The Processing of Free and Bound Gender-Marked Morphemes in Speech Production: Evidence From Dutch

Kristin Lemhöfer
University of Leipzig and Radboud University Nijmegen

Herbert Schriefers
Radboud University Nijmegen

Jörg D. Jescheniak
University of Leipzig

In many languages, the production of noun phrases requires the selection of gender-marked elements like determiners or inflectional suffixes. There is a recent debate as to whether the selection of freestanding gender-marked elements, such as determiners, follows the same processing mechanisms as the selection of bound gender-marked morphemes, such as adjectival suffixes. Most of the evidence on which this debate is based relates to the gender-congruency effect in picture-word interference experiments. In the present article, the authors address this issue with a pure picture-naming task, extending previous work in German (H. Schriefers, J. D. Jescheniak, & A. Hantsch, 2005). The results of the present study on noun phrase production in Dutch show that both types of gender-marked morphemes are selected via the same basic processing mechanisms.

Keywords: speech production, grammatical gender, free morphemes, bound morphemes

In recent years, experimental research on speech production has started to focus on higher order processes, such as the encoding of the syntactic features of a word. For example, during the production of a noun phrase in a language like Dutch, the speaker has to retrieve the noun’s grammatical gender in order to select the appropriate gender-marked elements, like “free” morphemes (e.g., determiners) or “bound” morphemes (e.g., inflectional suffixes).

Previous studies have shown that the selection of free gender-marked morphemes is subject to competitive processes. In a study by Schriefers, Jescheniak, and Hantsch (2002), German speakers named singular and plural objects by determiner–noun phrases (e.g., der Baum, ‘themasc,sg tree,’ or die Bäume, ‘thepl trees’).1 In German, the definite plural determiner for all genders is die, which is identical to the definite singular determiner for feminine nouns (e.g., die Tür, ‘thefem,sg door’). The results of this study showed that producing plural noun phrases was associated with a reaction time (RT) cost for masculine and neuter nouns, that is, in those cases in which singular and plural determiners did not match. However, for feminine nouns, for which singular and plural determiners are identical, this plural cost disappeared (see also Jansen & Caramazza, 2003, and Spalek & Schriefers, 2005, for parallel results in Dutch, and see Miozzo & Caramazza, 1999, for related results in Italian). These results suggest that during the production of plural noun phrases, the respective singular determiners become coactivated and compete with the plural determiners for selection, whereas the reverse is not the case (the singular-as-default hypothesis).

These results were obtained in a condition in which the relative proportion of the three determiner forms der, die, and das was balanced by the use of filler items, implying that there were more singular than plural trials. Schriefers et al. (2002) also showed that when the proportion of singular and plural trials was equal (leading to a higher proportion of the determinant die), the RT costs associated with plural noun phrases for masculine and neuter nouns were reduced or eliminated and a plural gain obtained for feminine nouns.

Recently, however, the question has emerged whether the observed competition of determiners, that is, of freestanding gender-marked elements, arises also during the selection of bound gender-marked morphemes, such as gender-marked adjectival inflections. Using the picture-word interference task, which has repeatedly been proven to be sensitive to determiner competition (La Heij, Mak, Sander, & Willeboordse, 1998; Schriefers, 1993; Schriefers & Teruel, 2000), two studies failed to find evidence for competi-

---

1 We use subscripts to indicate those elements that are marked for gender and/or number in the target language: masc = masculine, fem = feminine, neu = neuter, com = common gender (in Dutch), sg = singular, pl = plural.

Schiller and Caramazza (2003) investigated noun phrase production in German and Dutch. In both of these languages, a definite determiner carries all gender marking, with an additional adjective remaining uninfluenced (for the two gender classes, common and neuter, in Dutch: de rode tafel, ‘the_rode_red table,’ het rode boek, ‘the_rode_red book’; similar for German), whereas in absence of a determiner, the adjective takes over the function of gender marking by way of suffixes (rote tafel, ‘red (gender) table’; roed boek, ‘red (gender) book’). The results showed that in both languages, the production of phrases involving a singular definite determiner was slowed down by the presentation of a gender-incongruent distractor, compared with a condition in which the distractor had the same gender as the target noun. However, in adjective–noun phrases, where gender marking was carried by a bound morpheme in the adjective, this effect disappeared. This result stands in contrast to the findings by Schriefers (1993), who did observe a gender-congruency effect for Dutch adjective–noun phrases, which was, however, smaller than for phrases containing a gender-marked determiner.

Investigating the same issue of a potential difference between freestanding and bound gender-marked morphemes, Costa et al. (2003) had participants produce phrases in Croatian involving gender-marked personal pronouns or noun phrases involving a gender-marked possessive pronoun. The results showed a gender (in)congruency effect in the first case, in which gender marking was carried by freestanding morphemes, whereas no such effect was found for noun phrases containing inflected possessive pronouns.

On the basis of these results, Costa et al. (2003) and Schiller and Caramazza (2003) proposed that the selection of free versus bound gender-marked morphemes involves qualitatively different processes. More specifically, free gender-marked morphemes, such as determiners, were assumed to compete with each other for selection, whereas the selection of bound morphemes either might not be subject to competition at all (because they involve “phonological transformations” rather than the selection of independent morphemes; Costa et al., 2003) or competition might take place so late during noun phrase production that it can be resolved without affecting the final response latencies (Schiller & Caramazza, 2003).

Recent findings of Schriefers, Jescheniak, and Hantsch (2005), however, question this view. In three experiments, speakers of German were asked to produce singular and plural noun phrases of different formats in the context of a picture-naming task. In the noun phrases to be produced, gender marking was either carried by the determiner (determiner–adjective–noun phrases) or by suffixes of the adjective (adjective–noun phrases). In German, the pattern of gender-inflation of adjectives corresponds to the pattern that holds for determiners: The inflection indicating plural number (-e, as in rote Türe, ‘red (gender) door’) is identical across the three gender classes; furthermore, it is identical to the singular inflection for feminine nouns (rote Tür, ‘red (gender) door’). The results of this study showed similar behavioral patterns for both types of noun phrases: The RT difference between plural and singular trials was reduced for feminine nouns, that is, for those cases in which singular and plural gender-marking morphemes coincided. However, the effect was somewhat weaker for adjective–noun phrases, compared with noun phrases involving a determiner. This latter observation of an attenuated effect in the case of utterance non-initial bound morphemes is in line with Schriefers’s (1993) earlier finding.

Thus, to summarize, the empirical evidence concerning the selection of free versus bound gender-marked morphemes is still ambiguous. Whereas some (but not all) picture–word interference studies produced results that indicate competitive processes for the selection of freestanding morphemes such as determiners, but not for bound morphemes (e.g., suffixes), the findings by Schriefers et al. (2005), using a pure picture-naming task, do not support such a fundamental distinction.

In view of this situation, in the present study we sought to replicate the results obtained by Schriefers et al. (2005) in Dutch. Although the gender-marking system of Dutch is, on the whole, very similar to German, it is also different in some aspects. In particular, in contrast to German, in which almost all adjective inflections are realized by way of suffixes (-e, -en, -er), exceptions are a few color adjectives like lila (‘purple’) or rosa (‘pink’), which do not take inflectional suffixes, and the indefinite determiner ein (‘a’), which only takes an inflectional suffix in feminine gender), gender-marked adjectives in Dutch are either suffixed with a null morpheme (i.e., they remain in the citation form) or they are suffixed with -e (see Table 1). Thus, Dutch allows one to investigate whether the competition between bound

<table>
<thead>
<tr>
<th>Utterance format</th>
<th>Determiner + adjective + noun</th>
<th>Adjective + noun</th>
<th>Bare noun</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Singular</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Common</td>
<td>de grote hond</td>
<td>grote hond</td>
<td>hond</td>
</tr>
<tr>
<td></td>
<td>(‘the eg_common big dog’)</td>
<td>(‘big eg_common dog’)</td>
<td>(‘dog’)</td>
</tr>
<tr>
<td>Neuter</td>
<td>het grote paard</td>
<td>groot paard</td>
<td>paard</td>
</tr>
<tr>
<td></td>
<td>(‘the eg_neu big horse’)</td>
<td>(‘big eg_neu horse’)</td>
<td>(‘horse’)</td>
</tr>
<tr>
<td><strong>Plural</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Common</td>
<td>de grote honden</td>
<td>grote honden</td>
<td>honden</td>
</tr>
<tr>
<td></td>
<td>(‘the eg_pl big dogs’)</td>
<td>(‘big eg_pl dogs’)</td>
<td>(‘dogs’)</td>
</tr>
<tr>
<td>Neuter</td>
<td>de grote paarden</td>
<td>grote paarden</td>
<td>paarden</td>
</tr>
<tr>
<td></td>
<td>(‘the eg_pl big horses’)</td>
<td>(‘big eg_pl horses’)</td>
<td>(‘horses’)</td>
</tr>
</tbody>
</table>

*Note. sg = singular; com = common; neu = neuter; pl = plural.*
Vocabulary was created by listing all the nouns that enter the competition is a null morpheme or not. The experimental design was adapted from Schriefers et al. (2005). Participants had to name single or double objects in one of three utterance formats (see Table 1): (1) As determiner–adjective–noun phrases, in which gender marking is carried by the determiner; (2) as adjective–noun phrases, in which the adjective takes over the function of gender marking by way of suffixes (null or –e morpheme), or (3) as bare nouns, to control for possible differences between the items of the two gender categories. If the results of Schriefers et al. (2005) generalize to Dutch, we would expect that for determiner–adjective–noun phrases and adjective–noun phrases (Utterance Formats 1 and 2), the difference between plural and singular trials should be reduced for common gender words (for which the gender-marking element remains invariant in singular and plural). This should result in an interaction of gender and number for these utterances, but not for bare noun utterances (Utterance Format 3). On the other hand, given the mixed picture of singular and plural word frequency according to the CELEX database, the nouns of the two gender categories were matched item by item for the number of syllables of both the singular and plural word form, singular and plural word frequency according to the CELEX database (Baayen, Piepenbrock, & Gulikers, 1995), and plural dominance (i.e., the relative frequency of the plural compared with the singular). All experimental items are listed in the Appendix.

Eighteen additional object drawings (9 of each gender) were included as training and warming-up items. The line drawings were prepared in three different sizes (see Schriefers et al., 2005, for details): in medium size for the familiarization and subsequent training phase and in small and large size for the main experiment. The small and large pictures were used to induce noun phrases containing either the adjective klein (‘small’) or groot (‘big’). The use of two different adjectives avoids a situation in which the adjective + noun utterances always start with the same adjective stem, and it thus prevents strategic behavior on the part of the participants that might eliminate potential effects of competition of inflectional morphemes. In the main experiment, objects were always presented twofold (two identical objects next to each other) with the color black indicating the target(s). Participants were instructed to produce plural utterances if both objects were drawn in black, and singular utterances if one of them was drawn in black while the other was drawn in gray. This way, display complexity was held constant across singular and plural trials (for more details on stimulus construction, see Schriefers et al., 2005).

Procedure

The experimental session consisted of three parts: the familiarization phase, the training phase and the main experiment. In the familiarization phase, the participants saw each of the 66 (experimental and warming-up) line drawings as a single medium-sized object with its name (as a singular bare noun) written below it in the middle of a computer monitor, with the instruction to read the names aloud and to use these names in the remainder of the experiment. A new trial started after the participant had pressed a button on a button box. In the subsequent training, participants named the pictures (as singular bare nouns), which were now presented without their names. In the main experiment, participants named the pictures in one of the three utterance formats (see Table 1). Two practice blocks with the 18 warming-up items were followed by the two main blocks, each consisting of the 48 experimental pictures, preceded by 9 warming-up items. Singular and plural trials were mixed, with each object occurring once in the singular and once in the plural condition. Whether a given object occurred in a singular or a plural trial first was counterbalanced across participants. Furthermore, object size and, in singular trials, the position of the black line drawing (left vs. right) were counterbalanced. Altogether, 16 experimental lists were constructed. In these lists, no more than five trials with objects of the same size, determiner, gender category, or number followed each other. Furthermore, objects with names that shared the same onset or that were members of the same semantic category were not presented in immediate succession. A minimum of eight trials occurred in between the two presentations of the same object in singular and plural trials. Precisely the same 16 lists were used in each utterance format. The experimental session took about 35 min in total.

The participant was seated in a soundproof cabin in front of the microphone, the button box (used during the familiarization phase only), and the computer with a Pentium processor. Speech-onset latencies were measured to the closest millisecond with a voice key connected to the computer. In the training phase and main experiment, pictures were presented for 1,000 ms. The time between the offset of a picture and the onset of the next picture was 2,500 ms.

Results

For the analysis of RTs, trials with disfluent or incorrect answers and those with incorrect speech onset measures (e.g., due to a nonspeech sound prematurely triggering the voice key) were excluded. The mean percentage of errors was 7.4%. Outlier RTs that
lay more than 2 standard deviations away from the subject and item mean (per experimental condition) were also discarded (1.9% of the correct trials).2

The mean RTs and error rates for the three utterance formats are reported in Table 2. The data were analyzed with the factors format (between participants, within items), gender (within participants, between items), and number (within participants, within items).

The RT differences between plural and singular noun phrases as a function of utterance format and gender class are illustrated in Figure 1. In the analysis of variance (ANOVA) on the RTs, there was a significant main effect of format, $F(1, 45) = 3.6, p < .05$, $MSE = 25,608$; $F(2, 92) = 86.2, p < .001, MSE = 1,542$, with bare nouns being the fastest ($M = 690$ ms) and adjective–noun noun phrases being the slowest ($M = 763$ ms). The mean RT for determiner–adjective–noun noun phrases was $743$ ms. Furthermore, gender also significantly affected RTs, $F(1, 45) = 96.6, p < .001, MSE = 640$; $F(2, 46) = 8.8, p < .01, MSE = 11,167$, with longer RTs for neuter ($M = 750$ ms) compared with common gender gender words (there was a significant difference of $61$ ms with a 95% confidence interval of $20$ ms for determiner–adjective–noun noun phrases, and there was a significant difference of $26$ ms with a 95% confidence interval of $17$ ms for adjective–noun noun phrases), but not for neuter gender. In contrast, no significant differences between singular and plural utterances arose in the baseline condition with bare nouns.

The error rates were analyzed in analogy to the RTs. In the $3 \times 2 \times 2$ ANOVA, there was a main effect of format, $F(2, 45) = 29.3, p < .001, MSE = 0.004$; $F(2, 92) = 34.7, p < .001, MSE = 0.006$, with most errors for determiner–adjective–noun noun phrases (mean 11.9%) and the least errors for bare nouns (3.4%). The mean for adjective–noun noun phrases was 7.0%. The main effect of gender was also significant, $F(1, 45) = 9.3, p < .01, MSE = 0.003$; $F(1, 46) = 4.8, p < .05, MSE = 0.01$, with fewer errors on common gender words (6.2%) than on words with neuter gender (8.6%). The main effect of number was not significant (both $Fs < 1$); neither were any of the two-way interactions (all $ps > .13$). The three-way interaction of format, gender, and number was not significant either, $F(2, 45) = 1.51, p > .20$; $F(2, 92) = 1.25, p > .25$. The fact that the overall three-way interaction was not significant indicates that for error rates, baseline and

<table>
<thead>
<tr>
<th>Number</th>
<th>Determiner + adjective + noun</th>
<th>Adjective + noun</th>
<th>Bare noun</th>
</tr>
</thead>
<tbody>
<tr>
<td>Singular</td>
<td>RT (SE)</td>
<td>756 (25)</td>
<td>770 (25)</td>
</tr>
<tr>
<td></td>
<td>ER (SE)</td>
<td>11.5 (1.9)</td>
<td>14.1 (1.5)</td>
</tr>
<tr>
<td>Plural</td>
<td>RT (SE)</td>
<td>695 (20)</td>
<td>751 (22)</td>
</tr>
<tr>
<td></td>
<td>ER (SE)</td>
<td>8.3 (1.4)</td>
<td>13.5 (1.6)</td>
</tr>
<tr>
<td>Difference&lt;sup&gt;a&lt;/sup&gt;</td>
<td>RT (SE)</td>
<td>-61 (10)**/*</td>
<td>-19 (15)</td>
</tr>
<tr>
<td></td>
<td>ER (SE)</td>
<td>-3.2 (2.8)</td>
<td>-0.6 (1.7)</td>
</tr>
</tbody>
</table>

Note. Standard errors in parentheses. Probability is given for t test results, over participants/over items.

<sup>a</sup> Plural = singular.

* $p < .05$. ** $p < .01$. *** $p < .001$.

The critical interaction of gender and number (indicating that the difference between singular and plural trials was modulated by word gender) was significant, $F(1, 30) = 12.5, p < .01, MSE = 751$; $F(1, 46) = 11.6, p < .01, MSE = 1138$. We conducted t tests (see Table 2) that confirmed this pattern of results for these two utterance formats; the plural gain was significant for common gender words (there was a significant difference of 61 ms with a 95% confidence interval of 20 ms for determiner–adjective–noun noun phrases, and there was a significant difference of 26 ms with a 95% confidence interval of 17 ms for adjective–noun noun phrases), but not for neuter gender. In contrast, no significant differences between singular and plural utterances arose in the baseline condition with bare nouns.

The triple interaction Format $\times$ Gender $\times$ Number was no longer significant, $F(1, 30) < 1$; $F(2, 46) = 2.7, p > .10$. At the same time, the critical interaction of gender and number (indicating that the difference between singular and plural trials was modulated by word gender) was significant, $F(1, 30) = 12.5, p < .01, MSE = 751$; $F(1, 46) = 11.6, p < .01, MSE = 1138$. We conducted t tests (see Table 2) that confirmed this pattern of results for these two utterance formats; the plural gain was significant for common gender words (there was a significant difference of 61 ms with a 95% confidence interval of 20 ms for determiner–adjective–noun noun phrases, and there was a significant difference of 26 ms with a 95% confidence interval of 17 ms for adjective–noun noun phrases), but not for neuter gender. In contrast, no significant differences between singular and plural utterances arose in the baseline condition with bare nouns.

<sup>2</sup> In addition to the analyses reported here, where each individual incorrect trial and outlier was excluded, we also carried out all RT analyses with a “pairwise” exclusion of these trials (i.e., both the singular and the corresponding plural presentation of an excluded item were discarded for a given participant). The different procedures had only minor effects on the mean RTs and no effect at all on the results of the analyses of variance.
gender-marked conditions did not differ in terms of the critical 
Gender × Number interaction.

Discussion

The obtained results confirm and extend the previous findings 
on the selection of gender-marked elements reported by Schriefers 
et al. (2002, 2005). For gender-marked utterances, RTs in plural 
trials were significantly shorter than those in singular trials when 
the gender-marked elements of plural and singular were identical 
(i.e., for common gender), whereas this RT difference disappeared 
for nouns with nonmatching gender-marking elements in singular 
and plural (neuter gender). Crucially, this pattern was found for 
both free (determiners) and bound (suffixes) gender-marked mor-
phemes, whereas it was absent in the nongender-marked condition 
(bare noun naming). Descriptively, the effects were attenuated for 
bound morphemes as compared with free morphemes, but not 
significantly so. Thus, the data not only replicate the results 
concerning freestanding gender-marked morphemes (e.g., Schrief-
er et al., 2002) but also show that similar processes take place 
when bound gender-marking morphemes, such as adjective inflec-
tions, are concerned (see also Schriefers et al., 2005). The results 
therefore extend the findings by Schriefers et al. (2005) to a 
different language, Dutch, which differs from German with respect 
to its use of suffixes in gender-marked adjectives (see the intro-
duction). Together, the two studies, the present one and the one by 
Schriefers et al. (2005), provide converging evidence against the 
view that freestanding and bound gender-marking morphemes are 
processed in fundamentally different ways.

Still, there is some inconsistency among the extant data with 
respect to the effects obtained for bound morphemes, like adjective 
inflections. It is currently unclear what exactly is responsible for 
this situation. A first issue to consider is that both Schiller and 
Caramazza (2003) and Costa et al. (2003) made use of the picture-
word interference paradigm, which differs in many aspects from 
the simple naming task used here (for a more elaborate discussion 
of this point, see Schriefers et al., 2005). A second issue to 
consider is that in both the present study and Schriefers et al. 
(2005) the critical effect (the modulation of the plural–singular 
difference by gender) for adjective–noun phrases (with bound mor-
phemes) was descriptively smaller than that for determiner–adjective–noun phrases (with free morphemes). Thus, it is possible 
that the picture–word paradigm is less sensitive to the effects at 
question, so that the smaller competition effect for bound mor-
phemes cannot be picked up in this paradigm (but see Schriefers, 
1993). This suggestion could be tested by using precisely the same 
materials from a simple picture-naming experiment showing a 
Gender × Number interaction for utterances with bound gender-
marked morphemes in a picture–word interference experiment 
with gender-congruent and gender-incongruent distractors. It ap-
pears to be highly advisable to have this direct comparison empir-
ically tested before embarking on further discussions on the po-
tential communalities and differences in the processing of free and 
bound gender-marked morphemes.

Another issue evoked by a comparison of the present with 
previous results is that of the direction of the plural–singular 
difference. For the gender-marked utterance formats, that is, 
determiner–adjective–noun noun phrases and adjective–noun noun 
phrases, in the present study, there was no significant difference 
between singular and plural for nouns with different gender-
marked elements in plural and singular (neuter gender), and a 
significant plural gain for common gender nouns, for which the 
determiner is invariant across singular and plural. By contrast, both 
Schriefers et al. (2005) and Schriefers et al. (2002) had obtained 
plural costs for German neuter and masculine nouns in phrases 
involving a gender-marked determiner. However, as shown by 
Schriefers et al. (2002), whether the production of plural noun 
phrases creates an RT cost or a gain seems to depend on the 
proportion of singular and plural trials (and thus on the proportion
of the different gender-marked elements) in the experiment (see also Spalek & Schriefers, 2005, for further discussion). More specifically, there appears to be a growing tendency toward plural gains as the relative frequency of the word form of the plural determiner in the experiment increases (i.e., in the case of equal numbers of singular and plural trials in the experiment). This is also in line with the results obtained by Schiller and Caramazza (2003) with the picture–word interference paradigm. In their experiments, the proportions of singular and plural noun phrases were also equal, and the authors also observed plural gains in all Dutch (and some of the German) experiments.

To summarize, the results of the present experiment show that the selection of freestanding gender-marking morphemes (determiners) and that of bound morphemes such as adjective suffixes follow the same processing principles, a finding that is in line with previous evidence from German (Schriefers et al., 2005). The current data thus provide evidence against the view that the selection of free morphemes and the selection of bound gender-marked morphemes are qualitatively distinct processes. Furthermore, it appears that the picture–word interference paradigm is less sensitive for picking up the selection process of bound gender-marked morphemes than the pure naming task used in the present experiments, though this latter assumption is still lacking a direct empirical test.

References


Appendix

Names of Experimental Objects Used in the Experiment (Approximate English Translations in Parentheses) with Mean Item Statistics

Common gender

aap (monkey), bel (bell), bijl (axe), borstel (brush), bril (glasses), bus (bus), deur (door), douche (shower), hond (dog), jurk (dress), kast (cupboard), kroon (crown), lamp (lamp), lepel (spoon), riem (belt), schaar (scissors), sigaar (cigar), spiegel (mirror), spons (sponge), stoel (chair), tafel (table), tent (tent), vogel (bird), vos (fox).

Item statistics: mean word form frequency, singular form: 47.1 occurrences per million words (o.p.m.); plural form: 11.6 o.p.m.; mean number of syllables, singular form: 1.21; plural form: 2.04; values taken from the CELEX database (Baayen et al., 1995).

Neuter gender

anker (anchor), bed (bed), boek (book), dak (roof), eiland (island), gewicht (weight), hek (fence), hert (deer), kasteel (castle), konijn (rabbit), mes (knife), nest (nest), net (net), oor (ear), paard (horse), pistool (pistol), raam (window), skelet (skeleton), slot (padlock), tapijt (carpet), touw (rope), vuur (fire), zadel (saddle), zwaard (sword).

Item statistics: word form frequency, singular form: 4.8 o.p.m., plural form: 17.2 o.p.m.; mean number of syllables, singular form: 1.38, plural form: 2.29; values taken from the CELEX database (Baayen et al., 1995).