

Running Head: PHONOLOGY AND GENDER RETRIEVAL

Phonological Regularities and Grammatical Gender Retrieval
in Spoken Word Recognition and Word Production

pre-print version,

published in the Journal of Psycholinguistic Research (2008) 37:419-442

Abstract

Two experiments investigate whether native speakers of French can use a noun's phonological ending to retrieve its gender and that of a gender-marked element. In Experiment 1, participants performed a gender decision task on the noun's gender-marked determiner for auditorily presented nouns. Noun endings with high predictive values were selected. The noun stimuli could either belong to the gender class predicted by their ending (congruent) or they could belong to the gender class that was different from the predicted gender (incongruent). Gender decisions were made significantly faster for congruent nouns than for incongruent nouns, relative to a (lexical decision) baseline task. In Experiment 2, participants named pictures of the same materials as used in Experiment 1 with noun phrases consisting of a gender-marked determiner, a gender-marked adjective and a noun. In this Experiment, no effect of congruency, relative to a (bare noun naming) baseline task, was observed. Thus, the results show an effect of phonological information on the retrieval of gender-marked elements in spoken word recognition, but not in word production.

Keywords: phonology, grammatical gender, language production, language comprehension.

In a language with grammatical gender, every noun belongs to a gender class. Other words in a phrase have to agree with the noun's gender. For example, if a speaker of French wants to say “the green table”, he or she needs to know that table belongs to the class of feminine gender nouns, and it is therefore “la_{fem.} table verte_{fem.}”, whereas “the green bed” is translated as “le_{mas.} lit vert_{mas.}” into French.¹

Phonological and orthographic regularities might help a native speaker of French to determine the gender of nouns. For example, according to the database *lexique*, 81% of all French nouns ending in /kl/ (like *cercle*, circle) have masculine grammatical gender, and 86% of all French nouns ending in /ez/ (like *fraise*, strawberry) have feminine gender. In this respect, French is what Taft and Meunier (1998) have named a pseudo-regular language: Some noun endings are strongly associated with a particular gender (e.g., the above mentioned nouns), but there are also endings with a small predictive value; nouns with these endings are about equally distributed across the two gender classes. An example is the ending *-al(e)* (*ovale* - oval_{mas.}, *sandale* – sandal_{fem.}).

Tucker, Lambert, and Rigault (1977) established the predictivity of the written endings of all nouns occurring in the French dictionary *Petit Larousse*, and showed that native speakers of French use the predictivity of a noun's ending when they have to assign gender to non-words. Several studies have looked at the influence of orthographic regularities on grammatical gender in gender assignment to non-words, and in language comprehension (e.g., Desrochers & Paivio, 1990; Desrochers, Paivio, & Desrochers, 1989; Holmes & Dejean de la Bâtie, 1999; Holmes & Segui, 2004; 2006; Taft & Meunier, 1998). In contrast to these studies, in the present study we will focus on the potential role of phonological rather than orthographic regularities on gender processing. We will provide a comparison of auditory language comprehension (Exp. 1) and language production (Exp. 2), using the same materials in a comprehension task (gender

decision) and in a production task (picture naming). In both tasks, we looked at the retrieval of gender-marked elements, that is, words that have to agree with a noun's grammatical gender. For example, the form of the definite determiner in French is determined by the gender of the noun it belongs to (e.g., *le lit* [the_{mas.} bed] vs. *la table* [the_{fem.} table]), and the same holds, among others, for indefinite determiners (*un* [a_{mas.}] vs. *une* [a_{fem.}]), and for adjectives (e.g., *le grand lit* [the_{mas.} big_{mas.} bed] vs. *la grande table* [the_{fem.} big_{fem.} table]).

In the following, we will first review studies on the role of word form regularities (in most cases orthographic regularities) in gender retrieval in language comprehension, restricting the review to French, the language used in the present experiments. Then we will discuss studies that indicate a potential influence of word form regularities on gender retrieval in language production. Finally, we will report two experiments on the role of phonological word form regularities in the retrieval of gender-marked elements in comprehension and production.

Word form regularities in language comprehension

Two types of word form regularities relating form to gender are commonly studied in language comprehension. The first type of regularity is the predictive value of the noun's ending as discussed above. The predictive value of the ending can be strong. We will refer to nouns with a strong predictive ending as *congruent*, when they belong to the gender class predicted by the ending, and as *incongruent*, when they belong to the gender class not predicted by the ending. In addition, an ending can also be linked with about equal strength to both gender classes. We will refer to these types of nouns as *neutral*. The second type of regularity that links form to gender is the informative value of a noun's associated determiner. The determiner that is associated with a given noun can be regarded as an important information source in gender retrieval. Feminine nouns in French, the language tested in the present experiments, are generally preceded by gender-marked determiners *la* or *une*, and masculine nouns by the determiners *le* or *un*.

However, the definite determiner does not carry gender information when a noun starts with a vowel, because then both the feminine determiner *la* and the masculine determiner *le* are elided to *l'*. It is assumed that in these cases it is more difficult to retrieve gender information. Whether a noun's associated determiner is full or elided, is dependent on whether the noun starts with a vowel or a consonant. Therefore, we will refer to this variable as "noun onset".

Several researchers (e.g., Holmes & Segui, 2004; Tucker et al., 1977) claim that the association between a noun and its determiner(s) is the primary source of gender information. That is, when asked what the gender of the noun *table* (table) is, French speakers will implicitly retrieve the corresponding determiner (e.g., *la* in the case of *table*), and from this determiner they will derive the judgment that the noun belongs to a certain gender class (feminine in the case of *table*). As Holmes and Segui, (p. 428) put it: "When asked to classify a word's gender, people try to evoke implicitly the closest lexical associate, typically the definite article". Tucker et al. (1977) provide some anecdotal evidence supporting this idea: When asked to give the gender of a low-frequency noun in French, a child was heard to say "*un floraison ... une floraison ... c'est féminin.*" (a_{mas.} flowering ... a_{fem.} flowering ... it's feminine). All studies discussed below have looked at either the role of predictivity of a noun's ending, or the role of noun onset, or a combination of these two factors.

Desrochers and colleagues (1989) investigated the role of the predictive value of the orthography of a noun's ending. In both conditions, the nouns were congruent (i.e., they were all of the predicted gender), but the strength of the predictivity varied. Desrochers et al. compared endings with mean predictive values of 97% (high predictivity) and 71% (low predictivity), respectively. Participants either had to read aloud the visually presented nouns, or they had to make a gender decision using the indefinite determiners *un* and *une*, or they had to make a gender decision using the linguistic labels *masculine* and *feminine*. Desrochers et al. found no effect of

predictivity in reading. In contrast, in the gender decision task participants were faster for nouns with highly predictive endings than for nouns with less predictive endings. Finally, participants' reaction times were about 200 ms shorter when they had to choose the correct determiner than when they had to choose the correct linguistic label. Furthermore, the difference between highly predictive and less predictive endings was independent of the type of response (determiner or linguistic label).

Taft and Meunier (1998, Exp. 1) presented participants with written French nouns. Participants had to decide as quickly as possible whether the nouns had masculine or feminine gender. Pairs of French nouns were selected such that one member of a pair was congruent (called "typical" by Taft & Meunier, e.g., *folie*_{fem.}) and the other member was incongruent ("atypical") (e.g., *forêt*_{fem.}). The two members of a pair were matched on gender, word frequency, and approximate length. Congruency was based on the orthographic form of the respective noun endings. Taft and Meunier found shorter gender decision times for nouns with congruent endings than for nouns with incongruent endings.

Holmes and Dejean de la Bâtie (1999) investigated gender decision in native speakers of French and L2 learners of French. For our purposes, only the data of the native speakers are relevant. Holmes and Dejean de la Bâtie found an effect of congruency, that is, the reaction times on gender decisions were shorter for nouns with congruent endings than for nouns with incongruent endings. The endings had a predictive value of 95%, and the items either belonged to the predicted gender class (congruent) or to the opposite gender class (incongruent). The authors also found that the speakers who performed best (i.e., fastest and with the highest percentage of correct responses) showed the smallest influence of the ending.

Desrochers and Paivio (1990) crossed the factors ending and onset in a gender decision task (both with indefinite determiners and linguistic labels as response modes, see Desrochers et

al., 1989). They found shorter reaction times for nouns with predictive endings (predictivity value 97.5%) than for nouns with neutral endings (predictivity value 69.5%), and an effect of the onset (shorter reaction times for consonant-initial nouns than for vowel-initial nouns). The two effects were additive. No effects were obtained when participants had to read the words out loud.

Holmes and Segui (2004) also used a gender decision task, crossing ending and onset. They observed an interaction of ending and onset: A predictive ending resulted in faster gender decision times. Consonant onset also resulted in faster gender decision times, but this latter effect was only observed for nouns with neutral endings (see Barbaud, Ducharme, & Valois, 1981 for related findings from naturally occurring speech errors).

In summary, all studies show facilitation in gender decision for predictive endings, while the evidence concerning the role of noun onset and its potential interaction with the predictivity of the ending is less consistent. This predictivity effect is observed in tasks that require gender retrieval (i.e., gender decision), but not in tasks not requiring explicit gender retrieval (e.g., the reading task used as control condition by Desrochers et al. 1989 and Desrochers & Paivio, 1990).

All studies mentioned until now were conducted in the visual modality in which the entire orthographic representation presumably becomes available simultaneously, and thus word endings have a good chance of affecting gender retrieval. In the present study, we used auditory rather than visual stimuli which raises some interesting new questions. Colé, Pynte, and Andriamamonjy (2003) have claimed:

“As far as the auditory modality is concerned, an early influence of gender-related regularities is much more problematic. Speech stimuli are delivered over time, and for many words recognition occurs before the ending has had a chance to be processed (Marslen-Wilson & Welsh, 1978). In such circumstances, it can hardly be argued that final phonemes are used to predict the gender of the stimulus.” (p. 409)

Thus, it is important to put to an empirical test the question of whether gender retrieval for spoken words can be influenced by their endings.

Word form regularities in language production

Current models of language production assume that there are different representational levels for coding a word's semantics, syntax, and phonology. We will refer to these levels as conceptual level, lemma level, and word form level, respectively (see Levelt, 1999). Most researchers agree that there is a first step of lemma access and a second step of word form access. However, it is still a controversial issue (see Vigliocco & Hartsuiker, 2002, for a review) whether there is feedback from the word form level to the lemma level (interactive model) or not (non-interactive model).

For our present purpose, the opposition between non-interactive models (i.e., models without feedback from word form level to lemma level) and interactive models plays a crucial role. Non-interactive models assume that gender retrieval is a process that takes place at the lemma level. Because there is no feedback from the word form level to the lemma level (e.g., Levelt, 1999; Levelt, Roelofs, & Meyer, 1999), phonological information cannot affect gender retrieval. By contrast, interactive models of language production (e.g., Dell, 1986; Stemberger, 1985) allow for feedback from the word form level to the lemma level. Therefore, the phonological predictivity of grammatical gender could in principle affect the retrieval of grammatical gender. If, for example, a word like *nuage* (cloud_{mas.}) is accessed, activation will spread from the lemma of the noun *nuage* to the corresponding phonemes that make up its word form. These phonemes will then feed activation back to all lemmas sharing the phonemes of the ending *-age*. As most of the nouns with this ending have masculine gender, the syntactic property *masculine* which is represented at the lemma level will receive activation from most of the lemmas activated via the feedback loop while the syntactic property *feminine* will only become activated by the few feminine nouns that end in *-age*. Therefore, selection of the gender feature for the masculine noun *nuage* will be easier than the selection of the gender feature for a noun

that also ends in *-age* but has feminine gender (e.g., *plage*, *beach_{fem.}*). It should be noted in this context, that most interactive models do not make explicit predictions on the potential role of phonology in gender retrieval as they are based on data from English, which does not have grammatical gender. Thus, the prediction we derived for interactive models should be considered as a logical extrapolation².

To recapitulate: A non-interactive model predicts that phonology has no effect on gender retrieval. Therefore, if we find that, in production, a gender-marked element can be retrieved faster for a noun like *nuage* than for a noun like *plage*, the result would not be reconcilable with a strict-serial, non-interactive model and would instead support interactive models. By contrast, the absence of an effect of phonological congruency would allow for two different conclusions. First, it could mean that phonological information is not fed back to the level of gender retrieval, which would be in line with a non-interactive model. Second, it could mean that phonological information is fed back but does not influence the gender selection process. Indeed, for reasons of timing, the feedback from phonological information may be too slow to have an impact on gender processing.

Until now, most studies addressing the interaction between phonological word form level and lemma level in language production focussed on effects of phonology on lemma selection (Cutting & Ferreira, 1999; Damian & Martin, 1999; Dell & Reich, 1981; Ferreira & Griffin, 2003; Roelofs, Meyer, & Levelt, 1996; Starreveld & LaHeij, 1995; 1996). Only a few studies have investigated a potential interaction between the processing of phonology and syntactic properties like grammatical gender in language production.

Some evidence comes from studies of patients or from studies of tip of the tongue (TOT) states. Badecker, Miozzo, and Zanuttini (1995) describe an amomic patient, Dante, who can retrieve the gender of nouns correctly above chance level, while he is unable to access any word

form information. Importantly, the authors report that performance was not affected by phonological regularities (i.e., phonological predictivity of a noun's gender).

Vigliocco, Antonini, and Garrett (1997) found that Italian speakers in a TOT state were able to correctly identify the gender of a noun better than expected by chance. Caramazza and Miozzo (1997; Miozzo & Caramazza, 1997) looked at TOT states in Italian, too, using slightly different baselines, but essentially observing the same results as Vigliocco et al. However, Caramazza and Miozzo argue that in a system where gender is accessed before phonology (and where gender retrieval cannot be skipped), speakers in a TOT state should only be able to correctly recall some of the word form information (e.g., first phoneme) if they have also successfully retrieved the word's syntactic properties. In contrast to this prediction, they did not find a correlation between the retrieval of syntactic properties and the retrieval of parts of the phonological word form in TOT states. Therefore they propose that phonological form and syntactic properties can be accessed independently of each other. Note, that also this account predicts that there should be no effect of the phonological predictivity of an ending on gender retrieval, just as also as predicted by non-interactive models discussed above (see Vinson & Vigliocco, 1999, for further discussion of the (in-)dependence of phonological word form and syntactic properties).

Vigliocco, Franck, Antón-Méndez, and Collina (in revision) investigated the influence of the phonological predictivity of grammatical gender on the computation of gender agreement in a preamble completion task (Bock & Miller, 1991). In their task, participants saw an adjective and the beginning of a sentence (the so-called preamble), for example “*EVIDENT la raison du conflit*” (EVIDENT the cause of the conflict). They had to repeat the preamble and to complete it with the correctly inflected adjective, for example *la raison_{fem.} du conflit_{mas.} est evidente_{fem.}*. In these sentence preambles, the subject noun (e.g., *la raison*) is referred to as head noun and the

modifying noun (e.g., *du conflit*) as local noun. The authors looked at agreement errors of the type *la raison_{fem.} du conflit_{mas.} est evident_{mas.}*, where the adjective takes the gender of the local noun and not of the head noun. In three experiments in Spanish, Italian, and French, Vigliocco et al. manipulated the phonological predictivity of the head noun's ending for grammatical gender (predictive vs. non-predictive). They found that participants made fewer errors when the head noun's ending was predictive than when it was non-predictive. In other words, an adjective linked to a head noun with a predictive ending is more resistant to agreement errors than one linked to a head noun with a non-predictive ending. In a fourth experiment in Spanish, the authors demonstrated that head nouns with incongruent endings yielded more errors than head nouns with congruent endings. Note, however, that the task used by Vigliocco et al. is not a pure language production task (although arguments against this interpretation are put forward by the authors). Rather, it also contains a comprehension component, namely the comprehension of the preamble. This makes it difficult to judge whether the obtained effects are genuine language production effects.

In the following, we will present two experiments which investigate the potential role of phonological congruency of grammatical gender in spoken word recognition (Exp. 1) and in the production of phrases (Exp. 2). Both experiments consist of two tasks. One task provides a baseline of the basic lexical processing time for each item. The second task is designed to be as similar as possible to the baseline task, with one additional requirement, namely, the retrieval of a gender-marked element. In this way, we control for potential differences between the items in the different experimental conditions which should contribute equally to both tasks.

Experiment 1

Experiment 1 was a comprehension experiment. It consisted of a baseline lexical decision task and a gender decision task which required the retrieval of the gender-marked indefinite determiner *un lune*.

We used French words whose grammatical gender was either congruent or incongruent with their phonological endings. If phonological congruency has an effect, gender decisions should be faster for nouns with congruent phonological endings than for nouns with incongruent endings, relative to the reaction time pattern in the lexical decision task.

Method

Participants. Twenty-four native speakers of French, 16 women and 8 men, took part in the experiment. All participants were psychology students at the University of Geneva and received course credit for their participation. Participants were between 19 and 34 years of age, with a mean age of 23 years. Three participants were left-handed.

Material.

Words. We selected 60 experimental items. Thirty items had masculine gender and 30 items had feminine gender. Within a gender class, half of the items had a congruent ending and half of the items had an incongruent ending. The predictive value of an ending was computed using the internet database *lexique* (www.lexique.org). The computation was based on the last two or three phonemes.

In a language like French, one has to differentiate between morphological and phonological predictivity. Concerning morphological predictivity, nouns having the same derivational suffix always have the same gender (e.g., all nouns ending in the derivational suffix “-tion” are feminine). Thus, morphological predictivity is an absolute determinant of gender. For our experiments, we used phonological predictivity as independent variable, which is a

probabilistic rather than an absolute determinant of gender. For some of our experimental items, however, one might consider the ending of the noun having a morphological status (e.g., *-age*, *-ette*). But also these endings can appear in words of both masculine and feminine gender, that is, it displays the probabilistic predictivity of phonological endings and not the deterministic predictivity of morphological endings.

For computing the phonological predictivity value of an ending, we had to decide how many phonemes to include in this computation. Because the nouns used in the present experiments were about five phonemes long on average, we decided to use two phonemes as a standard for the computation of phonological predictivity (with a few exceptions where three phonemes were used). Thus, the predictive part of the word form was on average less than half of the number of phonemes of the whole word.

Two measures of predictivity were calculated, *token predictivity* and *type predictivity*. *Type predictivity* refers to the proportion of nouns of the dominant gender relative to all nouns with this ending (e.g., $n(\text{masculine nouns}) / [n(\text{masculine nouns}) + n(\text{feminine nouns})]$). *Token predictivity* refers to the cumulative frequency of nouns of the dominant gender relative to the cumulative frequency of all nouns with a given ending (e.g., $\sum \text{frequency}(\text{masculine nouns}) / \sum \text{frequency}(\text{masculine nouns} + \text{feminine nouns})$). An ending was defined as predictive if both measures were equal to or larger than .70. Items with the dominant gender for a given ending were chosen as congruent items, items with the non-dominant gender for this same ending were chosen as incongruent items. Only nouns referring to concrete, depictable objects were chosen so that the same items could also be used in the production experiment (Exp. 2). For each ending, one or two pairs of nouns were selected, with one noun of the dominant gender (congruent) and one of the non-dominant gender (incongruent) in each pair.

Using the same type and token measures, we also determined the predictivity of the onset cohorts. Noun onset was defined as either the first syllable of the noun or, for monosyllabic words, all phonemes up to and including the first vowel. For all four conditions, onset predictivities were slightly favouring masculine gender. But as far as the critical comparison between congruent and incongruent items is concerned, masculine nouns had a balanced onset - predictivity (.56 type predictivity, and .62 token predictivity, for congruent items and .63 and .68, respectively, for incongruent items). The same is true for feminine nouns (.45 and .46, respectively, for congruent items and .41 and .41, respectively, for incongruent items). All items with their corresponding predictivity measures for the endings are given in the Appendix.

ANOVAs with the two factors gender (feminine vs. masculine) and congruency (congruent vs incongruent) were carried out on type predictivity and on token predictivity. There was no significant effect of gender on type predictivity, $F(1, 56) = 3.15, p = .082, MSE = 0.006$ or on token predictivity, $F < 1$. The effect of congruency was significant for type predictivity, $F(1, 56) = 1361, p < .001, MSE = 0.006$, and token predictivity, $F(1, 56) = 1554, p < .001, MSE = 0.005$. There was no interaction of these two factors for type predictivity ($F < 1$) or for token predictivity ($F < 1$).

We tried to keep the following variables constant across the four conditions: groupsize (the number of French nouns with a given ending), logarithmic noun frequency (computed from the lexique database “*lemmes*”, frequency per million), theoretical uniqueness point (number of phonemes up to the uniqueness point)³, number of phonemes, number of syllables, and length of the stimulus in milliseconds. The last measure was obtained from the auditory stimuli used in Experiment 1. Table 1 shows the distribution of these variables across conditions.

[Insert Table 1 about here]

ANOVAs with the two factors gender and congruency were carried out on groupsize, logarithmic frequency, uniqueness point, number of syllables, number of phonemes, and length. The only significant effect was an interaction of gender and congruency for logarithmic frequency, $F(1, 56) = 5.52, p < .05, MSE = 0.455$. For all other variables, there were no significant main effects or interactions (all $ps > .10$).

Besides the 60 critical items, 88 more items were included. These items belonged to a different study but can be regarded as filler items in the present experiments. Four differently randomised experimental lists were created, such that successive nouns were neither semantically nor phonologically related and that no more than three successive nouns had the same gender. Four more nouns were used as warm-up items at the beginning of each list. These lists were used in the gender decision task. Furthermore, ten more nouns were chosen to serve as items in a training session.

Non-words. We created 152 non-words from the 152 nouns described above (60 critical items, 88 fillers, four warm-up items). The non-words were derived by splitting the nouns up into their constituent syllables, or, for monosyllabic nouns, into their constituent parts onset, nucleus, and coda. These parts were then randomly rearranged. None of the resulting combinations were existing French words and all were phonologically legal non-words of French, as confirmed by a native speaker of French. For the ten practice items, non-words were created in the same way. The four original experimental lists were interspersed with the non-words such that no more than five words or five non-words occurred in a row and that successive non-words or successive words and non-words were not phonologically similar. Words and non-words were recorded by a female native speaker of French. Participants were randomly assigned to one of the four experimental lists. These new lists were used in the lexical decision task.

Design. Participants performed both a lexical decision task and a gender decision task. Half of the participants started with the lexical decision task and then continued with the gender decision task, for the other half of the participants the order of the two tasks was reversed. Participants received differently randomised lists in the lexical decision and in the gender decision. Six participants were assigned to each of the four lexical decision lists, and six participants were assigned to each of the four gender decision lists.

Apparatus and Procedure. The experiment was controlled using the experimental software E-Prime 1.1, Psychology Software Tools. Experiments were run on a Dell Dimension XPS T500 computer. Visual instruction texts were presented on a Dell P990 monitor. Auditory stimuli were presented to the participants using Beyer dynamic DT 100 headphones.

Participants were tested individually in a soundproof booth. The instructions were presented in white letters on a dark blue screen. In the lexical decision task, participants were instructed to decide as quickly as possible whether an auditory stimulus was a French word or not by pressing a button on a button box. Half of the participants pressed the right button for words and the left button for non-words, for the other half this assignment was reversed. In the gender decision task participants were instructed to decide as quickly as possible whether an auditory stimulus required the indefinite determiner *un* (masculine) or *une* (feminine) by pressing a button on a button box. We used the *un - une* decision because it is a more natural task than the meta-linguistic assignment of the labels 'masculine' and 'feminine' (see Taft & Meunier, 1998) and because participants react more rapidly (Desrochers et al., 1989). Half of the participants were told to press the right button for *un* and the left button for *une*, for the other half this assignment was reversed.

When the participants had no more questions, the experimenter started the experiment. The instruction was replaced by a dark blue screen. At the beginning of each trial, participants

heard a 1kHz warning-tone for 250 ms, followed by the auditory target. From target onset onwards they could react for 2750 ms. Reaction times were measured from word onset. 2750 ms after target onset, the next trial started with the warning tone. Before the lexical decision task, participants had a practice phase with ten words and ten non-words. Before the gender decision task, participants had a practice phase with ten words.

Results

Reaction times from word onset

Reaction times for incorrect responses and reaction times that deviated by more than two standard deviations from a participant's and an item's mean were replaced by the corresponding cell mean. 4.8% of the reactions were classified as errors (6.4% in the lexical decision and 3.2% in the gender decision), and 1.7% as outliers (1.6% in the lexical decision and 1.8% in the gender decision). Table 2 shows the mean reaction times and mean error rates as a function of gender and congruency. Standard deviations are given in parentheses.

[Insert Table 2 about here]

We conducted ANOVAs with the factors task (lexical decision vs. gender decision), gender (masculine vs. feminine), and congruency (congruent vs. incongruent). Task, gender, and congruency were within-participant factors. Task was a within-item factor, gender and congruency were between-item factors. There was a significant main effect of task, $F_1(1, 23) = 118.6, p < .001, MSE = 8277$; $F_2(1, 56) = 221.83, p < .001, MSE = 2766$, with faster reactions in the lexical decision task than in the gender decision task. The main effects of gender and congruency were significant, $F_1(1, 23) = 27.34, p < .001, MSE = 3260$; $F_2(1, 56) = 3.93, p < .05$,

$MSE = 14186$, and $F_1(1, 23) = 64.12, p < .001, MSE = 1872$; $F_2(1, 56) = 5.29, p < .05, MSE = 14186$, respectively. The interaction of gender and congruency was significant in the participant analysis, $F_1(1, 23) = 12.09, p < .001, MSE = 2553$, but not in the item analysis, $F_2(1, 56) = 1.36, p > .20, MSE = 14186^4$.

[Insert Figure 1 about here.]

Most importantly, the interaction of task and congruency was significant, $F_1(1, 23) = 18.62, p < .001, MSE = 2272$; $F_2(1, 56) = 9.56, p < .01, MSE = 2766$, reflecting the fact that the difference between gender decision reaction times and lexical decision reaction times is larger for nouns with incongruent endings than for nouns with congruent endings. Panel a of Figure 1 illustrates this point by presenting the difference between the reaction times for gender decision and the lexical decision.

Finally, the interaction of task and gender, $F_1(1, 23) = 1.84, p > .15, MSE = 2821$; $F_2(1, 56) = 1.17, p > .25, MSE = 2766$, and the three-way interaction of task, gender, and congruency, both $F_s < 1$, were not significant.

Table 1 suggests that there might already be an effect of congruency in the lexical decision task. However, the interaction of task and congruency shows that the effect of the congruency of the ending is significantly stronger in gender decision than in lexical decision. Note that this result also indicates the importance of a baseline like lexical decision as a control for potential differences between the items in the different experimental conditions.

For lexical decision, the effect of gender was significant in the participant analysis, $F_1(1, 23) = 22.23, p < .001, MSE = 1155$, but not in the item analysis, $F_2(1, 56) = 3.11, p = .083, MSE = 5157$. The same was true for the effect of congruency, $F_1(1, 23) = 10.07, p < .01, MSE = 984$;

$F_2(1, 56) = 1.2, p > .25, MSE = 5157$, and the interaction of gender and congruency, $F_1(1, 23) = 10.29, p < .01, MSE = 1245$; $F_2(1, 56) = 1.55, p > .21, MSE = 5157$.

For gender decision, the effect of gender was significant in the participant analysis, $F_1(1, 23) = 13.94, p < .001, MSE = 4926$, but not in the item analysis, $F_2(1, 56) = 3.64, p = .062, MSE = 11795$. The effect of congruency was significant, $F_1(1, 23) = 48.24, p < .001, MSE = 3160$; $F_2(1, 56) = 8.08, p < .01, MSE = 11795$. The interaction of gender and congruency was only significant by participants, but not by items, $F_1(1, 23) = 5, p < .05, MSE = 3657$; $F_2 < 1$.

Reaction times from word offset

It has been argued (e.g., Lipinski & Gupta, 2005; Vitevitch & Luce, 2005) that for auditory stimuli, reaction times measured from stimulus offset rather than onset provide a better measure because they take into account differences in stimulus length. Therefore, we subtracted the length of a given stimulus from the reaction time to this stimulus for each participant. We re-entered these new values into the same ANOVA carried out above. Table 3 shows the distribution of reaction times across conditions.

[Insert Table 3 about here]

There was a significant main effect of task, $F_1(1, 23) = 102.66, p < .001, MSE = 8764$; $F_2(1, 56) = 193.77, p < .001, MSE = 2902$, with faster reactions in the lexical decision task than in the gender decision task. The main effects of gender and congruency were significant by participants, but not by items, $F_1(1, 23) = 5.04, p < .05, MSE = 3444$; $F_2 < 1$, and $F_1(1, 23) = 42.67, p < .001, MSE = 2152$; $F_2(1, 56) = 2.77, p > .10, MSE = 20711$, respectively. The interaction of gender and congruency was significant in the participant analysis, $F_1(1, 23) = 38.3, p < .001, MSE = 1778$, but not in the item analysis, $F_2(1, 56) = 2.06, p > .17, MSE = 20711$.

As in the analysis for reaction times from word onset, the crucial interaction of task and congruency was also significant for reaction times measured from word offset, $F_1(1, 23) = 10.44,$

$p < .001$, $MSE = 2488$; $F_2(1, 56) = 5.6$, $p < .05$, $MSE = 2902$. Panel b of Figure 1 presents the difference values between the gender decision and the lexical decision. The interaction of task and gender, $F_s < 1$, and the three-way interaction of task, gender, and congruency, $F_s < 1$, were not significant.

Errors

The same ANOVAs as on the reaction times were conducted on the error rates. The effect of task was significant by participants and showed a tendency towards significance by items, $F_1(1, 23) = 18.81$, $p < .001$, $MSE = 0.003$; $F_2(1, 56) = 3.67$, $p = .061$, $MSE = 0.008$, with higher error rates in the gender decision task than in the lexical decision task. The main effect of gender showed a tendency towards significance by participants, $F_1(1, 23) = 3.44$, $p = .076$, $MSE = 0.003$, but not by items, $F_2 < 1$. There was a significant effect of congruency, $F_1(1, 23) = 35.64$, $p < .001$, $MSE = 0.004$; $F_2(1, 56) = 7.24$, $p < .01$, $MSE = 0.012$. The interaction of gender and congruency was significant in the participant analysis, $F_1(1, 23) = 20.49$, $p < .001$, $MSE = 0.002$, but not in the item analysis, $F_2(1, 56) = 2.21$, $p > .10$, $MSE = 0.012$.

The interaction of task and congruency was significant by participants, $F_1(1, 23) = 7.27$, $p < .01$, $MSE = 0.002$, but not by items, $F_2(1, 56) = 1.36$, $p > .20$, $MSE = 0.008$. The same holds for the interaction of task and gender, $F_1(1, 23) = 5.09$, $p < .05$, $MSE = 0.002$; $F_2 < 1$. The three-way interaction of task, gender, and congruency was not significant, both $F_s < 1$.

Discussion

Experiment 1 reveals a congruency effect for the gender decision task. Participants are faster in making a gender decision for auditorily presented nouns with congruent endings than for nouns with incongruent endings. A weaker congruency effect was also observed in the lexical decision task. These findings were independent of the reference point used for the reaction time

measurements (word onset and word offset). In both tasks, however, the congruency effect was larger for masculine nouns when measured from onset and for feminine nouns when measured from offset. Finally, when using lexical decision as the baseline to eliminate potential inter-item differences and taking RT differences between lexical decision and gender decision tasks (see Figure 1), a clear congruency effect of approximately the same size for both gender classes is obtained. Importantly, this pattern looks almost identical for reaction times measured from word onset and from word offset; the critical interaction of task by congruency is significant for both measurement points while the triple interaction of task, gender, and congruency is not significant for either measure.

The correlation between mean item reaction times (over participants) for gender decision and lexical decision was .75, indicating that there is considerable overlap in the processing components required in the two tasks and that the lexical decision is indeed an appropriate baseline for controlling potential item-related differences. The pattern in the error rates largely follows the pattern of the reaction times.

The present experiment extends the previous findings in the visual modality discussed in the introduction by showing that an effect of gender congruency with a phonological ending is also obtained with auditory stimuli. This result is remarkable for two reasons. First, Taft and Meunier (1998) suggest that orthographic endings are more informative than phonological endings, because there exist a number of homophonic endings for both gender classes that are only disambiguated by their spelling. However, the present data show that also phonological cues can influence the retrieval of gender-marked elements. Second, whereas the stimulus information in a written noun is processed in parallel, auditory information is processed sequentially. As noted in the introduction, Colé and colleagues (2003) have claimed that given the earliness of spoken word recognition, gender information is retrieved even before the gender information

contained in the word ending is reached. In contrast to this hypothesis, the present data show that the ending of an auditorily presented noun can assist in retrieving the corresponding gender-marked indefinite determiner in comprehension.

It is interesting, in light of the hypothesis put forward by Colé et al. to look at the theoretical uniqueness point and the length of the nouns in the materials of the present experiment. The nouns had a mean length of 5.2 phonemes and their theoretical uniqueness point was reached on average after 4.8 phonemes. More to the point, in 55 of 60 nouns, the onset of the ending was earlier than the theoretical uniqueness point, in three cases, the onset of the ending and the uniqueness point coincided, and only for two nouns, the onset of the ending was after the uniqueness point. Thus, the present results are not necessarily in conflict with the claim made by Colé et al. (2003), because for the majority of nouns used in this study, the uniqueness point fell within the phonological ending. The distance between the position of the uniqueness point and the phonological ending of a noun may well be an important determinant of whether effects of phonological congruency on gender retrieval can be observed.

Another possible explanation for why we observe an influence of phonological congruency on gender retrieval lies in the duration of the observed reaction times. As Table 3 suggests, reaction times in the gender decision task occurred well after the end of the word, on average 260 ms after the end of the stimulus. Even if we allow for a period of 100 – 200 ms for initiating the motor response, participants most likely did process the whole word form before making their response and could therefore make use of the information provided by the noun ending.

Experiment 1 showed a role of phonological congruency on the retrieval of gender-marked elements in a gender decision task. Experiment 2 now addresses the question whether an analogous effect is also obtained for language production.

Experiment 2

Experiment 2 was a picture naming experiment. Participants either named pictures with bare nouns (baseline) or with a NP consisting of a gender-marked definite determiner, a gender-marked prenominal adjective, and a noun (i.e. *le grand* [*noun_{masc.}*] or *la grande* [*noun_{fem.}*]).

Method

Participants. Thirty-four participants took part in the experiment, all of them native speakers of French. All participants were students or scientific employees at the University of Geneva. They were between 20 and 36 years of age, mean age was 25. Three participants had to be excluded from all further analyses because they had made more than 20% errors.

Material and Design. The same lists were used as in the gender decision task of Experiment 1. The auditory stimuli were replaced by black on white line drawings corresponding to the picture names. Most pictures were taken from a picture database of the Max-Planck-Institute for Human Cognitive and Brain Sciences in Leipzig (Germany) and the remaining pictures were chosen from the google image database (www.google.com, images). Some pictures were simplified in order to have a set of pictures with comparable visual complexity. Pilot studies revealed that some pictures were named incorrectly very often. These items and their matched counterparts were excluded from the lists. The excluded items were two picture pairs with feminine-congruent phonological endings (*cloporte* - *tarte*; *grenouille* - *fenouil*) and two picture pairs with masculine-congruent endings (*mouchoir* - *bouilloire*; *dynamo* - *plumeau*)⁵. From the 88 filler items, another 8 items were excluded, reducing the complete item set to 132 items.

Apparatus and Procedure. The experiment was controlled using the experimental software E-Prime1.1, Psychology Software Tools. Experiments were run on a Dell Dimension XPS T500 computer. Visual stimuli were presented on a Dell P990 monitor. The voice key was triggered

using an ATR20 microphone. Participants' responses were recorded on a Sony ZA5ES DAT-recorder using head-worn microphones (Shure, SM10).

Before the experiment started, the voice key was tested using an E-Prime procedure. Participants had to name 20 randomly presented numbers. If all responses triggered the voice key, the experiment was started with this trigger level of the voice key, if not, the voice key trigger level was adjusted and the voice key test run again with the adjusted level.

The voice key test was followed by a familiarisation phase during which participants saw all pictures with the name of the picture being presented auditorily at picture onset. Participants were instructed to use these picture names during the main experiment. The familiarisation phase was self-paced. After the familiarisation phase, participants saw a written instruction on the screen which told them to name the pictures either with the corresponding (bare) noun or with an NP consisting of the definite determiner, the gender-marked adjective for 'big', and the noun (e.g., *la grande table*). Each participant received two experimental lists. Half of the participants were instructed to name the pictures with bare nouns in the first list followed by the second list in which they had to name the pictures with the phrase *le grand/ la grande [noun]*. For the other half of the participants, this order was reversed.

Speed and accuracy were both stressed in the instructions. Both blocks of the main experiment started with ten warming-up trials. At the beginning of each trial, a fixation cross was presented on the screen for 800 ms. It was followed directly by the picture that stayed on the screen for 2000 ms. Unknown to the participant, at picture onset a 1kHz-tone was recorded on the second channel of the DAT-recorder as a marker of stimulus onset for later measurements. The two tasks (bare noun naming and determiner adjective noun naming) differ in the variability of the onset phonemes of the responses. In the NP condition, all responses start with the same phoneme (/l/), whereas the responses in the bare noun condition start with different phonemes.

This could affect the observed reaction times in two ways: First, participants could already phonologically prepare their responses in the NP condition, giving them a head start for pronunciation. This is not problematic as it would only contribute to the overall difference between the two tasks, but it would not differentially affect the reaction time pattern within a task. Second, due to the constraints on material selection, we could not equate the distribution of noun onsets in the different conditions. Therefore, in bare noun naming, differences in the sensitivity of the voice key to different noun onsets could induce artificial differences between conditions. Therefore, we used the DAT-recordings to re-measure the onset latencies (i.e., the time between the onset of the 1kHz-tone (stimulus onset) and the beginning of the utterance) in the bare noun condition manually, using the software praat (www.praat.org).

Participants were asked to produce the sequence of determiner and gender-marked adjective before the noun in order to maximise the chance of finding a gender effect. In a pilot study, we had not seen an effect with simple determiner NPs. Due to the frequent association between a noun and its determiner, there might be a ceiling effect in the selection of the definite determiner. Correct inflection of the gender-marked adjective requires computation of grammatical gender above and beyond simple association.

Results

Reaction times

Disfluencies, self-corrections, use of a wrong name and no response at all were coded as errors. Reaction times that deviated more than two standard deviations from a participant's and an item's mean were replaced by the corresponding cell mean. 10% of the reactions were classified as errors, and 1.6% as outliers. Table 4 shows the mean reaction times and mean error rates split up

by utterance type (Bare noun vs. NP), gender, and congruency. Standard deviations are given in parentheses.

[Insert Table 4 about here]

We conducted ANOVAs with the factors utterance type (bare noun vs. determiner adjective NP), gender (masculine vs. feminine), and congruency (congruent vs. incongruent). Utterance type, gender, and congruency were within-participant factors. Utterance type was a within-item factor, gender and congruency were between-item factors. The main effect of utterance type was not significant in the participant analysis, $F_1(1, 30) = 1.26, p = .271, MSE = 37107$, but it was significant in the item analysis, $F_2(1, 48) = 7.35, p < .01, MSE = 2941$. The main effect of gender was not significant, both $F_s < 1$. There was a significant main effect of congruency by participants, but not by items, $F_1(1, 30) = 45.86, p < .001, MSE = 3371, F_2(1, 48) = 2.47, p > .10, MSE = 2308$. The interaction of utterance type and gender was significant in the participant analysis, $F_1(1, 30) = 5.42, p < .05, MSE = 4255$, but not in the item analysis, $F_2(1, 48) = 2.14, p = .15, MSE = 2941$. The interactions of utterance type and congruency, and of utterance type, gender, and congruency were not significant (all $p_s > .10$).

The absence of a significant interaction of utterance type and congruency reveals that any effects of congruency that are present in determiner adjective NP naming are also present in bare noun naming. While these congruency effects are descriptively slightly larger in determiner adjective NP naming (75 ms and 37 ms for masculine and feminine gender, respectively) than in bare noun naming (54 ms and 12 ms for masculine and feminine gender, respectively), the statistical analysis shows that this difference is not reliable. Figure 2 illustrates the data pattern.

[Insert Figure 2 about here].

While the interaction of utterance type and congruency was not significant, the data suggest that there might be a congruency effect in both bare noun naming and determiner adjective NP naming. Separate analyses for bare noun naming and determiner adjective NP naming confirmed that in both tasks there was a tendency towards a congruency effect. For bare noun naming, the main effect of congruency was significant in the participant analysis, $F_1(1, 30) = 9.58, p < .01, MSE = 3486$, but not in the item analysis, $F_2 < 1$. The interaction of gender and congruency was significant by participants, $F_1(1, 30) = 5.67, p < .05, MSE = 2313$, but not by items, $F_2 < 1$. The main effect of gender was not significant, $F_s < 1$. For determiner adjective NP naming, the main effect of congruency was significant in the participant analysis, $F_1(1, 30) = 32.4, p < .001, MSE = 3039$, and marginally significant in the item analysis, $F_2(1, 48) = 3.28, p = .077, MSE = 11331$. The interaction of gender and congruency was not significant, $F_1(1, 30) = 2.95, p = .096, MSE = 3780, F_2 < 1$. The main effect of gender was significant in the participant analysis, $F_1(1, 30) = 4.08, p < .05, MSE = 7497$, but not in the item analysis, $F_2 < 1$.

Errors

The same ANOVAs as on the reaction times were conducted on the error rates. There was a significant effect of utterance type in the error analyses, $F_1(1, 30) = 10.64, p < .01, MSE = 0.009$; $F_2(1, 48) = 19.73, p < .001, MSE = 0.002$. None of the interactions of utterance type and any of the other factors was significant.

Discussion

In contrast to Experiment 1 where the critical task (lexical decision vs. gender decision) by congruency interaction was significant, the corresponding interaction of utterance type (bare

noun vs. NP) by congruency interaction in the present experiment was not significant. This difference between the experiments is also obvious in the descriptive data pattern. In Experiment 1, the congruency effect for masculine nouns increases by 65 ms when going from the lexical decision task (43 ms congruency effect) to the gender decision task (108 ms congruency effect). For feminine nouns, this increase of the congruency effect has about the same size with 54 ms (-2 ms congruency effect for lexical decision vs. 52 ms congruency effect in gender decision). In Experiment 2, by contrast, the increase in the congruency effect when going from bare noun to NP production is much smaller, with 21 ms for masculine nouns (54 ms congruency effect for bare nouns vs. 75 ms congruency effect for NPs), and 25 ms for feminine nouns (12 ms congruency effect for bare noun vs. 37 ms congruency effect for NPs). Thus, while there was a significant difference in the congruency effect between lexical decision and gender decision in Experiment 1, the same materials did not show a corresponding difference in the congruency effect between bare noun and NP production (Exp. 2).

The correlation of mean item reaction times (over participants) between determiner adjective noun naming and bare noun naming was $r = .81$, indicating that bare noun naming is an adequate baseline condition for controlling potential differences between conditions due to the different items used.

So far, we have been focusing on the question whether a potential congruency effect is larger in a task that requires the retrieval of gender-marked elements (gender decision in Exp. 1, NP production in Exp. 2) than in some baseline task (gender decision in Exp. 1, bare noun production in Exp. 2), as shown by the interaction of congruency and task. But if we do not break the data by task (baseline vs. experimental task), there is an effect of congruency in Experiment 1 and a marginally significant effect of congruency in Experiment 2. Our conclusion is that the effect in the baseline tasks is due to item differences and only the interaction of task by

congruency has a meaningful interpretation. Given that the congruency effect was not larger in determiner adjective NP naming than in bare noun naming, we conclude that phonology did not affect gender retrieval in language production.

Two recent studies suggest that gender might also be retrieved during bare noun production, at least in other languages than in French. Plemmenou, Bard, and Branigan (2002) used a priming paradigm in Greek. Participants named black line drawings with a bare noun and coloured line drawings with a gender-marked colour adjective. The results showed that gender-marked colour adjectives were produced faster when the noun produced at the preceding trial (which was different from the target noun) had the same gender than when it had different gender. This gender priming by way of bare nouns suggests that gender is retrieved in the production of bare nouns.

In a study by Cubelli, Lotto, Paolieri, Girelli, and Job (2005), native speakers of Italian named pictures while ignoring written distractor words that were superimposed on the pictures. The results showed that it took longer to produce a bare noun when the distractor was of the same gender than when it was of different gender. Again, this result suggests that gender is not only retrieved when a gender-marked element has to be produced, but also when a bare noun is produced. However, the effect obtained by Cubelli et al. is in the opposite direction of the effects that are usually obtained with such tasks (e.g., Schriefers, 1993), and their results are in conflict with La Heij, Mak, Sander, and Willeboordse (1998) who showed an effect for NPs with a gender-marked determiner while no effect was obtained for bare noun production.

Do the studies by Plemmenou et al. (2002) and Cubelli et al. (2005) imply that the effect of phonological congruency in bare noun naming in the present Experiment 2 (significant by participants, but not by items) should also be considered as a reflection of gender retrieval rather than item differences between the phonologically congruent and incongruent conditions? Here, it

is important to note that both Greek and Italian nouns are morphologically marked for gender. Cubelli et al. explicitly argue that the effect they observe is due to this morphological gender-marking. This implies that when speakers produce a bare noun in Italian or Greek, they have to retrieve the noun stem and a gender-marked element, namely the gender-marked morpheme. Thus, the situation in Greek and Italian is clearly different from the French materials used in the present study in which gender was not marked by a suffix of the noun.

No matter how one interprets the effects in the baseline, we focus on the finding that there is no task by congruency interaction in Experiment 2. That is, in a condition where participants need to retrieve gender because they have to produce a gender-marked determiner and adjective, phonological congruency has no more effect than in a baseline condition where the retrieval of gender is in principle superfluous.

General Discussion

Experiment 1 extends earlier results on the role of orthographic predictivity on gender retrieval in language comprehension (Desrochers et al., 1989; Desrochers & Paivio, 1990; Holmes & Dejean de la Bâtie, 1999; Holmes & Segui, 2004; Taft & Meunier, 1998). All previous studies used visual stimulus presentation and manipulated the orthographic predictivity of gender. By contrast, in the present study auditory stimuli were used and the phonological predictivity of a noun's ending was manipulated. The observation of an effect of phonological predictivity in spoken word recognition is not trivial as it has been argued (Colé et al., 2003) that the sequential nature of spoken word recognition precludes such an effect.

Experiment 1 also provides the background against which to interpret the results found in the production experiment (Exp. 2). In Experiment 2, there is a weak (i.e., only significant by participants) effect of congruency. However, the interaction between task (bare noun naming vs. adjective determiner noun naming) and congruency was not significant, indicating that the

congruency effect was as big in the baseline task that did not, in principle, require gender retrieval as in the experimental task. Given that the critical task by congruency interaction was significant in Experiment 1, the absence of the corresponding interaction in Experiment 2 is not due to the selection of the materials and an insufficient difference in the phonological predictivity between the congruent and incongruent conditions.

As we have argued in the Introduction, the presence of a task by congruency interaction in Experiment 2 would have been clear evidence for feedback models of language production. The absence of such an interaction (as is always the case with null-results) is more difficult to interpret. In the following, we will discuss possible interpretations in turn.

Methodologically, subtracting the results of a baseline task from those of an experimental task with gender retrieval might have hidden a more general effect of phonological congruency that is already present in the baseline. We cannot completely exclude this possibility because of the studies mentioned in the discussion of Experiment 2 (Cubelli et al., 2005; Plemmenou et al., 2002) which found that gender is retrieved even for bare nouns. However, as we have discussed before, these effects have been found in languages where the production of a bare noun already entails retrieval of a gender-marked element, namely a gender-marked morpheme. This is not the case for French.

Our language production results are in line with feedforward models, but they do not necessarily exclude feedback models. The empirical demonstration of phonological feedback effects on gender retrieval in language production requires two conditions to be met. First, the critical phonological information must become available at a point in time at which gender retrieval has not yet been completed. Second the critical phonological information must have the adequate distribution in the language (i.e., the correlation of the word form with noun gender must be strong enough in order to act as a cue for gender retrieval). The second criterion

obviously is met by the present experiments: The effect of phonological congruency of noun gender in Experiment 1 empirically proves that, in our materials, the phonology of the nouns' endings is distributed such that phonological information can affect gender retrieval in language comprehension.

It is more difficult to determine, given our data pattern, whether this phonological information was available before gender retrieval processes were completed. Language production starts with a conceptual representation which has eventually to be mapped onto a form representation which eventually forms the input to articulation. Thus, phonological information becomes available relatively late. Strict-serial models of language production (e.g., Levelt et al., 1999) assume that access to a word's lemma information, including grammatical gender, precedes access to the word's phonological form. Therefore, gender retrieval will be completed before potentially relevant phonological information is available. By contrast, theoretically, interactive models (e.g., Dell, 1986) allow for activation of the phonological form of a noun before gender retrieval of this noun is completed, and thus, phonology could in principle affect retrieval of gender-marked elements by phonology-to-lemma feedback. However, potential feedback could actually be too slow to affect gender retrieval. At first sight, this possibility seems to be at odds with the fact that interactive models assume that feedback is fast enough to affect lemma selection processes, as, for example, reflected in the lexical bias effect in speech errors (Dell & Reich, 1981; Hartsuiker, Corley, & Martensen, 2005). This seeming contradiction can be resolved by taking a closer look at how grammatical gender might be represented. Levelt (1999; see also Levelt et al., 1999) assumes that the gender of a noun is represented as an abstract node at the lemma level, with all nouns of a given gender being connected to the same gender node. Feedback from the word form level to the lemma level is a one-step feedback. In contrast, in order to reach the gender node and thus to affect gender retrieval, feedback has to traverse one

additional link, the link from the noun lemma to its gender node. Or, as Vinson and Vigliocco (1999) put it, phonemes and syntactic properties like gender are linked indirectly by way of the lemma. Our data seem to be in conflict with the findings of Vigliocco et al. (in revision) who observed an influence of phonological predictivity of a noun ending in a sentence completion task (see introduction). How can these findings be reconciled with the present results? Here we have to keep in mind that the preamble completion task combines comprehension and production processes, that is, the critical noun has already been presented during preamble comprehension, before the production processes start. This implies that participants have early information about the noun's phonological form and can use it in their computation of agreement.

Recently, Holmes and Segui (2006) reported two production experiments in French, which appear to show that phonological information influences gender retrieval in language production. In their first experiment, Holmes and Segui presented pictures of objects with names that had either congruent or incongruent endings. The participants were instructed to think of the picture name and then decide whether it was masculine or feminine by pressing a button. Reaction times were faster for nouns with congruent endings than for nouns with incongruent endings. In their second experiment, they asked participants to translate English words into French. Participants were instructed to do so silently. After they had translated the word, either a semantic decision question (abstract or concrete?) or a gender decision question (masculine or feminine?) appeared on the screen. Again, participants were faster to make a gender decision for nouns with congruent endings.

However, one has to keep in mind that the tasks used by Holmes and Segui (2006) did presumably encourage the participants to compute the phonological form first, and to decide on gender only in a second step. For example, their task used in Experiment 2 can be seen as a

gender decision task on self-generated stimuli in which participants are generating a phonological form and then conduct a gender decision task based on this form.

This interpretation also fits well with the data by Vigliocco et al. (in revision): Phonological information does not affect gender retrieval in language production unless it is given a head start.

To conclude, it appears that the phonological predictivity of a noun's ending can affect the retrieval of gender-marked elements if the relevant phonological information becomes available before syntactic retrieval is completed. This is the case when determining a gender-marked element for an auditorily presented noun (Exp. 1), at least as long as the uniqueness point of the noun is not reached clearly before the noun's ending. This is also the case in Vigliocco et al.'s preamble completion task and the experiments reported by Holmes and Segui (2006). In contrast, in a pure production task like picture naming, phonological information is available too late to have an impact on the retrieval of gender-marked elements. This is compatible with the claims made by non-interactive models of language production. However, much of the evidence from the literature discussed here suggests that it is the impact of potential feedback underlies strict timing constraints – that is, phonology can only impact gender retrieval if it is available early enough. Further research is needed in order to investigate which timing parameters make effects from phonology on gender retrieval possible and which timing parameters do not. Such an investigation would shed some welcome light on how activation spreads within a given model architecture.

References

- 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.
- Barbaud, P., Ducharme, C., & Valois, D. (1981). L'usage du genre en canadien-français: Une étude syntaxique et sociolinguistique de la féminisation des noms à initiale vocalique [The usage of gender in Canadian-French: A syntactic and sociolinguistic study of the feminisation of vowel-initial nouns]. *Recherches Linguistiques à Montréal*, *17*, 1-42.
- Bock, J. K., & Miller, C. A. (1991). Broken agreement. *Cognitive Psychology*, *23*, 45-93.
- Caramazza, A., & Miozzo, M. (1997). The relation between syntactic and phonological knowledge in lexical access: Evidence from the 'tip-of-the-tongue' phenomenon. *Cognition*, *64*, 309-343.
- Colé, P., Pynte, J. & Andriamamonjy, P. (2003). Effect of grammatical gender on visual word recognition: Evidence from lexical decision and eye movement experiments. *Perception & Psychophysics*, *65*, 407-419.
- Cubelli, R., Lotto, L., Paolieri, D., Girelli, M., & Job, R. (2005). Grammatical gender is selected in bare noun production: Evidence from the picture-word interference paradigm. *Journal of Memory and Language*, *53*, 42-59.
- Cutting, J. C., & Ferreira, V. S. (1999). Semantic and phonological information flow in the production lexicon. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *25*, 318-344.
- Damian, M. F., & Martin, R. C. (1999). Semantic and phonological codes interact in single word production. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *25*, 345-361.

- Dell, G. S. (1986). A spreading-activation theory of retrieval in sentence production. *Psychological Review*, 93, 283-321.
- Dell, G. S., & Reich, P. A. (1981). Stages in sentence production: An analysis of speech error data. *Journal of Verbal Learning and Verbal Behavior*, 20, 611-629.
- Desrochers, A., Paivio, A., & Desrochers, S. (1989). L'effet de la fréquence d'usage des noms inanimés et de la valeur prédictive de leur terminaison sur l'identification du genre grammatical [The effect of frequency of usage of inanimate nouns and the predictive value of their ending on the identification of grammatical gender]. *Revue Canadienne de Psychologie*, 43, 62-73.
- Desrochers, A. & Paivio, A. (1990). Le phonème initial des noms inanimés et son effet sur l'identification du genre grammatical [The initial phoneme of inanimate nouns and its effect on the identification of grammatical gender]. *Revue Canadienne de Psychologie*, 44, 44-57.
- Ferreira, V. S., & Griffin, Z. M. (2003). Phonological influences on lexical (mis)selection. *Psychological Science*, 14, 86-90.
- Hartsuiker, R. J., Corley, M., & Martensen, H. (2005). The lexical bias effect is modulated by context, but the standard monitoring account doesn't fly: Related reply to Baars et al. (1975). *Journal of Memory and Language*, 52, 58-70.
- Holmes, V. M. & Dejean de la Bâtie, B. (1999). Assignment of grammatical gender by native speakers and foreign learners of French. *Applied Psycholinguistics*, 20, 479-506.
- Holmes, V. M., & Segui, J. (2006). Assigning grammatical gender during word production. *Journal of Psycholinguistic Research*, 35, 5-30.
- Holmes, V. M., & Segui, J. (2004). Sublexical and lexical influences on gender assignment in French. *Journal of Psycholinguistic Research*, 33, 425-457.

- Köpcke, K.-M., & Zubin, D. A. (1983). Die kognitive Organisation der Genuszuweisung zu den einsilbigen Nomen der deutschen Gegenwartssprache [The cognitive organization of gender assignment in monosyllabic nouns of contemporary German]. *Zeitschrift für germanistische Linguistik*, *11*, 166-182.
- La Heij, W., Mak, P., Sander, J., & Willeboordse, E. (1998). The gender-congruency effect in picture-word tasks. *Psychological Research*, *61*, 209-219.
- Levelt, W. J. M. (1999). Models of word production. *Trends in Cognitive Science*, *3*, 223-232.
- Levelt, W. J. M., Roelofs, A., & Meyer, A. S. (1999). A theory of lexical access in speech production. *Behavioral and Brain Sciences*, *22*, 1-38.
- Lipinski, J., & Gupta, P. (2005). Does neighborhood density influence repetition latency for nonwords? Separating the effects of density and duration. *Journal of Memory and Language*, *52*, 171-192.
- Miozzo, M. & Caramazza, A. (1997). Retrieval of lexical-syntactic features in tip-of-the-tongue states. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *23*, 1410-1423.
- Plemmenou, E., Bard, E. G., & Branigan, H. P. (2002). Grammatical gender in the production of single words: Some evidence from Greek. *Brain and Language*, *81*, 236-241.
- Radeau, M., Mousty, P., & Bertelson, P. (1989). The effect of the uniqueness point in spoken-word recognition. *Psychological Research*, *51*, 123-128.
- Roelofs, A., Meyer, A. S., & Levelt, W. J. M. (1996). Interaction between semantic and orthographic factors in conceptually driven naming: Comment on Starreveld and La Heij (1995). *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *22*, 246-251.

- Schriefers, H. (1993). Syntactic processes in the production of noun phrases. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 19, 841-850.
- Starreveld, P. A. & La Heij, W. (1995). Semantic interference, orthographic facilitation, and their interaction in naming tasks. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 21, 686-698.
- Starreveld, P. A. & La Heij, W. (1996). Time-course analysis of semantic and orthographic context effects in picture naming. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 22, 896-918.
- Stemberger, J. P. (1985). An interactive activation model of language production. In A. W. Ellis (Ed.), *Progress in the psychology of language* (pp. 143-186). London: Earlbaum.
- Taft, M., & Meunier, F. (1998). Lexical representation of gender: A quasiregular domain. *Journal of Psycholinguistic Research*, 27, 23-45.
- Tucker, G. R., Lambert, W. E., & Rigault, A. A. (1977). *The French speaker's skill with grammatical gender: An example of rule-governed behavior*. The Hague: Mouton.
- Vigliocco, G., Antonini, T., & Garrett, M. F. (1997). Grammatical gender is on the tip of Italian tongues. *Psychological Science*, 8, 314-317.
- Vigliocco, G., Franck, J., Antón-Méndez, I., & Collina, S. (in revision). The interplay of syntax and form in language production: An investigation of gender agreement in Romance languages. *Language and Cognitive Processes*.
- Vigliocco, G. & Hartsuiker, R. J. (2002). The interplay of meaning, sound, and syntax in sentence production. *Psychological Bulletin*, 128, 442-472.
- Vinson, D. P., & Vigliocco, G. (1999). Can independence be observed in a dependent system? The case of tip-of-the-tongue states. *Brain and Language*, 68, 118-126.

Vitevitch, M.S., & Luce, P.A. (2005). Increases in phonotactic probability facilitate spoken nonword repetition. *Journal of Memory and Language*, 52, 193-204.

Author Note

The Dutch Science Foundation (NWO) awarded the two travel grants NWO R 30-575 and NWO R 30-607 to Katharina Spalek. The research was partially funded by grant Nb 1114-068250 of the Swiss National Fund for Scientific Research to Ulrich Frauenfelder and Julie Franck.

Footnotes

¹ Throughout this article, we will use the following indices: mas. = masculine, fem. = feminine.

² It has generally been assumed that phonological influences on gender retrieval are due to word endings and not word onsets. However, if the phonology of onsets were correlated with gender, interactive activation models would presumably predict such effects, particularly since the phonological information comes early. Thus, for example, if 90% of all nouns starting with the cluster /fu/, for example, belong to the feminine gender class, gender retrieval for *fourchette* might be facilitated due to the existence of such an onset-cohort. Actually, Köpcke and Zubin (1983) report noun-initial *kn* as a phonological cue for masculine gender in German. In order to ensure that we were looking at the phonological regularity of the endings rather than other regularities, we computed the predictivity for noun-cohorts based on the initial syllables for nouns with more than one syllable and the initial CV for monosyllabic nouns. These values are reported in the Appendix.

³ Radeau, Mousty, and Bertelson (1989) say that for a gender decision task, the theoretical uniqueness point might have less impact than the “gender uniqueness point” which they describe as the point within a word where all the nouns of the remaining cohort are of the same gender. In their experiments, they only used items where theoretical uniqueness point and gender uniqueness point fell together. We computed both the gender uniqueness point and the theoretical uniqueness point for our study and assessed their correlation. As the two measures correlated with $r = .75$, both of them should affect the data pattern in a similar way. The cell with the earliest phonological UP (incongruent masculine nouns, 0.64 phonemes before the end of the word) also has the earliest grammatical UP (1.4 phonemes before the end of the word). The cell

with the latest phonological UP (congruent feminine nouns, 0.21 phonemes before the end of the word) also has the latest grammatical UP (0.47 phonemes before the end of the word). The grammatical UP lies between 0.17 phonemes (incongruent feminine) and 0.76 phonemes (incongruent masculine) earlier than the phonological UP.

⁴ We are aware of the fact that the interaction of an experimental factor with a factor “task” sometimes does not allow for a straightforward interpretation, in particular when the two tasks differ in relative difficulty. Therefore, we also analysed the logarithmically transformed reaction times. These analyses yielded the same pattern of results. Most importantly, the crucial interaction of congruency and task was significant for logarithmically transformed reaction times, too.

⁵ In order to ensure the comparability between experiments, we conducted an additional analyses on Experiment 1, in which the items we excluded from Experiment 2, were excluded from the data. The task by congruency interaction was still obtained.

Appendix

Experimental items and their predictivity measures.

Dominant Gender = masculine

Ending	Congruent Item	Type	Token	Incongruent Item	Type	Token
/to/	bateau [m]	.80	.79	moto [f]	.07	.15
/mo/	plumeau [m]	.78	.97	dynamo [f]	.09	.01
/lo/ /jo/ ^a	vélo [m]	.86	.97	radio [f]	.08	.23
/if/	canif [m]	.93	.97	griffe [f]	.06	.02
/UR/	carrefour [m]	.84	.87	tour [f]	.08	.04
/kl/	cercle [m]	.81	.97	boucle [f]	.19	.03
/aR/	cigare [m]	.87	.86	guitare [f]	.10	.09
/waR/	miroir [m]	.82	.77	bouilloire [f]	.16	.20
	mouchoir [m]	.82	.77	balançoire [f]	.16	.20
/tR/	cintre [m]	.77	.90	montre [f]	.14	.09
/εtR/	thermomètre [m]	.95	.87	fenêtre [f]	.04	.13
/œj/	fauteuil [m]	.75	.79	feuille [f]	.17	.21
/εZ/	manège [m]	.74	.75	neige [f]	.16	.22
/aZ/	nuage [m]	.98	.85	cage [f]	.01	.15
	village [m]	.98	.85	plage [f]	.01	.15
mean and sd		.85 (.08)	.86 (.08)		.10 (.06)	.13 (.08)
onset predictivity		.56 (.11)	.62 (.24)		.41 (.09)	.41 (.21)

Dominant Gender = feminine

Ending	Incongruent Item	Type	Token	Congruent Item	Type	Token
/in/	magazine [m]	.03	.16	cabine [f]	.95	.80
/ɛs/	cacatoès [m]	.17	.01	caisse [f]	.81	.89
/ot/	spot [m]	.17	.11	carotte [f]	.74	.87
/yR/	mur [m]	.12	.14	chaussure [f]	.87	.86
/as/	rapace [m]	.22	.02	contrebasse [f]	.70	.82
/jɛR/	lierre [m]	.06	.18	crinière [f]	.91	.79
	cimetière [m]	.06	.18	rivière [f]	.91	.79
/ɛn/	renne [m]	.24	.20	fontaine [f]	.73	.76
/ɛt/	squelette [m]	.08	.06	fourchette [f]	.89	.92
/ez/	trapèze [m]	.12	.11	fraise [f]	.86	.88
/uj/	fenouil [m]	.11	.02	grenouille [f]	.81	.97
/zõ/	poison [m]	.27	.09	maison [f]	.73	.91
	blouson [m]	.27	.09	prison [f]	.73	.91
/~ɛt/	labyrinthe [m]	.13	.02	plinthe [f]	.81	.80
/Rt/	cloporte [m]	.12	.00	tarte [f]	.88	1
mean and sd		.14 (.08)	.09 (.07)		.82 (.08)	.86 (.07)
onset predictivity		.63 (.13)	.68 (.27)		.45 (.13)	.46 (.30)

Note. Type = type predictivity. Token = token predictivity. The phonological transcriptions used are taken from the database *lexique*. ^aDue to the scarcity of depictable items that fulfilled all constraints, we paired a congruent (masculine) noun with the dominantly masculine ending /lo/ and an incongruent (feminine) noun with the dominantly masculine ending /jo/.

Non-words used in Experiment 1.

acasoir; adenesse; aduse; aémosse; aibernas; aint; aleil; ampie; anchette; anvale; arasol; arissol;
 artelle; asine; ason; âte; ateuil; atrennoir; augamette; aunêtre; aupeze; autoile; bacalnie; bapée;
 baquille; batenne; batruche; bétare; bloueil; borotte; bouca; bougateau; boutre; caceau; cacille;
 cairre; caisçoire; carlier; cataine; cathe; catoll; ceille; célançe; cenluta; céraçle; çhauçlam; çhéson;
 çhirte; çidiçon; çilon; çirorinthe; çlesse; çlodresse; çcoil; çconantie; çcormerelle; çcornouille; çcouçgle;
 çcracle; çcrio; çdyneau; çéçbouçge; çédrier; çenbyçgine; çengloumo; çenguette; çenne; çéçporte; çesnière; çéçsure;
 éçvière; faubine; faugare; feige; felloire; femozine; flomme; fonguitine; fourbeille; fral; graffe;
 grémobasse; guichoir; guilot; guirreau; hérage; hochelle; honnel; ilette; iseau; iteau; kilan; kilice;
 lasson; leige; lollage; lutre; maceau; madale; mainéoca; maraffe; mazaline; menifier; micriole;
 mopuche; mounane; mouzage; nine; nuson; oicluse; ommelier; onesse; onne; opelophe; orâte;
 ortalle; ouge; pinfour; piveçelle; plible; plour; plubeille; poiguille; poiroir; pounano; prinouil;
 ralogle; ramorette; reur; ringle; ripingle; sanrébic; sauteau; sæcle; spaï; squemimêtre; sule;
 susulle; thertique; tiès; tinse; tolle; toursiette; trateille; triteil; tupace; valo; véçbulance; verle;
 viseau; vitolon

Table 1.

Groupsize, logarithmic frequency, (phonological) uniqueness point, number of syllables, number of letters, and length in milliseconds for the materials in the four experimental conditions (sd in parentheses).

Ending		Masculine	Feminine
Congruent	Groupsize	246 (329)	226 (183)
	Log Freq	1.24 (0.57)	1.11 (0.73)
	Uniq. Point	4.93 (0.88)	5 (1.2)
	Nr. Syll.	1.93 (0.46)	1.8 (0.56)
	Nr. Phon.	5.4 (1.45)	5.2 (1.27)
	Length ms.	730 (120)	732 (125)
Incongruent	Groupsize	226 (183)	245 (329)
	Log Freq	0.69 (0.72)	1.38 (0.67)
	Uniq. Point	4.87 (1.19)	4.33 (1.05)
	Nr. Syll.	2.07 (0.88)	1.67 (0.72)
	Nr. Phon.	5.6 (1.72)	4.73 (1.53)
	Length	793 (162)	680 (149)

Table 2.

Experiment 1: Reaction times (in ms) measured from word onset and percentage errors as a function of task, gender, and congruency. Standard deviations in parentheses.

	Task			
	Lexical Decision		Gender Decision	
	Gender			
Congruency of the Ending	Masculine	Feminine	Masculine	Feminine
Congruent	848 (86)	838 (100)	969 (71)	944 (123)
	1.6% (3.5)	1.4% (2.8)	1.1% (3.2)	4.4% (4.7)
Incongruent	891 (108)	836 (96)	1077 (120)	996 (136)
	7.5% (7.9)	2.2% (3.8)	11.1% (7.2)	8.3% (5.6)
Incongruent minus Congruent	43	-2	108	52
	5.9%	-0.8%	10%	3.9%

Table 3.

Experiment 1: Reaction times (in ms) measured from word offset as a function of task, gender, and congruency. Standard deviations in parentheses.

	Task			
	Lexical Decision		Gender Decision	
	Gender			
Congruency of the Ending	Masculine	Feminine	Masculine	Feminine
Congruent	117 (94)	109 (99)	241 (74)	212 (121)
Incongruent	103 (106)	164 (104)	267 (112)	320 (137)
Incongruent minus Congruent	-14	55	26	108

Table 4.

Experiment 2: Reaction times (in ms) and percentage errors as a function of utterance type, gender, and congruency. Standard deviations in parentheses.

	Utterance Type			
	Bare Noun		Determiner Adjective Noun	
Congruency of the Ending	Gender			
	Masculine	Feminine	Masculine	Feminine
Congruent	781 (107)	809 (110)	817 (162)	805 (168)
	6% (7)	9% (9)	8% (9)	14% (11)
Incongruent	835 (113)	821 (110)	892 (159)	842 (178)
	12% (8)	6% (8)	17% (11)	9% (10)
Incongruent minus Congruent	54	12	75	37
	6%	-3%	9%	-5%

Figure Caption

Figure 1. Panel a) Reaction time difference for reaction times measured from word onset between the gender decision task and the lexical decision task as a function of gender and congruency.

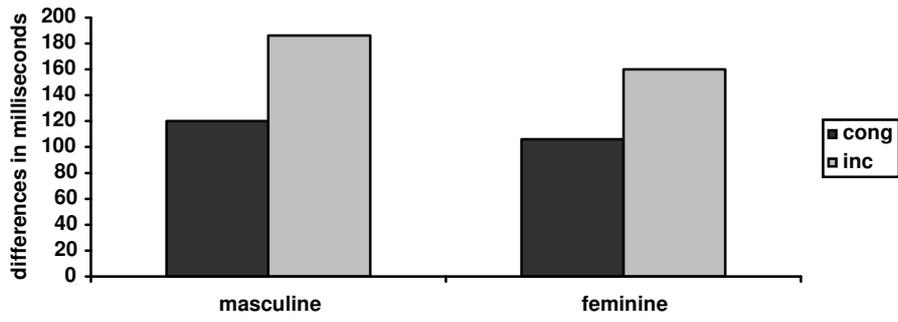
Panel b) Reaction time difference for reaction times measured from word onset between the gender decision task and the lexical decision task as a function of gender and congruency.

Figure 2. Reaction time difference between adjective determiner NP naming and bare noun naming as a function of gender and congruency.

Figures

Figure 1.

Panel a)



Panel b)

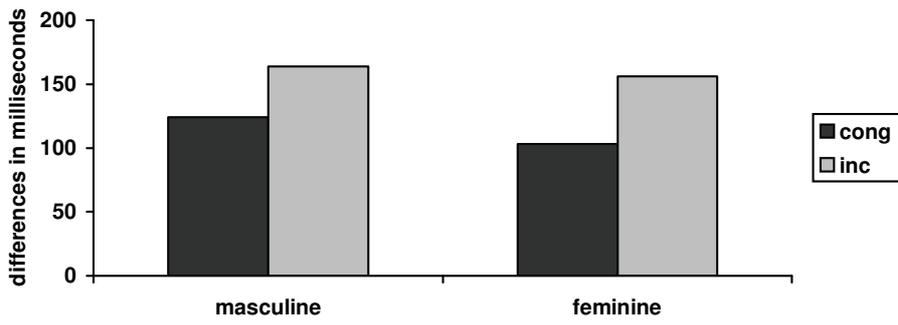


Figure 2.

