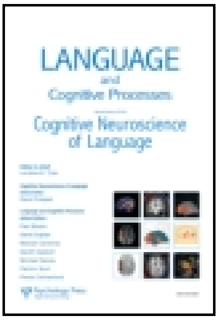

Language and Cognitive Processes
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/plcp20

The selection of closed-class elements during language production: A reassessment of the evidence and a new look on new data

Niels Janssen [a] , Niels O. Schiller [b] & F.-Xavier Alario [c]

[a] Departamento de Psicología Cognitiva, Social y Organizacional, Facultad de Psicología, University of La Laguna, La Laguna, Spain

[b] Leiden Institute for Brain and Cognition, Leiden University, Leiden, The Netherlands

[c] Aix-Marseille Université & CNRS, Marseille, France

Version of record first published: 11 Sep 2012.

To cite this article: Niels Janssen, Niels O. Schiller & F.-Xavier Alario (2012): The selection of closed-class elements during language production: A reassessment of the evidence and a new look on new data, Language and Cognitive Processes, DOI:10.1080/01690965.2012.693617

To link to this article: http://dx.doi.org/10.1080/01690965.2012.693617

LANGUAGE AND COGNITIVE PROCESSES, 2012, iFirst, 1–19

Psychology Press
Taylor & Francis Group

The selection of closed-class elements during language production: A reassessment of the evidence and a new look on new data

Niels Janssen[1], Niels O. Schiller[2], and F.-Xavier Alario[3]

[1]Departamento de Psicología Cognitiva, Social y Organizacional, Facultad de Psicología, University of La Laguna, La Laguna, Spain
[2]Leiden Institute for Brain and Cognition, Leiden University, Leiden, The Netherlands
[3]Aix-Marseille Université & CNRS, Marseille, France

Are closed-class items such as free-standing determiners and bound inflectional morphemes selected by a competitive or a noncompetitive mechanism? Jescheniak, Schriefers, and Lemhöfer provide a survey of the literature on this topic involving studies employing the picture–word interference and simple picture naming tasks. They claim that the extant evidence supports a competitive lexical selection mechanism for both types of closed-class items. In this commentary, we critically evaluate their proposal and present new empirical data. We come to the conclusion that the available data do not allow a straightforward interpretation in terms of a particular hypothesis of lexical selection. We highlight the need to consider answers to the theoretical question within a broader research context.

Keywords: Language production; Closed-class words; Lexical selection; Competition.

An ongoing debate in the language production literature concerns the mechanism of lexical selection. Currently, two basic hypotheses are contrasted. The *competitive lexical selection hypothesis* assumes that target selection time is a function of the activation level of lexical competitors. By contrast, the *noncompetitive lexical selection hypothesis* assumes that target selection time is a function of the target's own activation level irrespective of lexical competitors' activation. The majority of studies concerning this question have focused on the production of open-class words such as nouns (e.g. Abdel Rahman & Melinger, 2009; Janssen & Caramazza, 2011). The current lack of consensus regarding this issue is evidenced by the notable number of

---

Correspondence should be addressed to Niels Janssen, Grupo Neurociencia Cognitiva & Psicolingüística, Departamento de Psicología Cognitiva, Social y Organizacional, Facultad de Psicología, Universidad de La Laguna, Campus de Guajara s/n, La Laguna 38205, Tenerife, Spain. E-mail: njanssen@ull.es

This research was supported by grant RYC-2011-08433 from the Spanish Ministry of Economy and Competitivity to NJ. NOS was supported as a fellow-in-residence 2010/11 at the Netherlands Institute for Advanced Study (NIAS) in the Humanities and Social Sciences, Wassenaar, The Netherlands. The research leading to these results has received funding from the European Research Council under the European Community's Seventh Framework Program (FP7/2007–2013 Grant agreement n° 263575).

http://www.psypress.com/lcp http://dx.doi.org/10.1080/01690965.2012.693617

articles published in the last years that adjust, rebut, or otherwise comment on previously published theoretical proposals (cf. most recently Mahon, Garcea, & Navarrete, 2012; Mulatti & Coltheart, 2012). Jescheniak, Schriefers, and Lemhöfer (in press), valuably broaden the scope of the debate regarding competitive versus noncompetitive models by reviewing evidence from studies on the production of closed-class items, rather than nouns. Specifically, they focus on the production of free-standing (e.g. determiners) and bound (e.g. inflections) gender-marked morphemes. From their review, they conclude that both types of closed-class items are selected by the same mechanism, and that this mechanism is competitive. In this commentary, we present a discussion of Jescheniak et al.'s proposal as well as new empirical evidence which undermine their conclusions.

Jescheniak et al. consider the production of free-standing and bound morphemes in two different tasks: the picture–word interference task (PWI; Schriefers, 1993), and the simple picture naming task (SPN; Alario & Caramazza, 2002; Schriefers, Jescheniak, & Hantsch, 2002).

In the PWI task, studies have focused on the effect of *gender congruency*, in which pictures are named in the context of distractor words that either share or do not share the grammatical gender of the picture name. In Germanic and Slavic languages, there are reliable gender congruency effects in the production of free-standing determiners (Bordag & Pechmann, 2008; Costa, Kovacic, Fedorenko, & Caramazza, 2003; Schiller & Caramazza, 2003; Schriefers, 1993). For example, Dutch speakers named pictures with determiner + noun NPs (e.g. de $auto_{COM}$ [the car]) slower in the context of gender incongruent (e.g. $boek_{NEU}$ [book]) than gender congruent (e.g. $kerk_{COM}$[church]) distractor words.[1] Both competitive and noncompetitive hypotheses can account for this effect. The competitive selection hypothesis explains slower latencies in the incongruent condition in terms of competition between divergent determiner forms. The noncompetitive hypothesis explains faster latencies in the congruent condition in terms of priming between convergent determiner forms. In other words, as discussed by Jescheniak et al. (but see Lemhöfer, Schriefers, & Jescheniak, 2006; Schriefers, 1993; Schriefers, Jescheniak, & Hantsch, 2002, 2005), the gender congruency effect is ambiguous between a competitive and noncompetitive selection hypothesis.

Jescheniak et al. also re-interpret past evidence that they argue supports a competition interpretation of the gender congruency effect in the PWI task. Specifically, they make comparisons of numerical means of experimental conditions in two separate experiments (Foucart, Branigan, & Bard, 2010, Experiment 1; Schiller & Caramazza, 2003, Experiment 4b). However, these comparisons are clearly post-hoc, and more focused experiments are necessary to address this issue in full. More importantly, perhaps, Jescheniak et al. fail to consider in any detail a study by Alario, Ayora, Costa, and Melinger (2008) reporting a direct test of the competition account in PWI. Using a Stroop-like rationale, these authors used as distractors what are arguably the strongest competitors for determiners, namely other determiners of different gender, number, or definiteness. The results of five experiments showed that in different languages and at different Stimulus Onset Asynchronies (SOAs) the production of determiner NPs was sped up rather than slowed down by the presentation of such distractors. The absence of a discrimination cost between

[1] Such an effect is not seen in Romance languages (Caramazza et al., 2001). For French, the evidence is mixed; compare Alario and Caramazza (2002), Foucart, Branigan, and Bard (2010), and Schriefers and Teruel (1999).

conditions in which the distractor determiners were increasingly similar to (but different from) the target determiner stands in contrast with the predictions of the competition hypothesis. In short, the gender congruency effect cannot be used to distinguish between competitive and noncompetitive selection hypotheses, and experiments using other effects in the PWI task seem more in line with a noncompetitive account of closed-class word selection.

The SPN task seems a more promising paradigm for adjudicating between competitive and noncompetitive accounts, yet its use requires interpreting complex interaction patterns. This is because, contrary to the PWI task, the relevant contrasts in the SPN task involve between-item comparisons (e.g. words with different grammatical genders, or pictures depicting one vs. multiple items). This imposes that idiosyncratic differences must be partialled out from the relevant contrasts before the latter can be interpreted. While this constraint has been taken into account in previous reports, we will argue below that the details of its implementation may have been suboptimal.

Consider as an explicit example the Dutch version of the SPN task (see Table 1 for specific examples sentences). It involves four sets of pictures with: (1) common gender names that are named in singular, (2) neuter gender names that are named in singular, (3) common gender names that are named in plural, and (4) neuter gender names that are named in plural. This 2×2 design illustrates a property that has been used throughout these studies. Common gender nouns are used with the same determiner in singular and in plural, whereas neuter gender nouns are used with different determiners in singular and plural (see Table 1).

In this context, the competitive and noncompetitive selection hypotheses predict different types of gender by number interactions in the production of determiner NPs. A competitive hypothesis assumes that, in plural neuter NPs, increased competition will arise during determiner selection because of the activation of alternative determiner forms (i.e. "het" and "de"; Figure 1, left panel). Such increased competition will not be present in plural common NPs, nor in singular NPs irrespective of gender (Figure 1, right panel). Accordingly, competitive selection predicts a gender by number interaction where there is *an extra cost in the production of plural NPs with neuter gender nouns* (compared to plural common NPs).

By contrast, according to the noncompetitive selection hypothesis, plural common determiners will benefit from activation from both gender and number features (Figure 1, left panel). Such will not be the case for plural neuter determiners, nor in singular NPs irrespective of gender. Accordingly, noncompetitive selection predicts a

TABLE 1
Examples of Dutch definite gender-marked determiners (examples 1 and 2; as used in Experiment 1a) and distal demonstrative determiners (examples 3 and 4; as used in Experiment 1b) as a function of gender (common vs. neuter) and number (singular NPs vs. plural NPs). For both determiner types, only the determiner form of neuter gender nouns differs between singular and plural NPs

Example	*Number*	*Common*	*Neuter*
1	sg	*de* grote auto [the big car]	*het* kleine boek [the small book]
2	pl	*de* grote autos [the big cars]	*de* kleine boeken [the small books]
3	sg	*die* auto [that car]	*dat* boek [that book]
4	pl	*die* autos [those cars]	*die* boeken [those books]

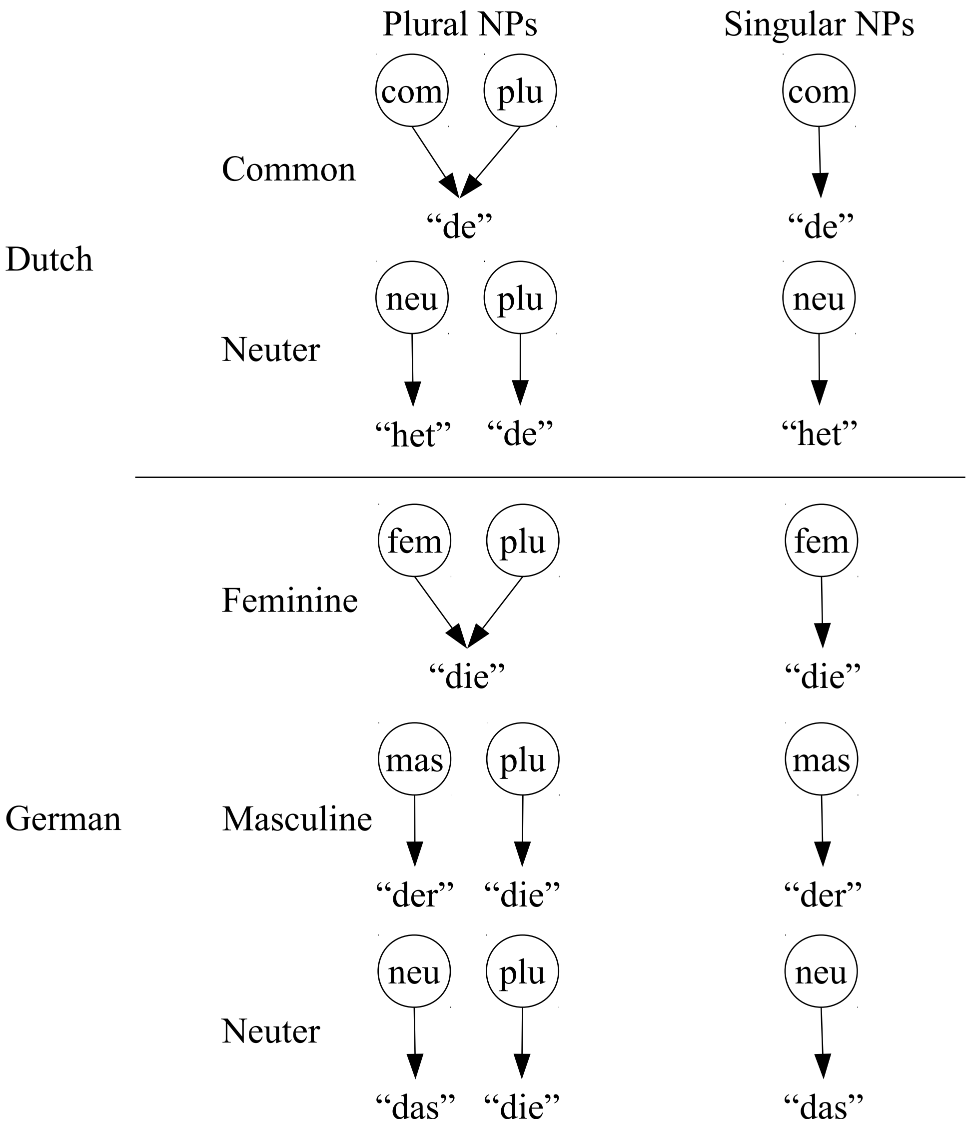

Figure 1. Graphical presentation of the relationship between the features grammatical gender, number and the form of the determiner in Dutch (top) and German (bottom) plural (left panel) and singular (right panel) NPs.

gender by number interaction where there is *a benefit in the production of plural NPs with common gender nouns* (compared to plural neuter nouns).[2]

In short, the competitive selection hypothesis predicts a gender by number interaction referred to here as *cost-type* interaction, and the noncompetitive

[2] In German, the determiner associated with feminine gender nouns plays a parallel role to the common gender in Dutch. The determiner used for singular feminine nouns is identical to that of plural nouns irrespective of gender (i.e. *die*). Accordingly, a competition hypothesis would predict a plural cost in the production of German plural determiner NPs with masculine and neuter gender nouns, and a noncompetitive hypothesis would predict a plural benefit in the production of plural determiner NPs with feminine gender nouns (see e.g. Schriefers et al., 2002 for a study using German).

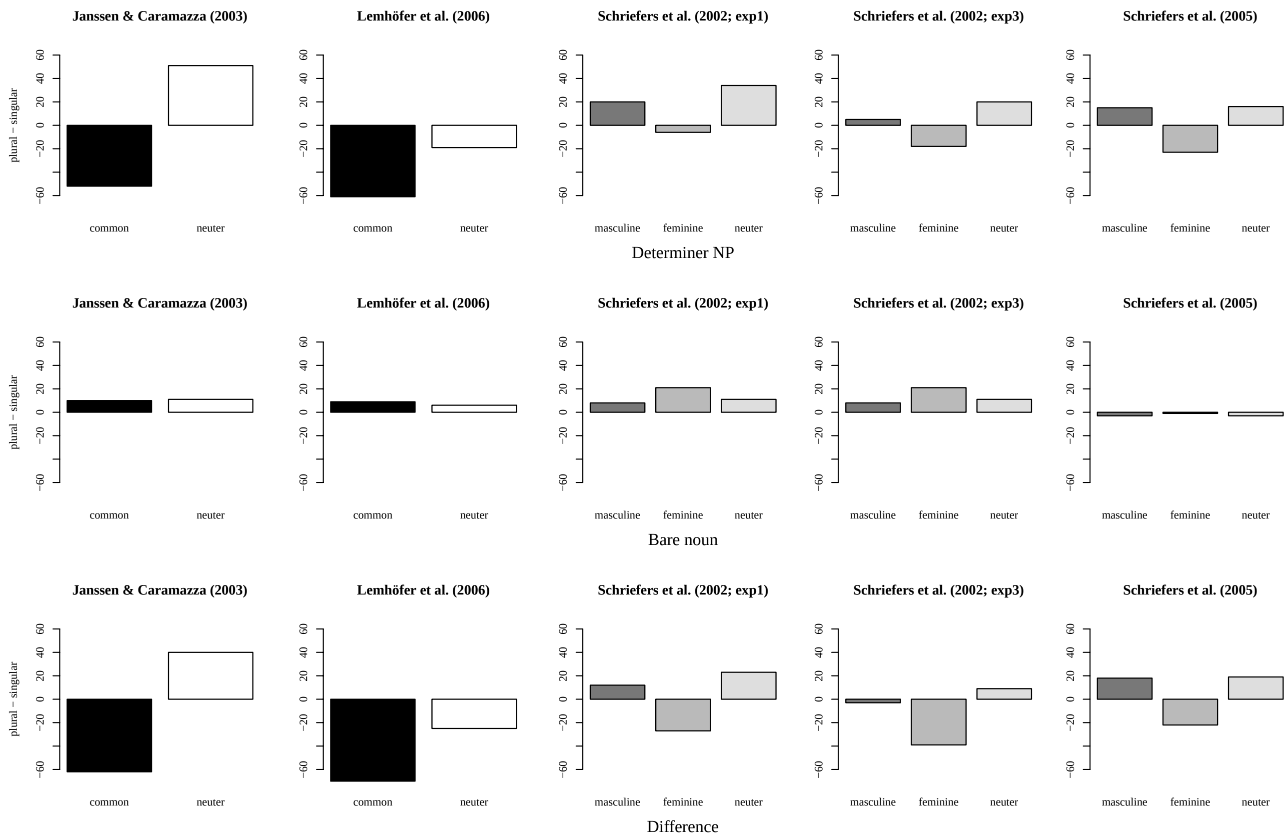

Figure 2. Summary of data reported for the determiner NP (top row) and bare noun (middle row) conditions of all Dutch (common and neuter gender) and German (masculine, feminine, and neuter gender) studies that have investigated determiner production in the SPN task. Difference (bottom row) refers to the determiner NP reaction times adjusted for the effect of number in the bare noun condition. The *y*-axis reflects the difference of performance between plural and singular NPs; positive numbers reflect a plural cost, and negative numbers reflect a plural benefit. There is little consistency in the direction of the most relevant effects, plotted under "Difference" (i.e. third row; See text for details).

hypothesis predicts a gender by number interaction referred to here as *benefit-type* interaction. In the following, we will show that the presence of gender by number interactions has been explored carefully, but that the exact shape of the interaction remains poorly understood. We will argue that the two hypotheses (cost- and benefit-type interactions) can be tested by analysing the data separately for each gender. Let us start by reviewing the available evidence in the way it has been previously reported. Figure 2 presents a graphical representation of the data from all available studies that have investigated determiner production using the SPN task. The relevant data at this point are plotted in the top row of Figure 2 under the label "determiner NP". Jescheniak et al. claim that a careful analysis of the results from the studies in Figure 2 support a competition hypothesis.

Consistent with the competition hypothesis, Schriefers et al. (2002) reported in their Experiment 1 using German a plural cost in the production of neuter gender nouns and a marginally significant plural cost in the production of masculine nouns ($p1 = .09$, $p2 = .06$). The plural benefit effect for feminine nouns was not significant. Janssen and Caramazza (2003), using Dutch, also revealed a plural cost in the production of neuter gender nouns, as well as a plural benefit in the production of common gender nouns. The remaining three data-points point in a different direction. In Schriefers et al. (2002; Experiment 3), the expected plural cost for masculine nouns was not significant, and the expected plural cost for neuter nouns was marginal ($p1 = .07$; $p2 = .17$); there was a plural benefit for feminine nouns. In Schriefers et al. (2005), using German, the expected plural costs for masculine and neuter gender nouns were again marginal (masculine: $p1 = .06$, $p2 < .05$; neuter: $p1 = .09$, $p2 < .05$); There was a plural benefit for feminine nouns. Finally, in Lemhöfer et al. (2006), using Dutch, the expected plural costs were absent, and only a plural benefit for common gender nouns was observed.

Jescheniak et al. claim that, despite this seemingly mixed set of results, a clear conclusion can be drawn if a specific observation made in Schriefers et al. (2002) is taken into account. That study revealed that the proportion of plural trials in an experiment affects the shape of the gender by number interaction. Specifically, Schriefers et al. found that in their Experiment 1 with 25% plural trials, a cost-type interaction was found, while in their Experiment 3 with 50% plural trials, a benefit-type interaction was found (see top row in Figure 2 for these studies). Schriefers et al. argued that the effect of the proportion of plural trials in the experiment was *additive* with the gender by number interaction. Specifically, it was assumed that the additive term reflected a speed up in the plural condition when it was more frequent (50%), which led to a reduction of plural costs, and an increase in plural benefits. Under such assumptions, both experiments revealed the same cost-type gender by number interactions, and hence Schriefers et al. (2002) interpreted the results in terms of a competitive mechanism for determiner selection. In Schriefers et al. (2005) and in Lemhöfer et al. (2006), the proportion of plural trials in the experiment was 50%. As mentioned above, these studies reported reliable plural benefits, and marginally or nonsignificant plural costs (Figure 2). Following the logic outlined above, these results were assumed to reflect a gender by number interaction of the cost-type that is affected (masked) by the additive effect of the high proportion of plural trials in the experiment. Following this logic, the results of Schriefers et al. (2005) and Lemhöfer et al. (2006) were thought to reflect a gender by number interaction of the cost-type with an additional additive term. Consequently, Jescheniak et al. interpreted the results from *all* studies obtained in the SPN task in terms of a competitive mechanism of determiner selection.

Jescheniak et al.'s interpretation of the results in terms of a competitive mechanism crucially hinges on the assumption that the various shapes of gender by number interactions observed within and across studies are driven by the proportion of plural trials in the experiment. However, there are at least two arguments against this view. First, the assumption that the different patterns of results reported in Figure 2 are driven by the proportion of plural trials in the experiment receives only weak empirical support. Only a single study has demonstrated that the proportion of plural trials affects the shape of the gender by number interaction (i.e. Schriefers et al., 2002), and neither Schriefers et al. (2005) nor Lemhöfer et al. (2006) have addressed this point explicitly. Given the importance of this assumption for adjudicating between competitive and noncompetitive models, one would want to see this assumption confirmed beyond a post-hoc proposal. Second, whereas Schriefers et al. (2002) showed an effect of this factor, other studies have not. In Experiment 1 of Janssen and Caramazza (2003), there were 40% plural trials, and in their Experiment 3 that included both diminutive and plural-diminutive conditions, there were 67% plural trials. However, this change in the number of plural trials did not have a critical impact on the shape of the gender by number interaction (the plural cost changed from 51 ms, $p1 < .001$, $p2 < .001$, to 41 ms, $p1 < .007$, $p2 < .001$). Finally, given that Schriefers et al. (2002) found a benefit type interaction with 50% plural trials, and that Janssen and Caramazza found a cost type interaction with 40% plural trials, one would need to conclude that a reduction of only 10% plural trials is sufficient for a complete reversal of the interaction type.

In short, Jescheniak et al.'s conclusion that the results from all SPN studies are consistent with a competitive selection hypothesis depends on the assumption that the proportion of plural trials in the experiment affects the gender by number interaction in an orderly manner. However, it is not clear that this assumption holds. The first goal of the experiments reported below was to test this assumption directly.

Before reporting that evidence, however, we need to take an even closer look at how the main effect of number modulates the shape of the interaction. As noted earlier, the SPN task relies on between item comparisons; in particular, singular and plural utterances are typically triggered by different pictures (e.g. with one vs. multiple instances of an item). To partial out any irrelevant effect of number from the theoretically relevant effects on determiner production, the determiner phrase data need to be contrasted with a condition involving the same materials but without determiner production (e.g. bare noun naming). Only differences in the shape of the gender by number interaction across these two formats will surely reflect determiner processing.

All studies appropriately report a bare noun control condition, in which the gender by number interaction should be absent if the effects are to be attributed to determiner rather than noun production.[3] However, while such a three-way interaction between gender, number, and format signals processes specific to determiner retrieval, it does not clarify in a straightforward manner whether the determiner pattern is of the cost- or benefit-type distinguished above.

To illustrate the important consequences of a main effect of number in bare noun naming for the shape of the gender by number interaction in determiner NPs, consider

[3] Lemhöfer et al. (2006) take into account this constraint by reporting significant three-way interactions between gender, number, and response format (in their design: bare noun, adjective noun, noun phrase). However, the critical three-way interaction in which only bare noun and noun phrases are considered was not reported.

the row labelled "Difference" in Figure 2 (bottom row). This row contains the latencies of the determiner NPs minus the main effects of number found in the bare noun naming condition (middle row) for each study. Note that, for example, in Schriefers et al. (2002; Experiment 1), a small benefit-type effect for feminine gender nouns in the determiner NPs increases in size once baseline differences between singular and plural conditions in the bare naming condition are taken into account. Given the importance of the benefit- versus cost-type direction of the effects for the interpretation in terms of a specific selection mechanism, it is clear that such main effects of number have to be taken into account. The available studies have not addressed this specific point in full.

A direct statistical test of this specific point can be performed by considering each grammatical gender separately. The core of the predictions of the competitive hypothesis concerns neuter nouns (in German, masculine, and neuter nouns), while the core of the predictions of the noncompetitive hypothesis concerns common nouns (in German, feminine nouns). With this distinction made explicit, the competitive account predicts *an extra cost in the production of plural NPs with neuter gender nouns compared to singular neuter NPs* once differences between singular and plurals are partialled out in a bare naming condition. In other words, this account predicts a format by number interaction of a specific shape. Symmetrically, the noncompetitive account predicts *a benefit in the production of plural NPs compared to singular common NPs*, a format by number interaction of the "opposite" shape. The second goal of the experiments reported below was to test these two predictions.

In the experiments below we evaluated whether differences in the proportion of plural trials in the experiment impacts the gender by number interaction in determiner NPs. In Experiment 1a, Dutch participants produced determiner NPs in an experiment in which there were 50% plural trials. On the basis of results obtained by Lemhöfer et al. (2006; and by Schriefers et al., 2005, in German), one would expect a gender by number interaction of the benefit type. In Experiment 1b, determiner NPs were produced in an experiment in which there were 40% plural trials. Given that Janssen and Caramazza (2003) have found a gender by number interaction of the cost-type with 40% plural trials, one would expect a gender by number interaction that mirrors that of Janssen and Caramazza in Experiment 1b.

As in previous reports, we also tested a condition in which participants produced bare nouns instead of determiner phrases. We computed the three-way interactions between format, gender, and number to test whether determiner retrieval *per-se* was indeed sensitive to the different combinations of gender and number features. In contrast with previous reports, we then tested two-way interactions between utterance format and number, *for each gender separately.* In this way we could directly test the critical theoretical predictions of the competitive and noncompetitive hypothesis, as they were stated two paragraphs earlier, that is for each gender separately. In support of this approach, it should be highlighted that, by construction, the sets of items of different gender are totally nonoverlapping; they are presented within a single experiment for practical reasons only (e.g. so that determiner choice alternates across trials). Given that the hypotheses tested do not rely on such mixture, their statistical test can be performed independently.

In Experiment 1a, participants produced determiner NPs with definite determiners, while in Experiment 1b, participants produced determiner NPs with distal demonstrative determiners (see Table 1 for examples). These different types of determiner NPs were used to ensure that the results would generalise between different types of determiner NPs.

Finally, following an expanding trend in psycholinguistics, the statistical analyses were based on linear mixed effect regression analyses conducted at the single trial level.

EXPERIMENT 1A—DETERMINER AND BARE NOUN NAMING WITH 50% PLURAL TRIALS

Participants

Forty native Dutch speakers, students at the Radboud University (Nijmegen, The Netherlands), took part in the experiment. These forty participants were divided into two groups and were randomly assigned to the determiner + adjective + noun and bare noun naming conditions. Each participant was paid three Euros upon completion of the experiment.

Materials and design

There were 48 pictures, half with a common gender and half with a neuter gender name. These materials were identical to those used by Lemhöfer et al. (2006; see Appendix 2). The 48 pictures appeared once in singular and once in plural. Thus, there were 96 items in the experiment, out of which 48 required a plural response (50% plural trials).

Singular NPs were elicited by the presentation of a single picture, and plural NPs were elicited by the presentation of two identical pictures (as in Janssen & Caramazza, 2003). Following the design of Lemhöfer et al. (2006), the adjective response was elicited by varying the size of the pictures. Note that in determiner + adjective + noun NPs, the adjective is not gender marked, and therefore does not impact our predictions regarding the selection of determiner forms.

The items were pseudo-randomised into stimulus lists with the following two constraints: (1) there were never more than three consecutive trials from the same gender, number, or size condition, and (2) there were no consecutive trials on which pictures names shared phonological onset or a semantic relationship. Finally, six additional pictures were selected for practicing the experimental task. From these six pictures, 12 practice trials were created.

Procedure

The experimental software was NESU (Nijmegen Experimental Set Up). Participants sat in front of a PC in a sound-proof booth. They wore headphones and spoke into a microphone. The microphone was connected to a button-box and provided voice-key measurements with 1 ms accuracy.

The experiment was preceded by a picture familiarisation phase, in which participants named the 48 experimental and 6 practice pictures. On each trial, a fixation appeared for 700 ms. Next, a picture appeared, and after 1 second the name of the picture appeared beneath the picture. The appearance of the picture's name was a cue for the participant to produce the name of the picture aloud. The picture and its name remained on the screen for 1 second. After a blank screen of 1,500 ms, the next trial started. Next, the participants practiced the experimental task. They were instructed to use singular or plural determiner + adjective + noun NPs, or singular and plural bare nouns, depending on their task group. On each trial, a fixation cross appeared for 700 ms, next the picture display appeared for 1,500 ms or until participants made a vocal response. Finally, there was a pause of 2,000 ms before the next trial started. After the practice phase, participants performed the experiment proper which was identical in trial structure to the practice phase. The experiment lasted about 20 minutes.

Analyses and results

Trials on which the participant produced a response that deviated from the response anticipated by the experimenter were removed from the analyses (8.9%). In addition, outliers identified by visual inspection of each individual participant's quantile–quantile plot were also discarded (1%). The remaining 3,459 data points were analysed using a linear mixed effects technique (Baayen, 2008), in which (untransformed) Reaction times (RTs) were fitted to a statistical model that included the fixed effect factors Format (determiner + adjective + noun vs. bare noun), Gender (common vs. neuter), Number (singular vs. plural), and their interactions, and random intercept estimates for Participants and Items. It was assumed that items (i.e. nouns) were crossed with the factors Format and Number.

Most importantly for us, there was a significant three-way interaction between utterance format, gender, and number [$t(3{,}451) = -2.61$, $p < .0103$; for a full description of the beta-coefficients, t- and p-values of each of the fixed effect factors and their interactions see Appendix 1]. This three-way interaction signals that the gender by number interaction was different in bare noun versus NP utterances, and hence that determiner retrieval per-se was sensitive to the combination of gender and number features. Figure 3 is a graphical presentation of the number effects (plural – singular) for each level of Gender across the two different values of Format in Experiments 1a (and 1b). We also plot the effects of number for the determiner NPs (corrected for the effects of number that may be present in the "baseline" the bare noun condition).

We then analysed the data for each gender separately (see Table 2). As argued above, the competitive hypothesis predicts an interaction between format and number of the cost-type. While there was a 19 ms interaction term, with a positive sign meaning cost-type, this term fell short of significance ($p = .11$). For common gender

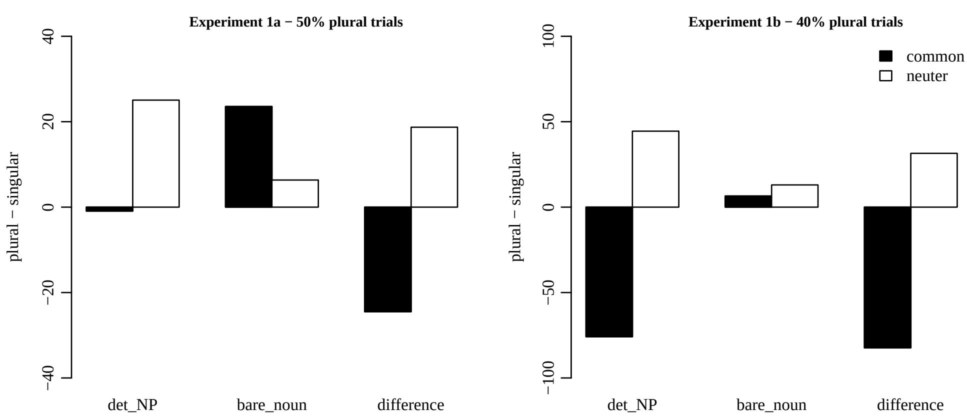

Figure 3. Graphical presentation of the effects in Experiments 1a and 1b. Det_NP = determiner + adjective + noun NP in Experiment 1a, and determiner + noun NP in Experiment 1b. Difference refers to the effects for the determiner NPs adjusted for the effects obtained for the bare nouns. Note that this graph presents effect sizes and not mean latencies.

TABLE 2
Beta-coefficients, *t*- and *p*-values for the format and number variables computed separately for each gender in Experiment 1a

Gender	*Variables*	*Beta-coefficients*	t[a]	p *(MCMC)*
Neuter	Intercept (bare, singular)	697.7	29.31	.0001
	Number = plural	6.3	0.75	.4498
	Format = det NP	50.1	1.70	.0428
	Format = det NP:number = plural	19.4	1.62	.1056
Common	Intercept (bare, singular)	658.1	31.63	.0001
	Number = plural	23.4	2.82	.0056
	Format = det NP	72.9	2.55	.0036
	Format = det NP:number = plural	−24.6	−2.09	.0422

Notes: [a]Degrees of freedom for common and neuter gender nouns were 1751 and 1701, respectively.

nouns, the noncompetitive hypothesis predicts that the format by number interaction should be of the benefit-type. This prediction was upheld, with a −25 ms significant interaction term. An overview of the mean RTs in each condition in Experiment 1a is given in Table 3.

EXPERIMENT 1B—DETERMINER AND BARE NOUN NAMING WITH 40% PLURAL TRIALS

Participants

Thirty-five participants were recruited from the same population as in Experiment 1a. Seventeen participants took part in the determiner + noun, and 18 in the bare noun naming condition. The bare noun naming condition was taken from Janssen and Caramazza (2003). None had taken part in Experiment 1a.

Materials and design

Materials and design were identical to those used by Janssen and Caramazza (2003; see Appendix 3), except that participants produced demonstrative instead of definite determiners. There were 60 pictures, half with common and half with neuter gender names. These 60 pictures appeared once in singular and once in plural, leading to a total of 120 experimental items. In these 120 experimental items, there are 90 items that require the determiner form "de" and 30 items that require "het". Following Janssen and Caramazza's design, another 30 filler pictures with neuter gender names that were named in singular were added. Thus, in total,

TABLE 3
Overview of the mean reaction times and error rates (between brackets) for the conditions in Experiment 1a

	Bare noun		*Determiner NP*	
	Common	*Neuter*	*Common*	*Neuter*
Singular	658 (5.6)	696 (6.9)	730 (10.4)	748 (12.1)
Plural	680 (9.0)	699 (8.1)	727 (6.9)	772 (13.1)

there were 150 items in the experiment, out of which 60 items required a plural response (=40% plural trials). The stimulus displays that were used to elicit the singular and plural NPs were created as in Experiment 1a. Finally, eight pictures that were not used in the previous set were selected as practice items. Half of these pictures had common and half had a neuter gender name. The items were pseudo-randomised as in Experiment 1a.

Procedure

The procedure was similar to the one used in Experiment 1a. The experimental software was Psyscope (Cohen, MacWhinney, & Flatt, 1993). In the determiner + noun condition, participants were told to use the gender-marked distal demonstrative determiner (e.g. die auto [that car]/dat boek [that boek]; die autos [those cars]/die boeken [those books]). Each naming condition lasted about 20 minutes.

Analyses and results

Analyses were identical to those in Experiment 1a. Trials on which the participants produced an unanticipated response were discarded (7.5%), as well as outliers identified based on visual inspection of the participants' quantile–quantile plots (2.1%). The remaining 3,806 data points were analysed with a statistical model that was identical to that used in Experiment 1a.

As was the case in the previous experiment, a significant three-way interaction between utterance format, gender, and number signals that determiner retrieval was sensitive to the combination of number and gender features [$t(3,798) = -5.15$, $p < .0002$, see Appendix 1 for statistical details of the full model]. We then analysed the data for each gender separately (see Table 4). The competition hypothesis predicts a cost-type interaction between format and number. The 22 ms interaction term was not significant ($p = .12$). For common gender nouns, the noncompetitive hypothesis predicts a benefit-type interaction. This prediction was upheld, with a −82 ms significant interaction term. An overview of the mean RTs in each condition in Experiment 1b is presented in Table 5.

TABLE 4
Beta-coefficients, *t*- and *p*-values for the format and number variables separately for each gender in Experiment 1b

Gender	*Variables*	*Beta-coefficients*	t^a	p *(MCMC)*
Neuter	Intercept (bare, singular)	751.2	31.37	.0001
	Number = plural	70.3	2.30	.0100
	Format = det NP	13.1	1.32	.1940
	Format = det NP:number = plural	22.3	1.54	.12
Common	Intercept (bare, singular)	742.9	30.59	.0001
	Number = plural	138.4	4.20	.0001
	Format = det NP	6.4	0.64	.5196
	Format = det NP:number = plural	−82.0	−5.65	.0001

Notes: [a]Degrees of freedom for common and neuter gender nouns were 1909 and 1889, respectively.

TABLE 5
Overview of the mean reaction times and error rates (between brackets) for the conditions in Experiment 1b

	Bare noun		*Determiner NP*	
	Common	*Neuter*	*Common*	*Neuter*
Singular	742 (3.3)	749 (4.3)	881 (11.8)	817 (8.4)
Plural	745 (3.7)	763 (6.0)	802 (8.8)	852 (13.9)

DISCUSSION

The results of the two experiments are at odds with Jescheniak et al.'s claim that reducing the proportion of plural trials in the experiment should change a benefit-type gender by number interaction into a cost-type interaction. In Experiment 1a, with 50% plural trials, we observed a gender by number interaction in which determiner bearing utterances showed plural benefits (with common gender nouns) and where plural costs were not significant ($p > .1$, with neuter nouns). The pattern was essentially the same in Experiment 1b, now with 40% plural trials, and a different type of determiner. Furthermore, a direct comparison of Experiment 1a and the data of Lemhöfer et al. (2006) show different interaction shapes for similar experimental designs with the same proportion of plural trials (50%), even when the main effects of number in bare noun naming are taken into account (compare Figures 2 and 3, study 2). Thus, our data do not support the claim that the proportion of plural trials in the experiment affects the shape of the gender by number interaction. To reiterate, our conclusion is also supported by the results of Janssen and Caramazza (2003), in which the same gender by number interaction was reported in two experiments despite clear changes in the proportion of plural trials in those experiments (40% vs. 67%).

In addition, and more importantly, the statistical evaluation based on the newly introduced tests conducted separately for each gender provided support for a noncompetitive account and no support for a competitive account. More generally, our reconsideration of the previously published evidence provides scarce support for a competitive mechanism. Such reconsideration could only be based on the evidence as it is available in the published report. Whether the same conclusions might be reached if separate tests were performed for each gender can only be known by reanalysing those data.

In short, this evidence undermines Jescheniak et al.'s conclusion that the data from the SPN task allow a consistent interpretation in terms of a competitive selection mechanism. Further research is necessary to uncover the reasons why these inconsistent patterns are observed. However, it should be clear that any interpretation of these data in terms of a competitive or noncompetitive selection mechanism is premature.

Up until now, the discussion has focused exclusively on the interpretation of effects in PWI and SPN tasks during the production of *free-standing* morphemes. According to Jescheniak et al., the available data on *bound morpheme* production suggest that these elements are also selected by a competitive selection mechanism. We agree with Jescheniak et al. that the gender congruency effect (inconsistently) observed in the PWI task during bound morpheme production (Bordag & Pechmann, 2008; Costa, Kovacic, Fedorenko, & Caramazza, 2003; Schiller & Caramazza, 2003; Schiller & Costa, 2006; Schriefers, 1993) remains ambiguous with respect to the alternative between competitive and noncompetitive mechanisms. Regarding the SPN task, Schriefers et al. (2005) and Lemhöfer et al. (2006) reported gender by number interactions of the benefit type during bound morpheme production, which Jescheniak et al. interpret in terms of competitive selection due to the proportion of plural trials in the experiment. However, as we have shown, the stability of the shape of the interaction remains to be ascertained more thoroughly. In addition, these studies also do not adequately take into account the number effects observed in bare noun naming. Thus, as with free-standing morpheme production, the results from the PWI and SPN tasks during bound morpheme production are inconclusive about the specific selection mechanism engaged.

Given that the available results from the PWI and the SPN tasks do not allow strong conclusions, it may be relevant to seek additional types of evidence. Here we will briefly mention two threads of research that should be considered on our way to an integrated model of closed-class production.

First, sentence production experiments have also been used to investigate the production of gender inflections (e.g. Badecker & Kuminiak, 2007; Franck, Vigliocco, Antón-Méndez, Collina, & Frauenfelder, 2008; Vigliocco & Franck, 2001; Vigliocco & Zilli, 1999). These studies involve experimental settings designed to elicit gender agreement errors (e.g. the equivalent of saying "rode" red$_{com}$ for "rood" red$_{neut}$). Semantic as well as morpho-phonological contextual information has been shown to contribute to gender inflection retrieval. Interestingly, the interpretation of these findings has been formulated in terms of agreement, a mechanism of syntactic processing involving features; it has not been explicit about the mechanism of bound morpheme selection (although see MacDonald and colleagues' investigations of number agreement: Haskell & MacDonald, 2003; Thornton & MacDonald, 2003). Thus, the agreement literature points to competition between features, a hypothesis that is largely rejected on the basis of the PWI and SPN data (see Jescheniak et al., in press). The relationship between determiner and bound morpheme retrieval with the gender and number agreement processes needs to be explicitly articulated (see also Schriefers & Jescheniak, 1999).

Second, there are a number of relevant studies on the role of morphology in the retrieval of familiar plural inflected nouns and verbs (e.g. Baayen, Levelt, Schreuder, & Ernestus, 2008; Bien, Baayen, & Levelt, 2011; Sereno & Jongman, 1997; Tabak, Schreuder, & Baayen, 2010). The conclusion reached on the basis of both comprehension and production studies is that familiar plural-inflected nouns and verbs are retrieved in their full-form, meaning that bound morphemes such as inflections are not retrieved independently from the stem. This conclusion is clearly at odds with the view of Jescheniak et al. that the production of a gender-marked adjective involves the concatenated retrieval of its stem and its gender-marked inflection (e.g. Levelt, Roelofs, & Meyer, 1999). Once again, an integrated discussion of these data together with those from the selection of gender-marked inflections would also move forward the research discussed here.

To conclude, Jescheniak et al. valuably broaden the discussion on the mechanism of lexical selection by considering evidence from the production of closed-class words. Their review of the literature on free-standing and bound morpheme production in PWI and SPN tasks suggests that there is a consistent pattern of results that supports a competitive mechanism of free-standing and bound morpheme selection. Our main goal in this commentary was to show that this is an overly optimistic portrayal of the present literature. We agree that the gender congruency effect in the PWI task is ambiguous between a competitive and noncompetitive account of free-standing and bound morpheme selection. We have argued, and empirically demonstrated, that the variable forms of the gender by number interaction observed in the SPN task do not warrant strong conclusions about the selection process. Clearly, in order to resolve these issues further empirical work is necessary that goes beyond the confines of a commentary. Such work should be complemented with a direct consideration of evidence from related research threads that have been construed independently up to now.

REFERENCES

Abdel Rahman, R., & Melinger, A. (2009). Semantic context effects in language production: A swinging lexical network proposal and a review. *Language and Cognitive Processes*, *24*, 713–734.

Alario, F.-X., Ayora, P., Costa, A., & Melinger, A. (2008). Grammatical and nongrammatical contributions to closed-class word selection. *Journal of Experimental Psychology: Learning, Memory & Cognition*, *34*, 960–981.

Alario, F.-X., & Caramazza, A. (2002). The production of determiners: Evidence from French. *Cognition*, *82*, 179–223.

Baayen, R. H. (2008). *Analyzing linguistic data. A practical introduction to statistics using R*. Cambridge, UK: Cambridge University Press.

Baayen, R. H., Levelt, W. M. J., Schreuder, R., & Ernestus, M. (2008). Paradigmatic structure in speech production. *Chicago Linguistics Society*, *43*(1), 1–29.

Badecker, W., & Kuminiak, F. (2007). Morphology, agreement, and working memory retrieval in sentence production: Evidence from gender and case in Slovak. *Journal of Memory and Language*, *56*, 65–85.

Bien, H., Baayen, R. H., & Levelt, W. J. M. (2011). Frequency effects in the production of Dutch deverbal adjectives and inflected verbs. *Language and Cognitive Processes*, *27*, 683–715.

Bordag, D., & Pechmann, T. (2008). Grammatical gender in speech production: Evidence from Czech. *Journal of Psycholinguistic Research*, *37*, 69–85.

Caramazza, A., Miozzo, M., Costa, A., Schiller, N. O., & Alario, F.-X. (2001). A cross-linguistic investigation of determiner production. In E. Dupoux (Ed.), *Language, Brain, and Cognitive Development: Essays in Honor of Jacques Mehler* (pp. 209–226). Cambridge, MA: MIT Press.

Cohen, J. D., Macwhinney, B., & Flatt, M. (1993). Psyscope: An interactive graphic system for designing and controlling experiments in the psychology laboratory using Macintosh computers. *Behavioral Research Methods, Instruments and Computers*, *25*, 257–271.

Costa, C., Kovacic, D., Fedorenko, E., & Caramazza, A. (2003). The gender congruency effect and the selection of freestanding and bound morphemes: Evidence from Croatian. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *29*, 1270–1282.

Foucart, A., Branigan, H. P., & Bard, E. G. (2010). Determiner selection in romance languages: Evidence from French. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *36*, 1414–1436.

Franck, J., Vigliocco, G., Antón-Méndez, I., Collina, S., & Frauenfelder, U. H. (2008). The interplay of syntax and form in language production: A cross-linguistic study of form effects on agreement. *Language and Cognitive Processes*, *23*, 329–374.

Haskell, T. R., & MacDonald, M. E. (2003). Conflicting cues and competition in subject–verb agreement. *Journal of Memory and Language*, *48*, 760–778.

Janssen, N., & Caramazza, A. (2003). The selection of closed-class words in noun phrase production: The case of Dutch determiners. *Journal of Memory and Language*, *48*, 635–652.

Janssen, N., & Caramazza, A. (2011). Lexical selection in multi-word production. *Frontiers in Psychology*, *2*, 81.

Jescheniak, J. D., Schriefers, H., & Lemhöfer, K. (2012). Selection of freestanding and bound gender-marking morphemes in speech production – A review. *Language and Cognitive Processes*. DOI: 10.1080/01690965.2012.654645

Lemhöfer, K., Schriefers, H., & Jescheniak, J. D. (2006). The processing of free and bound gender-marked morphemes in speech production: Evidence from Dutch. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *32*, 437–442.

Levelt, W., Roelofs, A., & Meyer, A. (1999). A theory of lexical access in speech production. *Behavioral & Brain Sciences*, *22*, 1–75.

Mahon, B. Z., Garcea, F. E., & Navarrete, E. (2012). Picture-word interference and the response exclusion hypothesis: A response to Mulatti and Coltheart. *Cortex*, *48*, 373–377.

Mulatti, C., & Coltheart, M. (2012). Picture-word interference and the response-exclusion hypothesis. *Cortex*, *48*, 363–372.

Schiller, N. O., & Caramazza, A. (2003). Grammatical feature selection in noun phrase production: Evidence from German and Dutch. *Journal of Memory and Language*, *48*, 169–194.
Schiller, N. O., & Costa, A. (2006). Different selection principles of freestanding and bound morphemes in language production. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *32*, 1201–1207.
Schriefers, H. (1993). Syntactic processes in the production of noun phrases. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *19*, 841–850.
Schriefers, H., & Jescheniak, J. D. (1999). Representation and processing of grammatical gender in language production: A review. *Journal of Psycholinguistic Research*, *28*, 575–600.
Schriefers, H., Jescheniak, J. D., & Hantsch, A. (2002). Determiner selection in noun phrase production. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *28*, 941–950.
Schriefers, H., Jescheniak, J. D., & Hantsch, A. (2005). Selection of gender-marked morphemes in speech production. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *31*, 159–168.
Schriefers, H., & Teruel, E. (1999). Phonological facilitation in the production of two-word utterances. *European Journal of Cognitive Psychology*, *11*, 17–50.
Sereno, J. A., & Jongman, A. (1997). Processing of English inflectional morphology. *Memory & Cognition*, *25*, 425–437.
Tabak, W., Schreuder, R., & Baayen, R. H. (2010). Producing inflected verbs: A picture naming study. *The Mental Lexicon*, *5*(1), 22–46.
Thornton, R., & MacDonald, M. C. (2003). Plausibility and grammatical agreement. *Journal of Memory and Language*, *48*, 740–759.
Vigliocco, G., & Franck, J. (2001). When sex affects syntax: Context influences in sentence production. *Journal of Memory and Language*, *45*, 368–390.
Vigliocco, G., & Zilli, T. (1999). Syntactic accuracy in sentence production: Gender disagreement in Italian language impaired and unimpaired speakers. *Journal of Psycholinguistic Research*, *28*, 623–648.

APPENDIX 1

Beta-coefficients (with standard error), *t*-, and *p*-values for each of the fixed effect factors (and their interactions) in Experiment 1a. *p*-Values were obtained from 10,000 Markov Chain Monte Carlo samples

Factors	β *(*SE*)*	t *(3,451)*	p *(MCMC)*
Intercept	681.6 (22.3)	30.48	.0001
Format = det NP	48.5 (29.6)	1.64	.0514
Gender = neuter	22.8 (13.9)	1.65	.0960
Number = singular	−23.6 (8.3)	−2.83	.0060
Format = det NP: Gender = neuter	21.2 (11.9)	1.78	.0710
Format = det NP: Number = singular	24.5 (11.8)	2.07	.0326
Gender = neuter: Number = singular	17.2 (11.8)	1.46	.1390
Format = det NP: Gender = neuter: Number = singular	−43.9 (16.8)	−2.61	.0102

Beta-coefficients (with standard error), *t* values, and *p*-values for each of the fixed effect factors (and their interactions) in Experiment 1b. *p*-Values were obtained from 10,000 Markov Chain Monte Carlo samples.

Factors	β *(*SE*)*	t *(3,798)*	p *(MCMC)*
Intercept	749.4 (24.1)	32.62	.0001
Format = det NP	56.5 (31.9)	1.77	.0504
Gender = neuter	14.8 (16.5)	0.90	.3594
Number = singular	−6.5 (9.9)	−0.66	.5084
Format = det NP: Gender = neuter	36.5 (14.5)	2.51	.0136
Format = det NP: Number = singular	82.4 (14.4)	5.70	.0001
Gender = neuter: Number = singular	−6.5 (14.0)	−0.46	.6580
Format = det NP: Gender = neuter: Number = singular	−105.5 (20.5)	−5.15	.0001

APPENDIX 2

Overview of items in Experiment 1a

	Number	
Gender	*Singular*	*Plural*
Common	borstel (brush)	borstels
	spons (sponge)	sponzen
	schaar (scissors)	schaar
	vos (fox)	vossen
	bijl (axe)	bijlen
	lepel (spoon)	lepels
	aap (monkey)	apen
	riem (belt)	riemen
	douche (shower)	douches
	sigaar (cigar)	sigaren
	bel (bell)	bellmen
	tent (tent)	tenten
	lamp (lamp)	lampen
	kroon (crown)	kronen
	bril (glasses)	brillen
	jurk (dress)	jurken
	bus (bus)	bussen
	vogel (bird)	vogels
	kast (cupboard)	kasten
	Spiegel (mirror)	spiegels
	hond (dog)	honden
	stoel (chair)	stoelen
	tafel (table)	tafels
	deur (door)	deuren
Neuter	skelet (skeleton)	skeletten
	hert (deer)	herten
	anker (anchor)	ankers
	konijn (rabbit)	konijnen
	tapijt (carpet)	tapijten
	zadel (saddle)	zadels
	zwaard (sword)	zwaarden
	net (net)	netten
	nest (nest)	nesten
	hek (fence)	hekken
	touw (rope)	touwen
	pistol (gun)	pistolen
	kasteel (castle)	kastelen
	mes (knife)	messen
	gewicht (weight)	gewichten
	oor (ear)	oren
	dak (roof)	daken
	eiland (island)	eilanden
	slot (lock)	sloten
	vuur (fire)	vuren
	paard (horse)	paarden
	raam (window)	ramen
	boek (book)	boeken
	bed (bed)	bedden

APPENDIX 3
Overview of items in Experiment 1b

	Number	
Gender	*Singular*	*Plural*
Common	arm (arm)	armen
	auto (car)	autos
	beer (bear)	beren
	berg (mountain)	bergen
	bijl (axe)	bijlen
	borstel (brush)	borstels
	bus (bus)	bussen
	deur (door)	deuren
	fiets (bike)	fietsen
	gieter (watering can)	gieters
	hond (dog)	honden
	jurk (dress)	jurken
	kaars (candle)	kaarsen
	kikker (frog)	kikkers
	krant (newspaper)	kranten
	lamp (lamp)	lampen
	leeuw (lion)	leeuwen
	molen (windmill)	molens
	pijl (arrow)	pijlen
	raket (rocket)	raketten
	riem (belt)	riemen
	sigaar (cigar)	sigaren
	stekker (plug)	stekkers
	stoel (chair)	stoelen
	trompet (trumpet)	trompetten
	vlieg (fly)	vliegen
	voet (foot)	voeten
	vogel (bird)	vogels
	vos (fox)	vossen
	wortel (carrot)	wortels
Neuter	anker (anchor)	ankers
	bed (bed)	bedden
	been (leg)	benen
	boek (book)	boeken
	bureau (desk)	bureaus
	fornuis (oven)	fornuizen
	hart (heart)	harten
	hek (fence)	hekken
	hert (deer)	herten
	kasteel (castle)	kastelen
	kompas (compass)	kompassen
	konijn (rabbit)	konijnen
	mes (knife)	messen
	nest (nest)	nesten
	net (net)	netten
	oor (ear)	oren
	paard (horse)	paarden
	pistol (gun)	pistolen
	potlood (pencil)	potloden
	raam (window)	ramen
	schaap (sheep)	schapen
	skelet (skeleton)	skeletten

APPENDIX 3 *(Continued)*

Gender	Number	
	Singular	*Plural*
	slot (lock)	sloten
	touw (rope)	touwen
	varken (pig)	varkens
	vergiet (colander)	vergieten
	vest (vest)	vesten
	wiel (wheel)	wielen
	zadel (saddle)	zadels
	zwaard (sword)	zwaarden