to name every object displayed on the screen, together with the appropriate color adjective, or, if the picture was in B and W, with just the picture name. Participants were asked to respond both fast and accurately, to try not to begin with /ro.../, /groe.../, /gee.../ or /blau.../ (the onsets of the Dutch color names) before they knew what they were going to say, and to try to use the names they had seen during the preview. Attention was not explicitly drawn to grammatical gender.

After a short practice session, the participants named 384 pictures with or without an appropriate color adjective. After a visual warning dot presented for 300 ms and a pause of 200 ms, the target picture was displayed, in color or B and W, for 1000 ms. Then the screen was cleared for another 1600 ms, after which the next naming trial began. The naming test lasted about 25 min.

2.1.5. Analysis

For the noun recency control effect, analyses of variance (ANOVAs) were carried out on the utterance onset times and percentage errors of adjective noun target responses, as a function of the first and second prime types (null-priming or noun-priming, within participants, within items), gender (de- or het-word; within participants, between items), and word frequency (LF, MF or HF; within participants, between items). For the critical gender recency effect, the equivalent analyses of variance were carried out for the second and third prime types (noun-priming or gender-priming, within participants, within items). And for the phrase recency control effect, the equivalent analyses were done for the third and fourth prime types (gender-priming or phrase-priming, within participants, within items). For this and the next experiment, *F*-tests were performed at an alpha level of 0.05, both over participants (*F*1) and over items (*F*2).

ANOVAs were carried out on the data from target presentations of experimental items only, but, for these analyses, all other responses were also scored on correctness. Correct naming responses were of the form /<adjective> <noun>/ for color trials, and of the form /<noun>/ for B and W trials. A naming response was treated as an error if the noun or the adjective were erroneous, if the response was interrupted or began with a hesitation sound, or if it was initiated too late, i.e., after 2000 ms from picture onset. Statistical analysis collapsed across error types.

Target presentation response times were discarded (1) if the response itself had been erroneous in the above way or the voice key had not triggered correctly, (2) if the corresponding prime or pre-target response had been erroneous or if the item name had erroneously been used in response to some other picture presented before, or (3) if the response time deviated from the mean of that condition by more than 3

SD. A total of 380 response times, 16.5% of the data, were treated in this way, 4.9% because of an error or voice key failure on the response itself, 10.6% because of a prime or pre-target error or some other earlier error that implicated the same item, and an additional 0.9% because of outlying response times. Such loss of data is inherent in a picture naming design in which the validity of target naming responses is made contingent on two earlier naming responses. As a side-effect of having a complete within-participants design, however, several participants did not have valid response times in all of the 24 cells. In order not to lose the complete data set of these participants in the ANOVA, discarded response times had to be replaced by estimates. Following Winer et al. (1991), a missing response time for some item was replaced by the mean response time for the item in the condition at hand, adjusted for the participant's overall mean response time deviation from the other participants that had responded to the item in that condition[5].

2.2. *Results and discussion*

Results pertaining to a noun recency effect, a gender recency effect, and a phrase recency effect will be reported separately, but Fig. 1 displays average reaction times and error rates for all three, collapsed across word frequency. Table 3 contains the exact numeric values, also for separate frequency subsets.

The results of this experiment did not support the gender recency hypothesis, which had predicted faster gender-marked target responses such as /rode auto/ after a gender-priming response such as /groene auto/, three to seven trials back, than after an unmarked noun-priming response such as /auto/. As can be seen in the middle panel of Fig. 1, the predicted benefit was not obtained, neither for de-words, nor for het-words. But, as the left and right panels of Fig. 1 show, the null effect of gender recency was flanked by considerable effects of noun recency and, most importantly, phrase recency. Because participants gave much faster target responses, e.g., /rode auto/, after the identical phrase-prime response, e.g., also /rode auto/, than after the gender-prime response, e.g., /groene auto/, the absence of a gender recency effect cannot be attributed to a general floor effect in response times.

2.2.1. *Noun recency*

The analysis of target presentation response times revealed a large noun recency effect: participants were much faster to produce an adjective noun response like / rode auto/ after having produced the same bare noun, e.g., /auto/, three to seven trials back than after having produced a control noun, e.g., /trommel/ (null-prime, 855 ms; noun-prime, 765 ms; $F1(1,47) = 105.96$, $MSE = 11099$, $P < 0.001$; $F2(1,42) = 62.37$, $MSE = 3143$, $P < 0.001$). This 90 ms effect need not be the result of lexical repetition only, and it is in fact highly plausible that seeing an object for the second or third time (including the preview) makes a difference as well. The size of the noun recency effect varied with the target noun's gender ($F1(1,47) = 12.41$, $MSE = 6246$,

[5]None of the relevant effects was significantly different in analyses with or without additionally estimated data points.

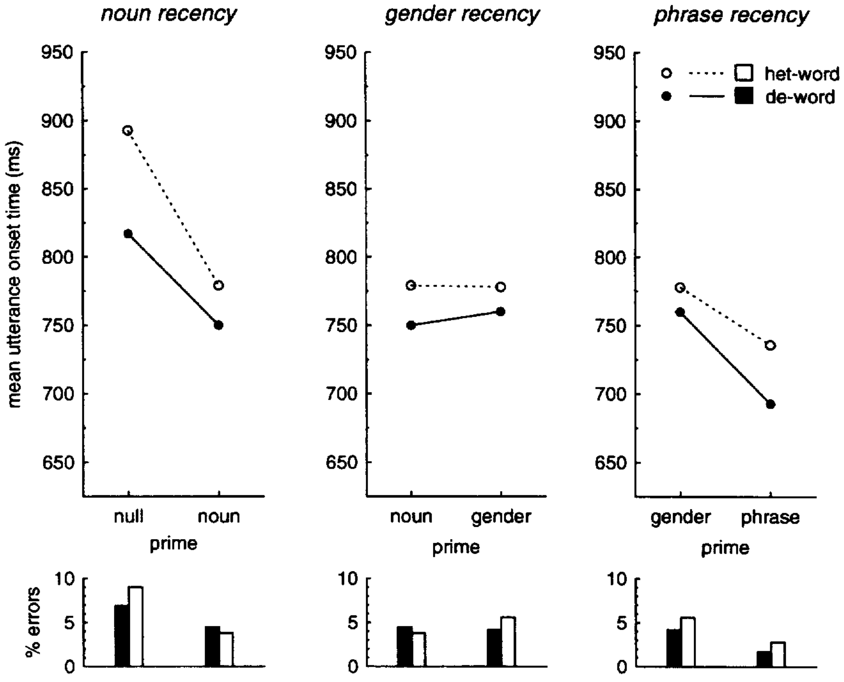

Fig. 1. Results of Experiment 1. Picture-based adjective noun naming response times and error rates for low-, medium- and high-frequent de- and het-word pictures after recent null-, noun-, gender- and phrase-prime naming responses.

$P < 0.01$; $F2(1,42) = 4.11$, $MSE = 3143$, $P < 0.05$), with its word frequency ($F1(2,94) = 8.94$, $MSE = 10393$, $P < 0.001$; $F2(2,42) = 4.93$, $MSE = 3143$, $P < 0.05$), and, marginally, with target noun gender and frequency considered together ($F1(2,94) = 6.42$, $MSE = 9047$, $P < 0.01$; $F2(2,42) = 3.08$, $MSE = 3143$, $P = 0.056$). But a simple effects analysis of the latter three-way interaction revealed a significant noun recency effect, in the right direction (of faster responses), for each of the six item sets. All interactions involving noun recency reflected the fact that the size of the effect was considerably larger for the low-frequent het-word set than for any of the other five item sets (see Table 3). Although a lexical source cannot be ruled out, this may well be the result of a disproportionate picture recognition gain, associated with five relatively complex pictures in that set.

As for error rates, participants made almost half the number of errors if the target item was presented for the second time (null-prime, 8.0%; noun-prime, 4.2%; $F1(1,47) = 6.98$, $MSE = 301.1$, $P < 0.05$; $F2(1,42) = 7.43$, $MSE = 47.1$, $P < 0.01$).

2.2.2. Gender recency

The analysis of target presentation response times confirmed the impression made by Fig. 1. Participants were as fast to produce a gender-marked adjective noun

Table 3
Results of Experiment 1

	de-words				het-words			
	LF	MF	HF	all	LF	MF	HF	all
	Absolute response measures							
Null-priming								
RT	871	797	783	817	1031	825	822	893
SD	141	126	120	110	206	159	129	128
%e	11.5	5.2	4.2	6.9	11.5	10.4	5.2	9.0
Noun-priming								
RT	796	724	730	750	829	738	771	779
SD	147	106	109	104	138	100	137	106
%e	7.3	3.1	3.1	4.5	8.3	1.0	2.1	3.8
Gender-priming								
RT	789	737	755	760	856	740	739	778
SD	140	118	114	103	127	115	100	99
%e	7.3	1.0	4.2	4.2	2.1	5.2	9.4	5.6
Phrase-priming								
RT	689	706	684	693	765	721	721	736
SD	96	99	98	89	100	109	123	98
%e	4.2	0.0	1.0	1.7	5.2	2.1	1.0	2.8
	Difference scores							
Noun recency effect								
RT	75	73	53	67	202	88	51	114
%e	4.2	2.1	1.0	2.4	3.1	9.4	3.1	5.2
Gender recency effect								
RT	7	−13	−24	−10	−26	−2	31	1
%e	0.0	2.1	−1.0	0.3	6.3	−4.2	−7.3	−1.7
Phrase recency effect								
RT	100	31	71	67	90	19	18	42
%e	3.1	1.0	3.1	2.4	−3.1	3.1	8.3	2.8

Picture-based adjective noun naming response times in milliseconds (RT), its standard deviation (SD), and percentage errors (%e) for low-, medium- and high-frequent (LF, MF and HF, respectively) de- and het-word pictures after recent null-, noun-, gender-, and phrase-prime responses.
Noun recency effect, null-primed – noun-primed; gender recency effect, noun-primed – gender-primed; phrase recency effect, gender-primed – phrase-primed; positive numbers represent faster responses and fewer errors with additional priming.
Apparent inaccuracies in difference scores or marginal means are the result of rounding.

response like /rode auto/ after having produced a gender-marked adjective noun response such as /groene auto/ three to seven trials back than after having produced an unmarked bare noun response like /auto/ (noun-prime, 765 ms; gender-prime, 769 ms; $F1(1,47) = 0.51$, $MSE = 5811$, $P > 0.40$; $F2(1,42) = 0.28$, $MSE = 1762$, $P > 0.50$). Having recently retrieved a noun's gender did not help to produce an utterance that required that same noun's gender again, even though the two retrieval events were separated by two to six other speech responses only. Furthermore, gender recency did not interact with the gender of the target noun ($F1(1,47) = 0.71$, $MSE = 6227$, $P > 0.40$; $F2(1,42) = 0.42$, $MSE = 1762$, $P >$

0.50), nor with its frequency ($F1(2,94) = 0.39$, $MSE = 6124$, $P > 0.60$; $F2(2,42) = 0.23$, $MSE = 1762$, $P > 0.70$). It also did not clearly interact with gender and frequency considered together ($F1(2,94) = 4.53$, $MSE = 5266$, $P < 0.05$; $F2(2,42) = 2.25$, $MSE = 1762$, $P > 0.10$). Numerically, the average effect of gender recency within each of these item sets ranged from an inhibitory –26 and –24 ms for low-frequent het-words and high-frequent de-words respectively to a facilitatory 31 ms for high-frequent het-words (see Table 3). But a simple effects analysis showed that none of these effects was significant (five out of six with $F1$ P-values > 0.10). The 31 ms trend for high-frequent het-words did approach significance ($F1(1,47) = 3.42$, $MSE = 6904$, $P = 0.071$; $F2(1,42) = 2.23$, $MSE = 1762$, $P = 0.143$), but, being the only interesting response time trend in the right direction, it was matched by over a 4-fold increase in error rates. It is clear that responses to none of the six item sets unambiguously gained from a recent gender retrieval.

The error rates analysis basically confirmed the results in response time. Most relevant, there was no effect of recent gender access, with participants making as many errors on a gender-marked adjective noun response like /rode auto/ after a recent gender-marked adjective noun response like /groene auto/ than after a recent unmarked bare noun response like /auto/ (noun-prime, 4.2%; gender-prime, 4.9%; $F1(1,47) = 0.36$, $MSE = 193.6$, $P > 0.50$; $F2(1,42) = 0.33$, $MSE = 35.3$, $P > 0.50$). The effect of gender repetition did marginally vary with word frequency ($F1(2,94) = 3.54$, $MSE = 181.4$, $P < 0.05$; $F2(2,42) = 3.03$, $MSE = 35.3$, $P = 0.059$). An inspection of the error rates in Table 3 suggests that words of high frequency were more inhibited by a recent gender-marked adjective noun response than those of low frequency. A simple effect analysis confirmed this impression, and also showed that the only significant gender recency effect was an *inhibitory* one obtained for high-frequent words.

2.2.3. Phrase recency

The analysis of target presentation response times revealed a substantial effect of phrase recency: participants were much faster to produce an adjective noun response like /rode auto/ after having produced the same adjective noun response, i.e., /rode auto/, three to seven trials back than after having produced a different adjective noun response, like /groene auto/ (gender-prime, 769 ms; phrase-prime, 714 ms; $F1(1,47) = 94.82$, $MSE = 4565$, $P < 0.001$; $F2(1,42) = 37.77$, $MSE = 1910$, $P < 0.001$).

Phrase recency also interacted with word frequency ($F1(2,94) = 12.39$, $MSE = 5057$, $P < 0.001$; $F2(2,42) = 5.47$, $MSE = 1910$, $P < 0.01$), although not with gender ($F1(1,47) = 5.29$, $MSE = 4220$, $P < 0.05$; $F2(1,42) = 1.95$, $MSE = 1910$, $P > 0.10$), or with the two considered together ($F1(2,94) = 1.51$, $MSE = 4709$, $P > 0.20$; $F2(2,42) = 0.62$, $MSE = 1910$, $P > 0.50$). An inspection of the means in Table 3 suggests that low-frequent words gained more than high-frequent ones. But simple effects analyses showed that facilitation from phrase recency was significant for words from all three frequency strata. And the only thing that mattered here was that responses to all item types could be further facilitated, irrespective of the size of the effect.

The error rates analysis yielded an effect of phrase recency only, with participants making fewer errors on an adjective noun response like /rode auto/ after recently having produced the same response, i.e. /rode auto/, than after recently having produced a different adjective noun response, like /groene auto/ (gender-prime, 4.9%; phrase-prime, 2.3%; $F1(1,47) = 7.78$, $MSE = 125.5$, $P < 0.01$; $F2(1,42) = 6.48$, $MSE = 25.1$, $P < 0.05$).

In all, the results did not support the gender recency hypothesis. Participants did not gain anything from recent gender retrieval, neither in response times, nor in error rates. But, as is clear from Fig. 1, they did gain from recent noun use as well as from recent phrase use, both in response times and error rates.

The sizable noun recency effect was as expected. But in spite of this substantial facilitation, the phrase recency result shows that additional gains could still be made, both in response times and error rates. Acting as a sensitivity check in the relevant response time range, this result clearly suggests that the absence of a benefit from recent gender retrieval was not caused by a general floor effect[6]. That alternative explanation would, incidentally, also be hard to reconcile with the hypothesized function of a gender recency mechanism. The experiment studied the effect of having retrieved the gender of a particular noun, some 9–22 s before, on the speed and accuracy with which speakers can produce a gender-marked noun phrase with the same noun. These are typical conditions under which a gender recency mechanism would be supposed to operate. If *noun* repetition were to eliminate a beneficial effect of recent gender retrieval, the mechanism would not do its job in a large part of its domain of operation, i.e., whenever re-access to a noun's gender would be for a nominal NP (noun phrase). It would then be left to facilitate gender-marked pronominal NPs only (an issue returned to in Section 4).

As pointed out by several reviewers, it is possible, in principle, that speakers did benefit from recent gender access, but that the facilitation effect was canceled out by a simultaneous similarity-based interference effect: speakers might find it more difficult to generate a phrase like /rode auto/ after just having produced the rather similar phrase /groene auto/ than after the less similar utterance /auto/. This possibility cannot be ruled out. But a closer inspection of the data does suggest that it is not a very likely account of the results at hand. Most importantly, if similarity-based interference were indeed a strong determinant of speech production here, one would expect to see not only slower responses (canceling out a gender recency effect) when producing utterances like /rode auto/ after /groene auto/ than after /auto/, but also more errors involving the *wrong color* (e.g., /groe...rode auto/). Such color errors did occur, but, in fact, there were *fewer* of these errors on, e.g., /rode auto/ after /groene auto/ (1.6%, with only 0.7% erroneously involving the prime's color) than after /auto/ (2.1%).

[6]Phrase recency effects have been established with gender-primed responses as a baseline, responses that already reflected noun *and gender* recency effects. This approach was consistent with the interpretation of the design as a stacking of recency ('identity priming', 'repetition') effects. One could argue, however, that, to control for a floor effect in noun-primed response times (the baseline for a gender recency effect), the phrase recency effect should have been established relative to noun-primed responses as well. A post-hoc analysis for which I redefined the phrase recency effect in this way yielded virtually identical results.

Furthermore, one might also expect such interference to be larger when only 2–3 unrelated trials, i.e., some 10 s, separated the target utterance from the interfering prime than when 5–6 trials, i.e., some 20 s, separated the two utterances. The design of Experiment 1 allowed for a test of the relevant priming × lag interaction in that every target followed every prime type after both a short lag (2–3 trials) and a long lag (5–6 trials). No such interaction was obtained, however, $F1(1,47) = 1.78$, $MSE = 10497$, $P = 0.19$; $F2(1,42) = 0.45$, $MSE = 6889$, $P = 0.51$.

It seems reasonable to conclude, therefore, that the production of a gender-marked noun phrase such as /rode auto/ was not facilitated by having retrieved the same noun's gender some 9–22 s before, even though conditions were similar to those under which one would typically expect a gender recency mechanism to do its job.

The production of an utterance like /rode auto/ is the result of a large number of coordinated subprocesses, only some of which may determine the resulting utterance onset latency. Earlier findings with gender-marked adjective noun phrase production in Dutch (Schriefers, 1993) clearly suggest that the gender retrieval process is amongst those that determine final response times for these utterances, this is why the null result of Experiment 1 can be considered informative. Furthermore, as just discussed, a gender recency mechanism is only going to contribute to the fluency of speech production if gender retrieval is indeed one of the time course bottlenecks. That is, if the functionality hypothesis of Jescheniak and Levelt is correct, there simply must be conditions under which the gender recency mechanism reveals itself in response times. Still, it would be reassuring to independently verify that gender processing is visible in utterance onset latency while testing for a gender recency effect in the very same experiment. This was realized in Experiment 2.

3. Experiment 2

Experiment 2 was an all-speech variant of Jescheniak's and Levelt's two-phase Experiment 5. In the manipulation block, one group of participants named a series of pictures with just the bare noun, e.g., /auto, huis, slak, .../, whereas another group named them with the gender-marked adjective 'klein(e)', 'small', before the noun, e.g., /kleine auto, klein huis, kleine slak, .../. In the assessment block, participants from both groups were tested on the speed with which they could retrieve the gender of the same picture names. It was here that Jescheniak and Levelt had used a push-button gender decision task. In the current experiment, however, the benefit of recent gender retrieval was assessed by means of another speech production task: all participants were asked to name the same pictures again, now together with the appropriate definite article, e.g., /de slak, de auto, het huis, .../.

To stay as close as possible to Jescheniak's and Levelt's Experiment 5, the current experiment did not only have a closely related design, but it also used the same item sets and trial randomizations. The obvious advantage, maximum comparability, was deemed to outweigh two drawbacks associated with these materials. First, in agreement with their slightly different research interest, Jescheniak and Levelt had only controlled the de-word items in their materials, and had used the (more variable) het-word items as fillers. Second, due to the blocking approach, some 20 min separated the critical retrieval of a particular noun's gender (the 'assessment' trial) from its most recent earlier retrieval in the experiment (the 'manipulation' trial). The current experiment thus unavoidably tested the effect of recent gender retrieval for de-words only, and after a relatively long time. Note, however, that the recency effect observed *had* been obtained with de-words across such a long interval (Jescheniak and Levelt, 1994, Experiment 5); under the current explanation for that effect, it should simply, or, rather, *especially*, replicate in an all-speech variant of the original study.

Like Experiment 1, the potential benefit of a recent gender retrieval was assessed by means of later gender-marked naming responses, in this case with a definite article. As for adjective noun phrases, earlier work by Schriefers (1993) had already indicated that speech production latencies for Dutch noun phrases beginning with a gender-marked definite article were sensitive to the time course of gender retrieval. To verify that this would indeed also be the case under current circumstances, Experiment 2 directly incorporated the gender-related manipulation with which Schriefers had obtained these results.

In his picture naming experiments, Schriefers had observed that the production of a simple gender-marked noun phrase, as a naming response to some picture, can be delayed by presenting a gender-incongruent distractor word at approximately the same time. In one of these experiments (Schriefers, 1993, Experiment 1), Dutch participants were asked to name a series of colored pictures with a definite article plus adjective plus noun response, e.g., /de rode auto, het groene bed, .../ ('the red car, the green bed, ...'). Not surprisingly, participants were slower to produce something like /de rode auto/ if a written distractor word was superimposed on the picture. But the *grammatical gender* of the distractor words also made a difference: congruent distractor words, i.e., those that had the same gender as the picture name, interfered much less with gender-marked picture naming than incongruent distractors, whose gender was a different one.

For the current picture example of 'auto', a de-word, the gender-congruent distractor would be another de-word, such as 'voet' ('foot'), and the gender-incongruent distractor would be a het-word, such as 'been' ('leg'). At a stimulus onset asynchrony (SOA) of 0 ms, i.e., with simultaneous picture-distractor presentation, there was an average 56 ms *gender distraction effect* in the production of phrases like /de rode auto/, and an average 31 ms effect in that of phrases like /rode auto/. Schriefers replicated these effects in a variety of languages, with both written and spoken distractor words, and at several picture-word SOA's (Schriefers, 1996). On the basis of his research, he concluded that the effect reflected a competition between two activated gender representations. During gender-marked noun phrase production, the selection of the head noun's lemma would activate the appropriate gender. The incongruent distractor word, however, would automatically activate a competing gender representation, and would thereby delay the selection of the head noun's gender information for inclusion into the noun phrase.

Assuming that this interpretation is correct, gender-incongruent distractor words interfere with the retrieval of grammatical gender for speech production. If this

Table 4
Example naming trial sequences for Experiment 2 (with gender distractors)

Prime condition	Critical elements of example trial sequence							
	Prime trial						Target trial	
Noun-priming, CON-distractor	…	/*auto*/	…	…	…	…	/de auto/ plaat	…
Gender-priming, CON-distractor	…	/klein*e auto*/	…	…	…	…	/de auto/ plaat	…
Noun-priming INC-distractor	…	/*auto*/	…	…	…	…	/de auto/ strand	…
Gender-priming, INC-distractor	…	/klein*e auto*/	…	…	…	…	/de auto/ strand	…

Only the intended prime utterance and the intended target utterance are shown, the latter with a gender-congruent (CON) or -incongruent (INC) written distractor word; prime response overlap with the target response is emphasized in italics.

gender distraction effect could be replicated under the same conditions as those used to test for the gender recency effect, it would unequivocally show that the gender retrieval process is visible in utterance onset response times. A *null* result of recent gender retrieval would then be very informative. Experiment 2 therefore combined two orthogonal manipulations of the gender retrieval process: *recency of earlier gender access*, realized, as in Experiment 1, by means of earlier noun- or gender-priming responses, and *gender distraction*, realized by means of written distractor words (cf. Schriefers, 1993). The resulting orthogonal design is schematically depicted in Table 4.

Table 4 exemplifies all relevant trials that involved the de-word ‘auto’, for each of four different participants. The first and second participants both produced the critical assessment response, /de auto/, in response to the picture of a car that was displayed together with the superimposed de-word ‘plaat’, ‘plate’, a gender-congruent distractor word (*CON*). In terms of gender distraction, therefore, the critical naming trials of these participants were equated. In terms of gender recency, however, they were not: whereas the first participant had named the same picture with the bare noun (*noun-priming*) some time before, only the second participant had named it with a gender-marked adjective noun phrase (*gender-priming*). The third and fourth participants differed on the recency of gender access in the same way. But *these* two participants produced the critical assessment response, /de auto/, in response to a car that was displayed together with the *het*-word ‘strand’, ‘beach’, a gender-incongruent distractor word (*INC*).

The predictions for this completely crossed design were as follows.

1. In line with Schriefers (1993), there should be a *gender distraction effect*, which would demonstrate that, under the circumstances in this experiment, variability in

the gender retrieval process was indeed visible in definite article noun naming response times.

2. In that case, the hypothesis of Jescheniak and Levelt (1994) clearly predicted a *gender recency effect*: participants who recently named pictures in a gender-marked adjective noun form should be faster to name the same pictures in a definite article noun form than participants who recently named the pictures in the unmarked bare noun form only.

3.1. Method

3.1.1. Participants

Forty-eight native speakers of Dutch were recruited from the Max Planck Institute participant pool. None of them had participated in Experiment 1, in any of the Jescheniak and Levelt (1994) experiments, or in any of Experiments 1–3 reported in van Berkum (van van Berkum, 1996, Chp. 4). They were paid for their participation.

3.1.2. Materials

The 48 experimental and 48 filler pictures, as well as their randomizations, were those of Jescheniak and Levelt (Jescheniak and Levelt, 1994, Experiment 5). The experimental items were 48 line drawings of simple objects (e.g., a star, a snail, a car, a broom), sampled from a database at the Max Planck Institute. All experimental items depicted monomorphemic de-words of one or two syllables (e.g., 'ster', 'slak', 'auto', 'bezem'). Picture name agreement had been checked in an informal pretest. Half of the experimental pictures were named by a LF de-word, like 'slak' or 'bezem', and the other half were named with HF de-words, such as 'ster' or 'auto'. Every LF picture name had a word frequency of less than 12 on a million, and every HF picture name had a word frequency of more than 60 on a million, as determined by *CELEX* lemma frequency counts on the *INL* corpus (Burnage, 1990). The two sets of items had been matched on average length in syllables and segments, and on the number of word-initial consonants (relevant for the automatic registration of speech onset latencies). Table 5 summarizes the properties of the experimental item sets; see Table A2 for a complete list of items.

Another 48 pictures, line drawings of simple objects from the same picture database, were used as filler items. The names of the filler items were all het-words, and they covered a wide frequency range. In addition to the 96 test items (48 experimental and 48 filler items), 10 more pictures, half of them with a de-word name and the other half with a het-word name, were used as practice items.

The written distractor words were sampled from the materials of three written word recognition experiments reported in van Berkum (van Berkum, 1996, Chp. 3). From the 120 words used in those experiments, 12 de-words and 12 het-words were selected to serve as gender-congruent (CON) and gender-incongruent (INC) distractors for the experimental pictures (which all had de-word names). All 24 experimental distractor words were mono- or bi-syllabic, between three and six letters

Table 5
Average characteristics of the experimental items in Experiment 2

	de-words	
	LF	HF
Number of items	24	24
Lemma frequency	6.0	150.7
Log lemma frequency	1.59	4.9
Number of syllables	1.1	1.1
Number of phonemes	3.7	3.7
Diphone frequency	9.2	9.4
Neighborhood density	12.2	11.5
Neighborhood frequency	18.7	17.7

long, and of at least medium lemma frequency (over 12 on a million). Words with a meaning that might interfere with that of any picture in the response set were avoided. Words were also avoided if they had attracted more than 10% erroneous responses in an unpublished word-based gender decision experiment (n = 24), if they had yielded relatively slow gender decision responses in that same experiment, or if they had yielded relatively slow lexical decision or word naming responses in the word recognition experiments mentioned above The two sets of distractor words were matched on average previous gender decision error rate, as well as on lemma frequency, word length in syllables and letters, bigram frequency, orthographic neighborhood density and neighborhood frequency. Table 6 summarizes the properties of the resulting distractor word sets.

Another 24 distractor words, 12 het-words and 12 de-words, were sampled from the same materials to serve as gender-congruent and gender-incongruent distractors for the filler pictures (which all had het-word names). Selection and matching criteria were as above. For use in practice trials, 30 more comparable distractor words were selected.

Table 6
Average characteristics of the critical distractor words in Experiment 2

	Distractors	
	CON (de-words)	INC (het-words)
Number of distractor words	12	12
Gender decision error rate	3.1	3.5
Lemma frequency	50.8	53.6
Number of syllables	1.3	1.3
Number of letters	4.7	4.6
Bigram frequency	9.9	10.1
Neighborhood density	7.3	7.3
Neighborhood frequency	12.9	14.5

CON, Gender-congruent distractors (for de-word target pictures); INC, gender-incongruent distractors (for de-word target pictures); gender decision error rates were determined in a separate word-based experiment (see text).

In the assessment block, experimental and filler pictures were presented in the order specified by the two different randomizations of Jescheniak's and Levelt's Experiment 5. These two pseudo-random trial sequences, each with 192 test trials, had been created, such that (1) the presentation of an experimental item was never preceded by the presentation of a phonologically, semantically, or associatively related item; (2) no more than five items of the same gender were presented in adjacent trials; and (3) all items were presented twice, with at least 20 trials separating the two presentations. The use of two presentations in this assessment block pertained to the research goals of Jescheniak and Levelt; for current purposes, only the first of these presentations is considered to be a critical trial.

For the current experiment, distractor words were pseudo-randomly allocated to the pictures specified in these two sequences, with each original sequence giving rise to two new assessment trial randomizations. Every resulting randomization contained 48 critical trials, 24 involving high-frequent de-word pictures, and 24 involving low-frequent de-word pictures, all occurring for the first time in the list. Twelve pictures of every subset were, for their critical first presentation, paired with a gender-congruent distractor word (a de-word), and the other 12 were paired with a gender-incongruent distractor word (a het-word). For their non-critical second presentation in the list, distractor allocation was reversed. Each of the 48 het-word filler pictures was likewise paired with a congruent (het-word) and an incongruent (de-word) distractor. Each of the four new assessment lists thus specified two presentations of 48 de-word pictures and of 48 het-word pictures, half of each paired with a congruent distractor first and an incongruent distractor second, and the other half with a reversed allocation. Across the four lists, every experimental de-word picture was (first) presented with two different gender-congruent and two different gender-incongruent distractors. Furthermore, congruently and incongruently distracted picture presentations were equated, across lists, on their average position in this block, and on their average distance to the last of two related trials in the preceding manipulation block (the critical lag that defined the most recent gender access). Picture and distractor names were never associatively or semantically related, nor did they share word-initial segments or actual syllables. Table A2 lists every experimental picture name together with its two congruent and two incongruent distractor words.

Each of the four new assessment block lists was preceded by a 192 picture trial randomization for the manipulation block, identical to those used in Jescheniak's and Levelt's Experiment 5. As in the assessment block, every picture featured twice here. This implied that, at the critical (first) presentation of a picture in the assessment block, participants would have named that particular picture twice before (either as a bare noun or together with a gender-marked adjective). In the gender distraction experiments of Schriefers (1993), however, participants repeated a much smaller set of pictures much more often, such that, on average, a picture had already been named 12 times before. To better approximate those conditions in the present experiment, under the time constraints imposed by a much larger picture set, the manipulation block was preceded by an extra preparation block. In this block, participants were to name each of the 96 pictures twice, at leisure, and with a bare noun. For this purpose, two different randomizations of 96 picture preparation trials preceded the first and second of the above lists in A–B order, and the third and fourth of those lists in B–A order. They were separated by a short pause, and preceded by a 10 trial practice series.

The four resulting experimental lists were such that all 48 experimental pictures occurred under both distractor conditions at their critical first assessment presentation, doing so at two different list positions with two different congruent-incongruent distractor pairs. As the gender recency manipulation was done by means of instructions for the manipulation block, all four lists, and hence all four critical trials of a given picture, were presented under both priming conditions.

3.1.3. Apparatus

The same apparatus was used as in Experiment 1. Display size of the pictures was approximately 8 × 8 cm, and viewing distance was roughly 60 cm. In all blocks of the experiment, pictures were presented as light-gray line drawings on a black background. In the assessment block, light-grey distractor words were simultaneously presented in a 32 × 13 mm black area 'cut out' from the middle of the picture. The largest distractor word had a display size of 27 × 8 mm.

3.1.4. Procedure

Each participant was randomly assigned to one of the two priming conditions, and to one of the four experimental lists. Participants were first shown all 106 pictures together with their names, using the same automated screen preview procedure as in Experiment 1.

3.1.4.1. Preparation block. Participants were asked to name every object presented on the screen with the name they had seen during the preview block, and to do so as accurately as possible. After a short practice session, they named 192 pictures with the bare noun. Trial timing was identical to that in Experiment 1. The preparation block lasted about 15 min.

3.1.4.2. Manipulation block. Next, participants were told they would see the same pictures again, now for a speeded naming task. In the *gender-priming* condition, participants were to name every object displayed on the screen together with "the adjective *klein(e)*", 'small'. In the *noun-priming* condition, they were to name every object by itself only. Participants were asked to respond both fast and accurately, to try not to begin with /klei.../ before they knew what they were going to say (gender-priming participants only), and to use the names they has seen during the preview block. After a short practice session, the participants named 192 pictures with the adjective 'klein(e)' (gender-priming) or with the bare noun only (noun-priming). Trial timing was identical to that in the preparation block. The manipulation block lasted about 15 min.

3.1.4.3. Assessment block. For the last part of the experiment, participants were told they would see the same pictures again, and that their task now was to name every object together with "the appropriate article *de* or *het*". Participants were also informed they would see words inside the pictures, but that this was irrelevant to their task. They were asked to respond both fast and accurately, to try not to begin with /de.../ or /het.../ before they knew what they were going to say, and to try again to use the names they has seen during the preview block. After a short practice session, the participants named 192 pictures with the appropriate definite article. Apart from the fact that every picture now contained a superimposed distractor word, trial timing was identical to that in the preceding blocks (and Experiment 1). The assessment block lasted about 15 min.

3.1.5. Analysis

ANOVAs were carried out on the utterance onset times and percentage errors of critical definite article noun responses, as a function of prime type (noun-priming or gender-priming, between participants, within items), distractor type (CON or INC; within participants, within items), and word frequency (LF or HF; within participants, between items). For every experimental picture, this critical response was the *first* response given in the assessment block; it was preceded by two unspeeded naming responses in the preparation block and two speeded naming responses in the manipulation block. Preparation and manipulation trial responses were also scored on correctness.

In the manipulation block, correct naming responses were of the form /kleine <noun>/ for participants in the gender-priming condition, and of the form /<noun>/ for participants in the noun-priming condition. In the assessment block, correct naming responses were of the form /de <noun>/ for all participants. In each block, a naming response was treated as an error if the noun, adjective, or article were erroneous, if the response was interrupted or began with a hesitation sound, or if it was initiated too late, i.e., after 2000 ms from picture onset. Statistical analysis collapsed across error types.

Critical first assessment response times were discarded (1) if the response itself had been erroneous in the above way or the voice key had not triggered correctly, (2) if any of the four corresponding preparation or manipulation responses had been erroneous or if the item name had erroneously been used in response to some other picture presented in any of the blocks before, or (3) if the response time deviated from the mean of that condition by more than 3 SD. A total of 487 response times, 21.1% of the data, were treated in this way (4.1% because of an error or voice key failure on the response itself, 15.7% because of a preparation or manipulation naming error or some other earlier error that implicated the same item, and an additional 1.3% because of outlying response times).

3.2. Results and discussion

Fig. 2 displays average reaction times and error rates for all eight combinations of prime type, distractor type, and word frequency; Table 7 contains the corresponding numeric values.

Like Experiment 1, the results of this experiment did not support the gender

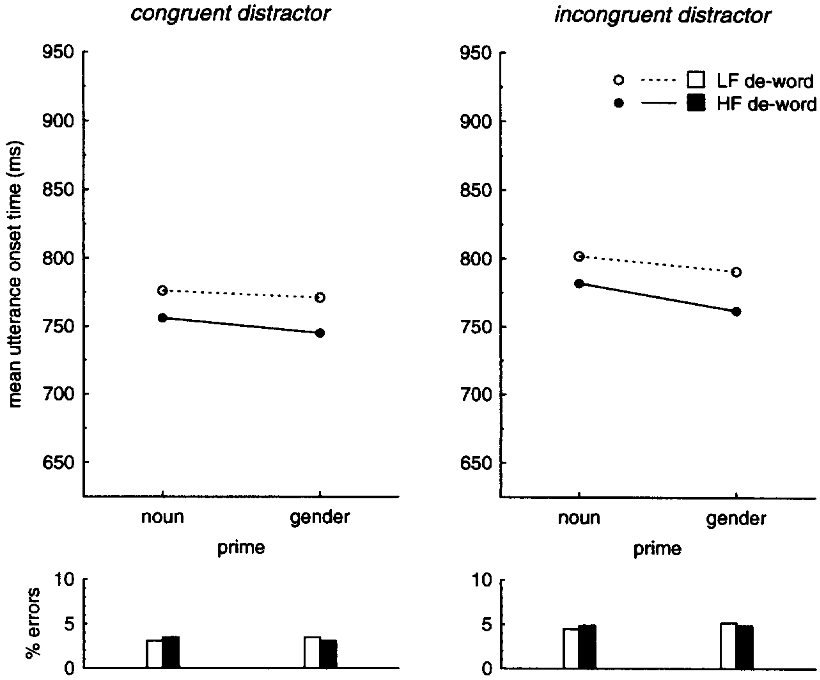

Fig. 2. Results of Experiment 2. Picture-based definite article noun naming response times and error rates for low- and high-frequent de-word pictures with superimposed gender-congruent (CON) and gender-incongruent (INC) distractor words, after recent noun- and gender-prime naming responses.

recency hypothesis. Although the effect was small, participants were generally slower in producing definite article noun phrases for a de-word picture if the distractor word was a het-word (INC) than if it was also a de-word (CON). This replication of the gender distraction effect reported by Schriefers (1993) showed that response times on the naming task at hand were co-determined by the gender retrieval process. But there was no benefit of recent access to the gender of the same picture names. Nor did the equally recent gender access for low- and high-frequent picture names in gender-primed naming lead to a subsequent reduction of the word frequency effect (an aspect of the gender recency effect that Jescheniak and Levelt had focused on).

The response time ANOVA revealed a small but clearly significant gender distraction effect. Definite article noun naming of de-word pictures was 22 ms slower if the picture was shown together with a gender-incongruent het-word than if it was shown with a gender-congruent de-word (CON, 762 ms; INC, 784 ms; $F1(1,46) = 17.55$, $MSE = 1376$, $P < 0.001$; $F2(1,46) = 10.83$, $MSE = 1754$, $P < 0.01$). In spite of the fact that their only task was to name the pictures, participants had obviously not been able to avoid processing the distractor words in a linguistic way, and to process the gender of those words as part of that. Importantly, distractor gender interfered with the retrieval of the picture name's gender for speech production in a way that was visible in utterance onset response times. This showed that, under the conditions of this experiment, the definite article noun picture naming task was indeed sensitive to the speed with which a noun's grammatical gender is retrieved for speech production.

The gender recency hypothesis had predicted that a noun's gender should be retrieved faster if that same noun's gender had been accessed some time before. But this was not the case overall: participants who had recently used the picture names in a gender-marked adjective noun phrase were no faster in the definite article noun picture naming task than participants who had recently used those names in a bare noun response (noun-prime, 779 ms; gender-prime, 767 ms; $F1(1,46) = 0.32$, $MSE = 21321, P > 0.50$; $F2(1,46) = 3.84, MSE = 841, P > 0.05$). The 12 ms trend

Table 7
Results of Experiment 2

	de-words			
	LF	HF	All	WF effect
	Gender-congruent distractors			
Noun-priming				
RT	776	756	766	20
SD	73	90	78	
%e	3.1	3.5	3.3	–0.4
Gender-priming				
RT	771	745	758	27
SD	82	84	80	
%e	3.5	3.1	3.3	0.4
Gender recency effect				
RT	5	12	8	
%e	–0.4	0.4	0.0	
	Gender-incongruent distractors			
Noun-priming				
RT	802	782	792	20
SD	72	82	74	
%e	4.5	4.9	4.7	–0.4
Gender-priming				
RT	791	762	777	29
SD	76	68	69	
%e	5.2	4.9	5.0	0.3
Gender recency effect				
RT	11	20	16	
%e	–0.7	0.0	–0.3	

Picture-based definite article noun naming response times in milliseconds (RT), its standard deviation (SD), and percentage errors (%e) for low- and high-frequent (LF and HF, respectively) de-word pictures with superimposed gender-congruent and gender-incongruent distractor words, after recent noun- and gender-prime responses.
WF (word frequency) effect, LF – HF; gender recency effect, noun-primed – gender-primed; positive numbers represent faster responses and fewer errors with additional priming.
Apparent inaccuracies in difference scores or marginal means are the result of rounding.

depended on one of 48 participants only. Reanalysis without this one participant yielded a non-significant 3 ms difference between the response times after noun-primed and gender-primed naming ($F1(1,45) = 0.03$, $MSE = 18708$, $P > 0.80$; $F2(1,46) = 0.00$, $MSE = 926$, $P > 0.90$), but a robust 20 ms gender distraction effect ($F1(1,45) = 14.42$, $MSE = 1369$, $P < 0.001$; $F2(1,46) = 9.20$, $MSE = 1750$, $P < 0.01$).

As in Experiment 1, gender recency effects in gender-marked noun phrase production were analyzed in terms of direct facilitation, if any, from recent gender-marked noun phrase use (gender-priming), relative to recent noun use only (noun-priming). In their Experiment 5, however, Jescheniak and Levelt had actually not evaluated *this* particular aspect of the data, at least, not in a direct statistical test. Instead, they had focused on whether experimentally equating low- and high-frequent picture names on the recency of gender access also equated them on the speed of subsequent gender retrieval. In their gender decision data, it did; whereas the gender decision response times to low- and high-frequent picture names differed if earlier naming had been with a bare noun only, this word frequency effect in gender decision times disappeared completely if earlier naming had been with a gender-marked noun phrase.

As is evident from Fig. 2, and as confirmed by the absence of a frequency by priming interaction ($F1(1,46) = 0.89$, $MSE = 855$, $P > 0.30$; $F2(1,46) = 1.29$, $MSE = 841$, $P > 0.20$), there was no such collapse in the results of the current experiment. After having named the pictures with an unmarked bare noun, the average definite article noun naming response with a high-frequent picture name was 20 ms faster than that with a low-frequent picture name. But after having named the pictures with a gender-marked adjective noun phrase, responses to high-frequent picture names were still 28 ms faster. This means that equating low- and high-frequent picture names on the recency of gender access did not equate them on the speed with which their gender could be retrieved again.

It should be noted that the overall 24 ms word frequency effect was not significant by items (LF, 785 ms; HF, 761 ms; $F1(1,46) = 32.15$, $MSE = 855$, $P < 0.001$; $F2(1,46) = 3.36$, $MSE = 8596$, $P = 0.073$). One might therefore argue that there was not anything to disappear. But simple effects analysis of the word frequency effect at the two different levels of the prime type factor revealed that, although there was no clearly significant frequency effect in definite article noun naming after recent bare noun naming (LF, 789 ms; HF, 769 ms; $F1(1,46) = 11.18$, $MSE = 855$, $P < 0.01$; $F2(1,46) = 1.91$, $MSE = 4912$, $P > 0.10$), such an effect did clearly emerge after recent gender-marked adjective noun naming (LF, 781 ms; HF, 753 ms; $F1(1,46) = 21.86$, $MSE = 855$, $P < 0.001$; $F2(1,46) = 4.55$, $MSE = 4524$, $P < 0.05$). This does not support the gender recency hypothesis, because experimentally induced equally recent gender access for low- and high-frequent words should, under that account, reduce or perhaps even eliminate, rather than enlarge, differences in later access time.

Up until now, the presence of a gender distraction effect has been taken to demonstrate that the task at hand was sensitive to the gender retrieval process under all conditions tested in the experiment. Strictly speaking, however, the presence of a gender distraction effect has only shown that the process of gender retrieval was visible in definite article noun naming responses to pictures with incongruent distractors. This is because part of the true distractor interference effect may have been used to slow down the gender retrieval process to the extent that it would just *begin* to determine final onset response times. On this account, an *overall* gender recency effect was not necessarily expected. But a gender recency effect should then at least have been observed on trials with incongruent distractors, i.e., in cases where the process of gender retrieval demonstrably determined response times. The data showed otherwise. Prime type did not interact with distractor type ($F1(1,46) = 0.49$, $MSE = 1376$, $P > 0.40$; $F2(1,46) = 1.86$, $MSE = 1080$, $P > 0.10$), indicating there was as little facilitation from recent gender retrieval on incongruent distractor trials as there was on congruent distractor trials.

The interaction between word frequency and prime type did not vary across distractor condition either ($F1(1,46) = 0.01$, $MSE = 1162$, $P > 0.90$; $F2(1,46) = 0.09$, $MSE = 1080$, $P > 0.70$). As can be seen in the right panel of Fig. 2, equating low- and high-frequent picture names on the recency of gender access (gender-priming) did not equate them on the speed with which their gender could be retrieved again, not even under conditions where the assessment task was demonstrably sensitive to the gender retrieval process (INC distractor).

There were no significant effects in the corresponding error rates analysis. Importantly, participants made as many errors on gender-marked definite article noun naming after a prior gender-marked adjective noun naming block as they did after prior bare noun naming (noun-prime, 4.0%; gender-prime, 4.2%; $F1(1,46) = 0.02$, $MSE = 65.6$, $P > 0.80$; $F2(1,46) = 0.04$, $MSE = 39.1$, $P > 0.80$). Prime type also did not interact with distractor type ($F1(1,46) = 0.04$, $MSE = 38.9$, $P > 0.80$; $F2(1,46) = 0.04$, $MSE = 36.2$, $P > 0.80$), nor with picture name frequency ($F1(1,46) = 0.21$, $MSE = 27.8$, $P > 0.60$; $F2(1,46) = 0.15$, $MSE = 39.1$, $P > 0.70$) or distractor type by picture name frequency considered together ($F1(1,46) = 0.00$, $MSE = 37.0$, $P > 0.90$; $F2(1,46) = 0.00$, $MSE = 36.2$, $P > 0.90$). Although participants made, on average, slightly more errors when naming pictures that had gender-incongruent distractors than when naming pictures that had gender-congruent distractors, this effect failed to reach significance by participants (CON, 3.3%; INC, 4.9%; $F1(1,46) = 3.01$, $MSE = 38.9$, $P = 0.089$; $F2(1,46) = 4.48$, $MSE = 26.1$, $P < 0.05$). Nevertheless, the tendency made sense. An examination of the different error types showed it to be entirely caused by an increase in the percentage of article selection errors, e.g., as evident in a response like /het... de auto/. Whereas such errors had been made on only 12 out of the 1152 trials with gender-congruent distractor words (1.0%), they featured on 32 out of 1152 trials with gender-incongruent distractor words (2.8%). This suggests that gender-incongruent distractors did not just slow down the retrieval of the correct gender for the picture name, but also caused the incorrect gender to be selected on almost three times as many occasions.

The main result of this study, then, is that the production of a gender-marked definite article noun phrase such as /de auto/ was not facilitated by recent access to the same noun's gender, e.g., for /kleine auto/ rather than /auto/, even though the same response could be successfully delayed by gender-incongruent distractor words.

4. General discussion

Two experiments explored the gender recency hypothesis of Jescheniak and Levelt (1994). These authors had suggested that the speed with which a noun's grammatical gender information is retrieved for speech production may depend on the recency of access to that same noun's gender, as the result of a lexical retrieval mechanism that is dedicated "to facilitate gender-marked anaphoric reference to recently introduced discourse entities, therefore contributing to the fluency of the utterance" (Jescheniak and Levelt, 1994, p. 841). This explained the results of their Experiment 5, where recent use of a noun in a gender-marked picture naming response, such as /*de* auto/ or /*het* huis/, facilitated a later DE-HET gender decision for the same picture name.

In the current experiments, however, recent gender access never facilitated later gender-marked *speech production* responses. Experiment 1 showed that the production of a gender-marked adjective noun phrase, such as /rod*e* auto/, or /rood huis/, could not be facilitated by a recent gender retrieval, e.g., for the production of /groene auto/ or /groen huis/, even though recent access had been only a few seconds before, and even though, in the same experiment, a comparable control condition revealed that participants had not yet performed at ceiling. Likewise, Experiment 2 showed that the production of a gender-marked definite article noun phrase, such as /*de* auto/, or /*de* slak/, did not benefit from a recent gender retrieval, e.g., for the production of /klein*e* auto/ or /klein*e* slak/, even though it *was* affected by the gender of a distractor word superimposed on the picture to be named (cf. Schriefers, 1993). In this experiment, moreover, recent gender access for low- and high-frequent words did not eliminate, nor even diminish, a word frequency effect in later gender-marked naming response times, a critical prediction of the gender recency hypothesis.

It is important to emphasize again that, even though Jescheniak and Levelt had suggested the existence of a gender-dedicated memory retrieval mechanism to account for results obtained in a picture-based gender decision task, they meant it to be an hypothesis about the retrieval of gender for *speaking*. Experiments 1 and 2 thus tested the gender recency account in its proper domain: gender-marked speech production. Their null results seem to indicate that, whatever underlies the complicated pattern of gender decision results obtained before (Jescheniak and Levelt, 1994, Experiment 5; van van Berkum, 1996, Experiments 1 and 2), it is probably not a gender-dedicated retrieval mechanism geared to the specific demands of speaking.

Before we can accept the latter proposition, however, several potential counter arguments should be addressed. One might suggest, for example, that recent gender access possibly only helps in a real-life meaningful discourse, where repeated reference is being made to the same discourse entity, or that it perhaps only helps in the production of a later *pronominal* NP, the domain that Jescheniak and Levelt had exemplified with their discourse fragment 'I saw a big horse ... it crossed the road'. But note that such domain restrictions would not be in the spirit of the hypothesis at hand. The reason is that the mechanism, a recency-sensitive connection between a noun's lemma and a grammatical gender representation, has clearly been defined at the level of the 'functional architecture' (Pylyshyn, 1984) for speaking. That is, it would be part of the relatively invariant processing machinery that supports speech production. This is precisely the level of system organization that would presumably be *indifferent* to such things as whether or not one is having a meaningful discourse, or whether the retrieved gender is going to be for an adnominal adjective or an independent pronoun. In this respect, it is much more plausible to assume that the results obtained with nominal NP production tasks (e.g., with responses such as /rode auto/ or /de ster/) generalize to the domain of pronominal NP production.

Because native speakers of Dutch do not (sufficiently) reliably mark *grammatical* gender on independent anaphoric pronouns (van van Berkum, 1996, Chp. 2), the relevant generalization test cannot be done in this language. But with its relatively reliably gender-marked personal pronouns 'er', 'sie' and 'es', the German language, for example, does allow for such a test. In fact, Jescheniak (1996) has recently conducted a pronominal variant of Experiment 1 to this end. The results showed that native speakers of German were no faster to produce a gender-marked pronoun after recent use of the antecedent noun in a gender-marked NP (e.g., /*ein* Schuh/... /*Er*/) than after recent use of this noun in an unmarked NP (e.g., /Schuh/... /*Er*/). In line with the theoretical argument just discussed, this German result confirms the generality of the results reported here, both with respect to the domain of gender-marked utterances, and with respect to the language under study.

As a second alternative explanation for the current results, one might argue that the control condition for a gender recency effect, recent production of the target noun in an unmarked NP, had provided an inappropriate baseline. Although the production of a bare noun does not *require* gender, this does not necessarily imply that speakers can avoid *automatic access* to that noun's gender. The spreading-activation architectures of recent network models of the mental lexicon (Dell, 1986; Berg, 1992; Roelofs, 1992; see also Schriefers, 1993) would in fact lead one to expect such automatic gender access, as a mere side-effect of word node activation. But if such automatic access does indeed occur, then the participants in Experiment 2 must have accessed the gender of target nouns on both the gender-prime *and the noun-prime* responses. A dedicated gender retrieval mechanism would then have operated in both conditions. This explanation obviously cannot be ruled out. But if it is correct, then it must by the same token apply to the baseline of Jescheniak's and Levelt's Experiment 5, which means that even *their* results can not be taken to support the gender recency account. In fact, empirical tests might then be impossible to devise.

An important third issue remains. The current results, although apparently problematic for the hypothesis at hand, are all null results. Hence, some might say, "intrinsically uninformative". As extensively argued by Frick (1995), however, null results are not necessarily uninformative: if the results are sufficiently clear, and if the experiment qualifies as a "good effort" to reject the null hypothesis, this hypothesis can be legitimately accepted. Such an experiment uses many participants and many trials per participant, controls for major sources of variance, uses manipulations that are as effective as possible and a measurement that is as sensitive as possible, and avoids floor and ceiling effects (Frick, 1995, p. 135).

Relative to the original study that gave rise to the gender recency hypothesis, Jescheniak's and Levelt's Experiment 5, the current experiments do seem to qualify as "good efforts". Both experiments, for instance, used twice the number of participants as the original experiment had. Assessing a gender recency effect in terms of direct facilitation rather than, as Jescheniak and Levelt had done, as a modulation of word frequency effects, implied that, for the critical comparison, every item served as its own control. Furthermore, to obtain a more effective gender recency manipulation, Experiment 1 had reduced its distance to the measurement from some 20 min to 9–22 s. And, importantly, Experiments 1 and 2 had included relevant control conditions, such that null results could not easily be attributed to a lack of sensitivity.

As pointed out by one of the reviewers, though, the critical gender recency comparison had been between-, rather than within-subjects, in Experiment 2; this will certainly have yielded lower power than what would have been possible in a within-subjects design. Could it be that we therefore failed to detect a small gender recency effect? Note, first of all, that Experiment 1 did use a within-subjects design, but still failed to detect the relevant effect. Also note that the original experiment, Jescheniak's and Levelt's Experiment 5, had used the same between-subjects design as that of Experiment 2, with half the number of subjects. More importantly, the effect had not been that small: in Jescheniak's and Levelt's Experiment 5, low-frequent words gained a substantial 59 ms as a function of the (between-subjects) gender recency manipulation. The current 5 and 11 ms trends for (the very same) low-frequent words in gender-congruent and gender-incongruent distractor conditions respectively (see Table 7) come nowhere near that original effect size[7]. Even more telling, the *qualitative* pattern of results in Experiment 2 simply does not match the critical result of Jescheniak and Levelt: whereas the latter featured a reduction of the word frequency effect in gender-determined response times as a function of the gender recency manipulation, Experiment 2 actually featured a trend in the opposite direction (see Table 7, rightmost column).

Taken together, then, the results of these two speech production experiments do speak against the gender recency hypothesis, as do the recent pronominal speech production results in German (Jescheniak, 1996). In addition, one should note that the original *gender decision* result (Jescheniak and Levelt, 1994, Experiment 5) did not replicate in a large-scale follow-up experiment with virtually identical metho-

[7]Although this effect size comparison involves different response tasks, it can be justified with the same assumption that allowed (Jescheniak and Levelt, 1994, p. 831) to interpret gender decision response times as an accurate reflection of the time course of early processes in speech production.

dology (but a different item set; van van Berkum, 1996, Chp. 4, Experiment 2; although see ibid., Experiment 1). This suggests that a reinterpretation of that original finding, if required, should perhaps await further empirical verification.

What implications does all this have for our understanding of the speech production process, and the retrieval of grammatical gender therein? First of all, fluency in speech production is apparently not supported by a dedicated mechanism for the retrieval of grammatical gender. This is a somewhat negative result, unfortunately, and the converse would have been a lot more interesting. But it is the most parsimoneous account of the results.

The fact that knowledge of grammatical gender does not appear to be stored in a way that is responsive to very recent, incidental use of that knowledge in the course of speaking is also interesting by itself, because, as already suggested before, there are lots of reasons why we might expect the retrieval of a noun's grammatical gender to benefit from earlier retrieval. Indeed, virtually every human activity benefits from repetition (Newell and Rosenbloom, 1981), and lexical processing is no exception to the rule (e.g., Wheeldon and Monsell, 1992; Kirsner and Speelman, 1993). The ubiquity of word frequency effects in lexical performance can be taken to reflect the structural, long-lasting effects of repeated exposure to or use of the words at hand. In addition to such structural effects, most models of lexical processing also assume relatively ephemeral effects of a single repetition at short lag (often associated with the concept of 'residual activation', e.g., Dell, 1986).

Although the experiments had included relatively low-frequent words, the absolute amount of pre-experimental practice that adult native speaker of Dutch will have had with the gender of words like 'slak' or 'anker' ('snail', 'anchor') may be such that one, or, in Experiment 2, two recent access events are not enough to result in relatively structural changes. But why is the retrieval of a noun's grammatical gender not primed by short-term repetition either? One, in my opinion somewhat remote, possibility is that the gender of a noun is *computed* from scratch again and again, and this process is not responsive to short-term repetition. On the other hand, it is also possible that the gender of a noun is *stored* in a format that is simply not, or no longer, responsive to one or two recent access events, not even in the case of words that have a relatively low frequency of use. Thus, although the results at hand have ruled out all versions of either a storage or a computation account that would generate a benefit of incidental recent gender access, they have not decided the issue of storage versus computation itself.

Beyond having these relatively immediate implications, the current findings also bear on the more general issue addressed by Jescheniak and Levelt (1994), the locus of the word frequency effect in speech production. For these researchers, the gender recency account was not an isolated hypothesis, but part of a larger chain of inferences, whose outcome had suggested a particular locus for the word frequency effect in speech production. In their Experiment 1, Jescheniak and Levelt had obtained the effect of interest, a word frequency effect in picture naming, and one that did not diminish in size across repetition. The results of their Experiments 2 and 3 indicated that this stable effect was most likely a lexical effect. It was as yet unclear, though, whether the phenomenon reflected a stable difference in the retrieval time of lemmas (i.e., words as lexical-syntactic entities), or in that of the associated wordforms. It was the *gender recency interpretation* of the results of their Experiment 5 (and Experiment 4), an *unstable* word frequency effect in lemma-determined gender decision response times, that allowed Jescheniak and Levelt to infer that the *robust* effect had to be related to the wordform (rather than, as proposed by Dell (1990), to the lemma). Now that this interpretation cannot be maintained, the particular chain of inferences in which it took part may need to be revised as well[8].

The two experiments reported here were designed to evaluate a particular hypothesis about the retrieval of grammatical gender for speech production. But Experiment 2 has also yielded new evidence for another phenomenon associated with gender retrieval for speaking: the gender distraction effect (Schriefers, 1993). As will be recalled, Schriefers had observed that, if Dutch speakers were asked to name a picture with a simple gender-marked noun phrase, their naming response could be delayed by presenting a *gender-incongruent* distractor word at about the same time. This distraction effect was interpreted as the result of a competition between two activated gender representations, only one of which was to be projected onto the syntax of the utterance under construction. The Dutch experiments of Schriefers had shown that such interference can delay the onset of (color) adjective noun phrases that do or do not begin with a definite article, such as /rode ster/, or /de rode ster/.

With Experiment 2, the phenomenon has now also been established for definite article noun phrases *without* an adjective, such as /de ster/ (in line with other recent findings; La Heij et al., 1997). And it has been established with a much larger set of pictures, each of which was named a few times only. This is not a trivial extension. The experiments of Schriefers had used only 10 target pictures in an elaborate repeated measures design, such that a participant's average critical naming response had in fact been his or her 13th naming response to that particular picture. The extent to which such thorough repetition had contributed to the gender distraction effect was still unknown. The results at hand now show that it is possible to selectively tap into the production of grammatical gender agreement under conditions that somewhat better approximate those for real speaking.

Amongst other things, the gender recency account of Jescheniak and Levelt (1994) was of interest because it could perhaps help us understand how native speakers so fluently coordinate words into larger linguistic units, and how they produce agreement as part of that. In this respect, one might say, the gender recency effect seems to have let us down. The gender *distraction* effect, however, turned out to be a robust phenomenon. Perhaps *this* phenomenon, rather than gender recency, will provide us with a window on the process of speaking in more than one word.

[8]The word frequency theory of Jescheniak and Levelt was also grounded in the results of their Experiment 6, which showed that low-frequent homophones (e.g., 'steeds', 'urban') inherited the accessing speed of their high-frequent homophone twins (e.g., 'steeds', 'always'). Because homophones only share their wordform, and not their lemma, this suggested a wordform-level origin of the robust frequency effect.

Acknowledgements

The research reported in this paper was part of a Ph.D. project supported by the Max-Planck-Gesellschaft zur Förderung der Wissenschaften, München, Germany. I thank Pim Levelt for supervising the project, Jörg Jescheniak and Pim Levelt for letting me use the materials of their 1994 study in Experiment 2, Herbert Baumann, Ger Desserjer, Inge Doehring, John Nagengast, and Paul Verleg for technical assistance, Monika Baumann, Kay Bock, Uli Frauenfelder, José Kerstholt, Paul Meijer, Antje Meyer, Molly Potter, Ardi Roelofs, Bernadette Schmitt, Herbert Schriefers, Miranda van Turennout and Linda Wheeldon for helpful discussion, and Antje Meyer, Marianne Wissink, and three anonymous reviewers for their comments on the manuscript.

Appendix A

Table A1
Materials for Experiment 1

LF de-word	Control	MF de-word	Control	HF de-word	Control
bijl (axe)	mond (mouth)	appel (apple)	beker (mug)	auto (car)	trommel (drum)
fluit (flute)	aap (monkey)	beer (bear)	hand (hand)	fles (bottle)	bloem (flower)
kassa (till)	trommel (drum)	bril (glasses)	aap (monkey)	hond (dog)	weg (road)
kegel (ninepin)	beker (mug)	kaars (candle)	weg (road)	koning (king)	beker (mug)
kikker (frog)	lepel (spoon)	lamp (lamp)	bloem (flower)	trap (steps)	hand (hand)
tol (top)	vrouw (woman)	motor (motorbike)	trommel (drum)	trein (train)	vrouw (woman)
vaas (vase)	weg (road)	pijl (arrow)	vrouw (woman)	voet (foot)	aap (monkey)
zwaan (swan)	hand (hand)	tent (tent)	mond (mouth)	zak (bag)	mond (mouth)

LF het-word	Control	MF het-word	Control	HF het-word	Control
anker (anchor)	kussen (pillow)	hek (fence)	schaap (sheep)	bed (bed)	oog (eye)
bekken (pelvis)	masker (mask)	kasteel (castle)	cadeau (present)	blad (leaf)	hoofd (head)
frame (frame)	oog (eye)	kruis (cross)	oog (eye)	brood (loaf)	huis (house)
hert (deer)	boek (book)	mes (knife)	boek (book)	bureau (desk)	cadeau (present)
kanon (gun)	cadeau (present)	spook (ghost)	ei (egg)	eiland (island)	kussen (pillow)
koor (choir)	huis (house)	varken (pig)	kussen (pillow)	paard (horse)	ei (egg)
luik (hatch)	ei (egg)	wiel (wheel)	hoofd (head)	raam (window)	boek (book)
vlot (raft)	hoofd (head)	zwaard (sword)	huis (house)	vuur (fire)	schaap (sheep)

Low-, medium- and high-frequent (LF, MF and HF, respectively) de- and het-word picture names are shown together with the name of their control picture (for null-prime trials).

Table A2
Materials for Experiment 2

LF de-word	CON distractor	INC distractor	HF de-word	CON distractor	INC distractor
bezem (broom)	arts (doctor)	doel (target)	auto (car)	plaat (plate)	strand (beach)
kano (canoe)	partij (party)	hotel (hotel)	arm (arm)	knop (button)	graf (grave)
hark (rake)	maat (measure)	vel (skin)	bank (sofa)	arts (doctor)	doel (target)
harp (harp)	kroeg (pub)	scherm (screen)	boom (tree)	plaat (plate)	strand (beach)
kam (comb)	partij (party)	hotel (hotel)	bloem (flower)	wond (wound)	wijf (woman)
krab (crab)	maat (measure)	vel (skin)	broek (trousers)	partij (party)	hotel (hotel)
peer (pear)	cel (cell)	vak (section)	boot (boat)	partij (party)	hotel (hotel)
bijl (axe)	keizer (emperor)	ritme (rhythm)	trap (steps)	arts (doctor)	doel (target)
rups (caterpillar)	kroeg (pub)	scherm (screen)	brief (letter)	cel (cell)	vak (section)
fluit (flute)	cel (cell)	vak (section)	kerk (church)	kroeg (pub)	scherm (screen)
slee (sledge)	knop (button)	graf (grave)	fles (bottle)	ramp (disaster)	veld (field)
zaag (saw)	heuvel (hill)	nummer (number)	ster (star)	heuvel (hill)	nummer (number)
snavel (beak)	ramp (disaster)	veld (field)	mond (mouth)	knop (button)	graf (grave)
schaar (scissors)	wond (wound)	wijf (woman)	deur (door)	keizer (emperor)	ritme (rhythm)
step (scooter)	lijst (list)	plein (square)	neus (nose)	lijst (list)	plein (square)
slak (snail)	knop (button)	graf (grave)	muur (wall)	ramp (disaster)	veld (field)
tang (tongs)	keizer (emperor)	ritme (rhythm)	stoel (chair)	heuvel (hill)	nummer (number)
tol (top)	plaat (plate)	strand (beach)	schoen (shoe)	wond (wound)	wijf (woman)
uil (owl)	plaat (plate)	strand (beach)	tafel (table)	kroeg (pub)	scherm (screen)
pauw (peacock)	arts (doctor)	doel (target)	vinger (finger)	maat (measure)	vel (skin)
vaas (vase)	heuvel (hill)	nummer (number)	vis (fish)	maat (measure)	vel (skin)
worst (sausage)	ramp (disaster)	veld (field)	zak (bag)	lijst (list)	plein (square)
zwaan (swan)	wond (wound)	wijf (woman)	voet (foot)	keizer (emperor)	ritme (rhythm)
spin (spider)	lijst (list)	plein (square)	hond (dog)	cel (cell)	vak (section)

Low- and high-frequent (LF and HF, respectively) de-word picture names are shown together with their gender-congruent (CON) and -incongruent (INC) distractor words.

Every picture name has two different congruent distractors and two different incongruent distractors, which it shares with another comparable picture name (shown immediately above or below).

References

Anderson, J.R., 1993. Rules of the Mind. Erlbaum, Hillsdale, NJ.

Badecker, W., Miozzo, M., Zanuttini, R., 1995. The two-stage model of lexical retrieval: evidence from a case of anomia with selective preservation of grammatical gender. Cognition 57, 193–216.

Berg, T., 1992. Prelexical and postlexical features in language production. Applied Psycholinguistics 13, 199–235.

van Berkum, J.J.A., 1996. The psycholinguistics of grammatical gender: studies in language comprehension and production. Doctoral dissertation, University of Nijmegen, The Netherlands.

Bickerton, D., 1984. The language bioprogram hypothesis. Behavioral and Brain Sciences 7 (2), 173–221.

Bock, J.K., 1995. Sentence production: from mind to mouth. In: Miller, J.L., Eimas, P.D. (Eds.), Handbook of perception and cognition: Vol. 11. Speech, language, and communication. Academic Press, San Diego, CA, pp. 181–216.

Bock, J.K., Eberhard, K.M., 1993. Meaning, sound, and syntax in English number agreement. Language and Cognitive Processes 8, 57–99.

Bock, J.K., Miller, C.A., 1991. Broken agreement. Cognitive Psychology 23, 45–93.

Burnage, G., 1990. CELEX: a guide for users. Centre for lexical information, University of Nijmegen, Nijmegen, The Netherlands.

Corbett, G.G., 1991. Gender. Cambridge University Press, Cambridge.

Corbett, G.G., 1994. Gender and gender systems. In: Asher, R.E., Simpson, J.M.Y. (Eds.), The Encyclopedia of language and linguistics. Pergamon Press, Oxford, Vol. 3, pp. 1347–1353.

Dell, G.S., 1986. A spreading-activation theory of retrieval in sentence production. Psychological Review 93 (3), 283–321.

Dell, G.S., 1990. Effects of frequency and vocabulary type on phonological speech errors. Language and Cognitive Processes 5 (4), 313–349.

Fodor, J.A., 1983. The Modularity of mind: an essay on faculty psychology. MIT Press, Cambridge, MA.

Frick, R.W., 1995. Accepting the null hypothesis. Memory and Cognition 23 (1), 132–138.

Henaff Gonon, M.A., Bruckert, R., Michel, F., 1989. Lexicalization in an anomic patient. Neuropsychologia 27 (4), 391–407.

Jescheniak, J.-D., 1994. Word frequency effects in speech production. Unpublished doctoral dissertation, Max Planck Institute for Psycholinguistics, Nijmegen, The Netherlands.

Jescheniak, J.-D. 1996. The production of gender-marked pronouns: a study of primed object naming. Manuscript submitted for publication. Freie Universität Berlin, Germany.

Jescheniak, J.-D., Levelt, W.J.M., 1994. Word frequency effects in speech production: retrieval of syntactic information and of phonological form. Journal of Experimental Psychology: Learning. Memory and Cognition 20 (4), 824–843.

Kempen, G., Huijbers, P., 1983. The lexicalization process in sentence production and naming: indirect election of words. Cognition 14, 185–209.

Kirsner, K., Speelman, C., 1993. Is lexical processing just an 'ACT'? In: Collins, A.F., Gathercole, S.E., Conway, M.A., Morris, P.E. (Eds.), Theories of memory. Erlbaum, Hove, UK, pp. 303–326.

La Heij, W., Mak, P., Sander, J., Willeboordse, E., 1997. The gender-congruency effect in picture-word tasks. Manuscript submitted for publication. Univ. of Leiden, The Netherlands.

Lapointe, S.G., Dell, G.S., 1989. A synthesis of some recent work in sentence production. In: Carlson, G.N., Tanenhaus, M.K. (Eds.), Linguistic structure in language processing. Kluwer, Dordrecht, The Netherlands, pp. 107–156.

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

Mallinson, G., Blake, B.J., 1981. Language typology: cross-linguistic studies in syntax. North-Holland, Amsterdam.

Nagengast, J., Baumann, H., 1994. NESU experiment specification and execution software, version 4 [computer program]. Max Planck Institute for Psycholinguistics, Nijmegen, The Netherlands.

Newell, A., Rosenbloom, P.S., 1981. Mechanisms of skill acquisition and the law of practice. In: Anderson, J.R. (Ed.), Cognitive skills and their acquisition. Erlbaum, Hillsdale, NJ, pp. 1–55.

Nicol, J.L., 1995. Effects of clausal structure on subject-verb agreement errors. Journal of Psycholinguistic Research 24 (6), 507–516.

Pinker, S., 1994. The language instinct: the new science of language and mind. Allen Lane, Penguin Press, London.

Pinker, S., Bloom, P., 1990. Natural language and natural selection. Behavioral and Brain Sciences 13, 707–784.

Pylyshyn, Z.W., 1984. Computation and cognition: toward a foundation for cognitive science. MIT Press, Cambridge, MA.

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

Schriefers, H., 1993. Syntactic processes in the production of noun phrases. Journal of Experimental Psychology: Learning. Memory, and Cognition 19 (4), 841–850.

Schriefers, H., 1996. The production of nounphrases in German and French: crosslinguistic studies. Manuscript submitted for publication. Univ. of Nijmegen, The Netherlands.

Uit den Boogaart, P.C., 1975. Woordfrequenties in Geschreven en Gesproken Nederlands. Oosthoek, Scheltema and Holkema, Utrecht, The Netherlands. (Computerized as the Eindhoven corpus, expanded version distributed by the Free University Amsterdam, 1989).

Vigliocco, G., Antonini, T., Garrett, M.F., 1997. Grammatical gender is on the tip of Italian tongues. Psychological Science, in press.

Vigliocco, G., Butterworth, B., Semenza, C., 1995. Constructing subject-verb agreement in speech: the role of semantic and morphological factors. Journal of Memory and Language 34, 186–215.

Wheeldon, L.R., Monsell, S., 1992. The locus of repetition priming of spoken word production. Quarterly Journal of Experimental Psychology 44A (4), 723–761.

Winer, B.J., Brown, D.R., Michels, K.M., 1991. Statistical principles in experimental design. McGraw-Hill, New York.

Zubin, D.A., Köpcke, K.-M., 1981. Gender: a less than arbitrary grammatical category. In: Hendrick, R.A., Masek, C.S., Miller, M.F. (Eds.), Papers from the seventh regional meeting of the Chicago linguistic society. University of Chicago Press, Chicago, IL, pp. 439–449.