

# Representation and selection of determiners with phonological variants

Audrey Bürki, Tea Besana, Gaëlle Degiorgi, Romane Gilbert, F.-Xavier Alario

▶ **To cite this version:**

Audrey Bürki, Tea Besana, Gaëlle Degiorgi, Romane Gilbert, F.-Xavier Alario. Representation and selection of determiners with phonological variants. Journal of Experimental Psychology: Learning, Memory, and Cognition, In press, 45 (7), pp.1287-1315. 10.1037/xlm0000643 . hal-02094540

**HAL Id: hal-02094540**
**https://hal.science/hal-02094540**

Submitted on 9 Apr 2019

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Representation and selection of determiners with phonological variants

Audrey Bürki[1,2], Tea Besana[1], Gaelle Degiorgi[1], Romane Gilbert[1], & F.-Xavier Alario[3]

[1] Faculty of Psychology and Educational Sciences, University of Geneva, Geneva, Switzerland

[2] Department of Linguistics, Cognitive Sciences, University of Potsdam, Potsdam, Germany

[3] Aix Marseille Université, CNRS, LPC UMR 7290, Marseille, France

**Abstract**

The aim of this study is to contribute to a better understanding of cross-linguistic differences in the time course of determiner selection during language production. In Germanic languages, participants are slower at naming a picture using a determiner + noun utterance (*die Katze* 'the cat') when a superimposed distractor is of a different gender (gender congruency effect). In Romance languages in which the pronunciation of the determiner also depends on the phonology of the next word, there is no such effect. This difference is traditionally assumed to arise because determiners are selected later in Romance languages (*late selection hypothesis*). It has further been suggested that in a given language, all determiners are either selected late or early (*maximum consistency principle*). Data on French have challenged these two hypotheses by revealing a gender congruency effect when participants name pictures using the definite singular determiner *le-la* (*l'* before vowels) and a noun, at positive Stimulus Onset Asynchrony (SOA), i.e., when there is a delay between the presentation of the picture and that of the distractor. We examined this finding further and investigated whether it generalises to the indefinite determiner *un-une*. Results of four picture-word interference experiments reveal that gender congruency effects in French are not restricted to the definite determiner or positive SOAs, but can be hard to detect in experiments which do not account for the variability in reading and naming times across participants and trials. We discuss the implications of these results for the modelling of determiner selection across languages.

**Introduction**

Languages differ in many respects, including their phoneme inventory, the way sounds are actually pronounced, their morphology, or their syntax. Psycholinguists have long recognized that cross-linguistic differences provide a unique window into the cognitive mechanisms and representations underlying the processing of speech (e.g., Cutler, Mehler, Norris, & Segui, 1983; Cutler, Sebastián-Gallés, Soler-Vilageliu, & van Ooijen, 2000; Hawkins, 2007; Mehler, Dupoux, Pallier, & Dehaene-Lambertz, 1994). In research on language production, processes are however often assumed to be similar across languages. Despite a recent increase in the interest for cross-linguistic comparisons (e.g., Brown-Schmidt & Konopka, 2008; Dumay & Damian, 2011; Ganushchak, Konopka, & Chen, 2014; Norcliffe, Konopka, Brown, & Levinson, 2015; see also Jaeger & Norcliffe, 2009), most language production issues are not examined from a cross-linguistic perspective. One notable exception concerns the production of determiners (e.g., Caramazza, Miozzo, Costa, Schiller, & Alario, 2001). Studies on determiner production have repeatedly pointed to important differences across languages in the data and therefore in the processes underlying determiner selection and noun phrase planning (see Caramazza et al, 2001, for a review). In the present study, we report an empirical investigation about existing inconsistencies in both the observed findings and the theoretical accounts regarding these differences.

Most of the empirical evidence on cross-linguistic differences in determiner selection comes from the alternation between the presence and absence of a grammatical gender congruency effect in the picture-word interference task (data from the alternative bare noun naming task are reviewed in Janssen, Schiller, & Alario, 2014). In a seminal study, Schriefers (1993) asked Dutch participants to name pictures using determiner-adjective-noun utterances or adjective-noun utterances in response to a visually presented picture with a superimposed written distractor. In these noun phrases, the grammatical gender of the noun is reflected in the pronunciation of the determiner or adjective (e.g., ***de*** *groene stoel*$_{masc}$ 'the green chair' *vs*. ***het*** *groene bed*$_{neutral}$ 'the green bed'; ***groene*** *stoel*$_{masc}$ 'green chair' *vs*. ***groen*** *bed*$_{neutral}$ 'green bed'). In Schriefers' study, the distractor was either of the same or a

different gender as the target word. Gender incongruency between targets and distractors slowed down naming latencies (i.e., gender congruency effect). Schiller and Caramazza (2003) further observed that this effect only surfaced when the pronunciation of the determiner was modulated by the noun's gender, as in the Dutch examples above, or in German singular noun phrases, e.g., ***der*** *Tisch*$_{masc}$ 'the table' *vs*. ***die*** *Katze*$_{fem}$ 'the cat'. There was no such effect for determiners whose pronunciation is the same across genders (e.g., German plural noun phrases with the definite determiner, e.g., ***die*** *Tische*$_{masc}$, ***die*** *Katzen*$_{fem}$). The authors concluded that the gender congruency effect originates during determiner form selection (phonological encoding). Determiner forms receive activation from both the target word and distractor word, and the selection process is slowed when two different determiner forms receive activation. The gender congruency effect has been replicated several times in Dutch and German (Bürki, Sadat, Dubarry, & Alario, 2016; Heim, Friederici, Schiller, Rüschemeyer, & Amunts, 2009; La Heij, Mak, Sander, & Willeboordse, 1998; van Berkum, 1997). By contrast, the majority of studies in Romance languages reported an absence of effect (Alario & Caramazza, 2002; Costa, Sebastian-Galles, Miozzo, & Caramazza, 1999; Miozzo, Costa, & Caramazza, 2002). The gender congruency effect is absent even when tested at different SOAs (i.e., stimulus onset asynchrony, or time lag between the presentation of the target picture and that of the distractor word; Miozzo et al., 2002).

Caramazza et al. (2001) explained this discrepancy across languages with the so-called *late selection hypothesis* (see also Costa, Alario, & Sebastián-Gallés, 2007). In Dutch or German, the determiner form can be selected as soon as the gender of the noun is known because it does not depend on the phonological properties of the following word (e.g., the feminine form of the German singular definite determiner is always *die*, irrespective of the phonology of the following word). By contrast, in Romance languages, both the gender of the noun and the phonology of the following word are required to select the appropriate determiner form, at least for a subset of the determiners (henceforth "alternating determiners"). For instance, in Italian, the masculine definite singular determiner is pronounced (and spelt) *l* before vowels, *lo* before /ʃ/, /s + consonant/, /gn/, /ts/, and

/dz/, and is pronounced *il* in all other cases. The feminine forms are *la* before consonants and *l'* before vowels. In Spanish, the definite masculine determiner is always *el,* while the definite feminine determiner is *la* in all phonological contexts but one, i.e., when the next word starts with a stressed /a/. Similarly, in French, the pronunciation of most determiners varies with both the gender of the noun and the phonology of the next word. For instance, the masculine and feminine singular definite determiners are realized as *le* and *la* before consonants and as *l'* before vowels. According to the *late selection hypothesis* (Caramazza et al., 2001), during the production of noun phrases with alternating determiners, by the time the phonological properties of the noun have been accessed and determiner selection occurs, the activation of other determiner forms generated by the processing of the distractor has decayed and no longer interferes. Costa et al. (1999) observed that gender congruency effects in Romance languages are also absent for determiners whose pronunciation does not depend on the phonology of the next word (e.g., the Spanish masculine definite determiner). In order to account for this finding, Caramazza et al. (2001) further hypothesized that the time point at which a given determiner form is selected in a given language is the same for all determiners. If one determiner must be selected later because its pronunciation depends on the phonological context, all determiners in the language will be selected later. They termed this hypothesis the "maximum consistency" or "highest common denominator" principle.

The *late selection hypothesis* and *maximum consistency principle* offer a straightforward account of determiner selection across languages. They further provide a theoretical framework to study and model several fundamental issues in linguistic/psycholinguistic research. For instance, the *late selection hypothesis* allows deriving predictions about the representation of alternating determiners. It implicitly assumes that the two pronunciations of a given determiner (e.g., *le* and *l* for the masculine definite determiner) have a corresponding abstract phonological representation (i.e., lexeme) in the speaker's long term memory, challenging assumptions in both psycholinguistic models of word production (e.g., Garrett, 1980) and in many phonological accounts of words with alternating pronunciations (e.g., Dell, 1975; Encrevé, 1988; Selkirk, 1978; Tranel, 2000). The *late selection*

*hypothesis* further makes specific claims regarding the time course of phonological planning within utterances, by assuming that, in Romance languages, the phonological encoding of the noun precedes that of the determiner. This assumption challenges the long standing view that the phonological encoding process operates from begin to end ("left to right"; e.g., Meyer, 1990; 1991). Finally, the *maximum consistency principle* makes clear assumptions regarding the principles that govern phonological planning within languages, by assuming that the time course of phonological planning is the same for all determiners.

Notably, the *late selection hypothesis* and *maximum consistency principle* are challenged by empirical findings on the gender congruency effect in French with the singular definite determiners *le* and *la* (Foucart, Branigan, & Bard, 2010; Schriefers & Teruel, 1999, see below for details). The available empirical data and their interpretation suffer from inconsistencies, however. In the present study, we address these inconsistencies by attempting to replicate and extend the empirical findings and discussing the data in the broader context of phonological variation. In the remainder of the Introduction, we summarize the available evidence regarding the gender congruency effect in Romance languages, describe the potential of gender congruency manipulations to address important issues in phonology, and introduce the current study in more details.

*Gender congruency effects in French*

As mentioned above, gender congruency effects are usually absent in Romance languages. Data from French offer however a more complex picture. Alario and Caramazza (2002) did not observe the effect. Schriefers and Teruel (1999) manipulated the gender congruency between target and distractor for two types of French utterances. Both started with the definite singular determiner, which is realized as *le* before consonant-initial masculine nouns (e.g., *le chat*$_{masc}$ 'the cat'), as *la* before consonant-initial feminine nouns (e.g., *la table*$_{fem}$ 'the table'), and as *l'* before masculine and feminine vowel-initial nouns (e.g., *l'âne*$_{masc}$ 'the donkey', *l'armoire*$_{fem}$ 'the cupboard'). The first type of utterances had a vowel-initial noun and therefore always started with the determiner form *l' (e.g.,*

*l'arbre*$_{masc}$ *rouge/vert* 'the red/green tree'; l'armoire$_{fem}$ rouge/verte 'the red/green cupboard'*)*, the second type of utterances had a consonant-initial noun and therefore started with *le* or *la* depending on the gender *(e.g., le bureau*$_{masc}$ *rouge/vert* 'the red/green desk'; *la table*$_{fem}$ *rouge/verte* 'the red/green table'*)*. Schriefers and Teruel tested different SOAs (-150 ms, 0 ms, 150 ms, and 300 ms). They observed a gender congruency effect at an SOA of 300 ms for utterances starting with *l'*. For the second type of utterances, a gender congruency effect was significant by Participant (but not by Item) at an SOA of 0 ms. The authors concluded that the gender congruency effect occurs irrespective of whether the determiner form varies with the gender of the noun. They suggested that differences in the SOAs at which the gender congruency effect occurs across noun phrase types might be related to the different sets of nouns used in the two noun phrase types. Foucart et al., 2010) also focussed on the French definite determiners *le* and *la,* testing utterances in which the full forms *le* or *la* have to be produced. They observed a gender congruency effect at 200 ms but not at 0 SOA.

At first sight, these results can be explained, at least partly, within the *late selection hypothesis*. As mentioned above, this hypothesis assumes that determiner selection occurs later in French (and other Romance languages). Consequently, if the distractor is processed later in time, the determiner form activated by the processing of the distractor should have a higher chance of influencing the selection of the to-be-produced determiner. If this was the case, however, a gender congruency effect should be observed at later SOAs for other French determiners and, beyond that, in other Romance languages.

Foucart et al. (2010) provide an account of their results which allows reconciling the presence of a gender congruency in French and its absence in other Romance languages. Their account diverges from the *late selection hypothesis* and *maximum consistency principle*. First, their proposal introduces "adjustment" processes in addition to selection processes (see also Garrett, 1980; 1984). In their view, the forms [le] and [la] are stored in the mental lexicon with a single underlying representation each, and the form [l] produced before vowel-initial words results from a late

phonetic[1] adjustment. For instance, when a speaker says *le chat* 'the cat' or *l'âne* 'the donkey', s/he first accesses the determiner form */le/*; the vowel is deleted later on, as a result of a late adaptation process. The presence of a gender congruency effect at +200 SOA occurs because the adjustment process delays the selection of the determiner. In their own words, "*le/la would always be selected by 'default' but could only be 'fully' selected once the post-selection rule had been checked with reference to the local context"* (p. 1416). Secondly, their account postulates that determiner selection processes may not be homogenous within a given language. According to Foucart et al., the adjustment process is unlikely to apply to all French determiners. *Le* and *la* share properties that make them good candidates for a late adjustment process. The two pronunciations of these determiners are closely related phonologically, the form used before vowel-initial words being a mere reduction of the form used with consonants. Moreover, both genders call for the same form before vowel-initial words (i.e., [l]). Foucart et al. (2010) argue that determiners which do not share these properties are unlikely to be governed by a late adjustment rule and should thus not show congruency effects at +200 SOAs. Accordingly, the processes and representations underlying the production of determiners are not necessarily uniform within languages.

Importantly, the data currently available for French and their interpretation call for further investigations. First, there is some inconsistency in the findings between Foucart et al. (2010); Schriefers and Teruel (1999), and Alario and Caramazza (2002) with regard to the SOA at which the effect surfaces for the full forms *le* and *la*. The difference between the full and reduced pronunciations (i.e., [l]) reported by Schriefers and Teruel also requires empirical confirmation. Most crucially, none of these studies report robust statistical evidence that the gender congruency effect

[1] The term *phonetic* is borrowed from Foucart et al. (2010). Note that in the context of dominant psycholinguistic models of language production (e.g., Levelt, Roelofs, & Meyer, 1999), where the phonological encoding process deals with the generation of abstract phonological sequences of words and the phonetic process with the computation of articulatory gestures, the adjustment rule they describe would likely take place during the phonological encoding process, given that it influences the selection of the determiners' phonological representations.

interacts with the time at which the distractor is presented. For instance, in Foucart et al., the interaction between SOA and gender congruency is not significant in the by-participant analysis, suggesting that it might not be replicated with a different group of participants. Schriefers and Teruel (1999) did not report interactions between gender congruency and SOA, or between gender congruency and utterance type. The data collected in the present study were partly meant to address these caveats.

*Gender congruency effects as a window into phonological variation phenomena*

The study of gender congruency effects in Romance languages is not only informative about determiner selection and the planning of noun phrases; it also has the potential to contribute to fundamental debates in Phonology. A major issue in the field concerns the representations and mechanisms involved in phonological variation (e.g., Anderson, 1982; Bybee, 2001; Côté, 2005; Dell, 1985; Encrevé, 1988; Levelt, 1989; Selkirk, 1978; Tranel, 2000). The terms *Phonological variation* describe cases where a word has two (sometimes more) pronunciations. The determiners tested in psycholinguistic studies on Romance languages, whose pronunciation varies with the phonological properties of the next word, are instances of phonological variation phenomena. Phonological variation can be found in most languages, and can take different forms. The question of the representations and mechanisms involved arises for each specific phenomenon. The French language has many examples of phonological phenomena, and many of them can be found in determiners. The deletion of vowels in some contexts (as in the French definite determiner *la* realized as [l]) is an example of phonological variation. Schwa deletion is another example. In French, when a monosyllabic word ending with the letter "e" (e.g., *le*) is produced in isolation, the "e" is realized as [ə] (or schwa). When the same word occurs before a vowel-initial word, the schwa is not pronounced. According to most textbooks, the schwa must be pronounced before consonant-initial words but in practice, French native speakers often also delete the schwa before consonants in conversation utterances (e.g., *j'ai pris l'train* 'I took the train'). Similarly, French polysyllabic words

whose first syllable ends with an unaccented "e" (*cheval* 'horse' or *fenêtre* 'window'), can be produced both with and without the schwa (e.g., [fənɛtR] / [fnɛtR]). French liaison is yet another example of phonological variation. Many words ending with a written consonant are pronounced without this consonant when the word is produced in isolation or is followed by a consonant-initial word but are produced with the consonant when they are followed by a vowel-initial word, e.g., *grand chat* [gRãʃa] 'big cat' vs. *grand ami* [gRãt̲ami] 'great friend'). Important issues here are whether words with more than one pronunciation are represented in memory with only a default form or whether both pronunciations are represented in memory, and whether this depends on the type of phenomenon.

The *late selection hypothesis* implicitly assumes a representation for each determiner form, whereas Foucart al.´s adjustment process assumes that one form (here *l'*) is not represented but derived via a local adjustment process. The study of gender congruency thus offers us a unique mean to test how determiners are represented in the speakers' lexicon. This is particularly valuable given the scarcity of paradigms that allow testing the predictions of contrasting accounts of variation phenomena (see also Bürki, Frauenfelder, & Alario, 2015).

*The current study*

In four experiments, we examine the production of French determiners. We investigate further the hypotheses that the selection of some determiners in French requires a late adjustment followed by a checking mechanism, and that the time course of determiner selection depends on the properties of the determiner.

French offers an ideal test case to examine these issues. Most determiners in the language have several pronunciations but with varying combinations of possibilities with regard to (1) how phonological and gender information combine to determine the pronunciation of the determiner and (2) with regard to the type of change that occurs between the two pronunciations of a given determiner (see also Alario & Caramazza, 2002). The only French determiner for which a gender

congruency effect was tested so far is the definite singular determiner. As mentioned above, the two genders share the same pronunciation before vowels, and the alternation between the "full" and "reduced" pronunciations can easily be accounted for by a general phonological process (see Côté & Morrison, 2007 for a summary of the phonological literature on *le* and other clitics). Other determiners in French differ from the singular definite determiners in one or several dimensions. The French indefinite determiner for instance is realized (and written) *un* [œ̃] before consonant-initial words while the feminine form is "*une*" [yn]. Before vowel-initial words, the masculine form is realized with an additional liaison ([œ̃n] or [œn]). The addition of a liaison consonant is a general phenomenon and can easily be (and has repeatedly been) accounted for by an adjustment process. Unlike for the definite determiner however, the two genders do not share a pronunciation before vowel-initial words.

In the present study, we focus on the definite and indefinite determiners *le-la* and *un-une*. Our motivation for examining the indefinite determiner *un-une* in addition to *le-la-l'* is twofold. First, this determiner allows contrasting different hypotheses regarding the time course of determiner selection in a given language and how the properties of the determiners may influence this time course. Foucart et al. explicitly predicted that when "the phonetic adjustment rule affects only one of the genders, the *maximum consistency principle* is applied and the determiner is selected late" (p. 1419). Accordingly, no gender congruency effect should be found for this determiner, irrespective of the gender (assuming that the gender congruency effect at 200 SOA only arises when a late adjustment process followed by a checking mechanism occur)[2]. According to an alternative hypothesis, determiner forms are selected as soon as the relevant information is available. The

[2] Foucart et al. offer yet a third possible account. This account builds on the assumption that the indefinite determiner does not vary at all with the local context (p. 1419). Under this assumption, determiner forms can be selected early and both genders should show a gender congruency effect at earlier SOAs. Note that the fact that the pronunciation of the masculine indefinite determiner *un* varies with the following context is undebated and we doubt here that Foucart et al. (2010)'s intention was to question this. We assume that Foucart et al. implicitly refer to the lack of alternation of the determiner in the spelling. The change in pronunciation of the masculine indefinite determiner *un*, unlike for the definite determiner, is not reflected in the spelling. One could assume a difference in processing between determiners whose alternating pronunciations are reflected by different spellings and those whose alternating pronunciations share a spelling.

feminine determiner *une* would be selected as soon as the gender of the noun is known (i.e., as early as in German languages). By contrast, and following the logic of Foucart et al., for the masculine determiner, the default form /œ̃/ would be selected by default, but only fully selected once the phonological context has been checked. This alternative hypothesis predicts a gender congruency effect for the masculine determiner at 200 SOA, and a gender congruency effect for the feminine determiner at 0 SOA.

Secondly, the indefinite determiner *un-une* was chosen because the alternation between the two pronunciations for the masculine determiner is a specific instance of the French liaison process. By studying this determiner, we further hope to gain novel insights into the representations and mechanisms underlying this process. If this determiner shows a gender congruency effect at 200 SOA for the masculine form, this would suggest that the production of liaison consonants results from the application of a phonological rule that insert (or delete) the consonant (e.g., Tranel, 1981).

**Experiments 1a-1b**

The aims of the first experiment were to replicate the gender congruency effect at 200 SOA for the French definite determiners *le-la*, test whether this effect interacts with the actual pronunciation of the determiner (full form *vs*. reduced form), and assess whether the effect extends to the indefinite determiners *un-une*.

***Participants***

Twenty-six participants, all French speaking psychology students at the University of Geneva, participated in the study. None of them reported hearing or language disorders. They were paid or received course credit for their participation. In this and subsequent experiments, participants were given details about the experimental procedure and provided their informed consent before the experimental session. All experiments received ethical approval by the "Comité d'éthique" of the University of Geneva.

***Materials***

We selected 96 pictures; half corresponded to vowel-initial nouns and half to consonant-initial nouns (see Appendix A). In each of these two groups, half the nouns were feminine and half were masculine. The number of occurrences per million (as given for films by the database *Lexique*, (Matos, Ferrand, Pallier, & New, 2001) for these words ranged from 0.3 to 389, with a mean of 23.9. We also selected 148 words to be used as distractors. Each picture was paired with two distractors. Some distractors were paired with more than one picture. The first distractor was of the same gender as the picture (congruent condition), the second was of a different gender (incongruent condition). The two distractors and the picture had the same type of onset, i.e., they all started either with a vowel or with a consonant. Distractors in the congruent and incongruent conditions were balanced in terms of lexical written frequency (as given for books by the database *Lexique*, 87 vs. 84 occurrences per million for congruent and incongruent distractors, respectively), number of letters (5.6 vs. 5.4 for congruent and incongruent distractors, respectively), and number of phonemes (4.1 vs. 4.3 for congruent vs. incongruent distractors). There was no semantic or phonological overlap between pictures and distractors. We also added a baseline condition, where the pictures appeared with a superimposed line of Xs.

In all conditions, pictures appeared as a black outline on white screen, framed in the same-size black square. A distractor word or line of Xs on a white rectangle was superimposed on each picture. The position and size of the white rectangle were identical across conditions for a given picture. As in Miozzo and Caramazza (1999) the position of the distractor differed slightly among picture targets. Three experimental lists were created. Pictures were identical across lists but distractor type (congruent, incongruent and line of Xs) were differently distributed over the pictures in each list. Each picture appeared once and in a different condition in each list. Each participant was presented with one of the three lists. The order of presentation of the pictures was randomized for each participant. Examples of stimuli for each condition are presented in Figure 1.

- Insert Figure 1 about here -

***Procedure***

The experiment was implemented in the DMDX software (Forster & Forster, 2003). Participants were tested at the University of Geneva. They completed the experiment individually, in a quiet booth. They were first asked to perform a familiarization task. During this task, they were presented with all the pictures of the experiment and their corresponding written nouns, and were told that they would later be asked to name these pictures. Participants were then asked to name the pictures while ignoring a printed distractor or line of Xs superimposed on them. Each trial had the following structure: a fixation cross was shown at the centre of the screen for 500 ms, followed by the presentation of the picture. 200 ms after onset picture presentation, the distractor (or line of Xs) appeared on the picture and both stayed there for another 2000 ms. A 500 blank screen interval separated trials.

In Experiment 1a participants were instructed to produce the noun preceded by the definite determiner *le* or *la* 'the' depending on the gender of the noun. In Experiment 1b, they were instructed to use the indefinite determiner *un* or *une* 'a' depending on the gender of the noun. The order of the two experiments was counterbalanced across participants.

***Data analyses***

In all experiments, we first analysed the naming latencies (time interval between picture onset and vocal response onset) with mixed-effects regression models (e.g., Baayen, Davidson, & Bates, 2008; Goldstein, 1987) using the statistical software R (R Development Core Team, 2017). Alpha was set at .05. The dependent variable was the (untransformed) naming latency. The predictors of interest were Gender congruency, Determiner (Experiments 1a-b, 2a-b), Gender (Experiment 1b), SOA (0 ms *vs*. 200 ms, Experiments 2 a-b, 3 and 4), the onset of the noun (Experiment 1a), and continuous measures (centered around the mean) accounting for the participants' processing speed in reading the distractor, or for differences in processing speed in naming the picture and reading the distractor

(in Experiment 3 and 4). For these two continuous variables, we tested both the linear term and the quadratic term. To include the quadratic term in the model, we computed the log of the square of the continuous measure. The log transformation was applied in order to avoid convergence problems due to scale differences between predictors.

All dichotomous variables were dummy coded. Several interactions between these variables were also tested. Non-significant interaction terms were always removed from the models, the results we present all stem from models without these non-significant terms. The position of the trial in the experiment (uncentered) was always entered as a covariate to account for changes in performance across the experiment. Statistical details for this variable are not reported. In all but one statistical model, this variable was significant, naming latencies were longer for items presented later in the experiment. Participants and items were entered as crossed random effects, including intercepts and slopes to allow these variables to be differently influenced by the predictors of interest. The correlation components between slopes and intercepts were not entered in the model. Random slopes were removed when the model failed to converge or when their variance was equal to 0.

Following Baayen (2008), model residuals larger than 2.5 standard deviations were considered outliers and removed, independently for each model. The results reported stem from the models without these outliers. Denominator degrees of freedom and p-values for F-tests were computed based on Satterthwaite's approximations with the Rpackage *lmerTest* (Kuznetsova, 2014).

The statistical analyses described above rely on null-hypothesis testing (see for instance Wagenmakers, 2017 or Vasishth & Nicenboim, 2016) and are often referred to as “frequentist statistics”. In this framework, the null hypothesis (i.e., no difference between conditions) can only be rejected, it can never be accepted. Bayesian analyses offer an alternative framework, in which it is possible to quantify the evidence for competing hypotheses. Here, we relied on Bayes factors (see for instance Rouder, Speckman, Sun, Morey, and Iverson, 2009; or Jarosz & Wiley, 2014 for an introduction). Bayes factors consist in comparing the likelihood of the data under competing

hypotheses or models and can be used to compare the likelihood of the data under the researcher´s hypothesis and under the null hypothesis. The Bayes factor is computed by dividing the likelihood of the data given H0 by the likelihood of the same data under the researcher´s hypothesis and quantifies the evidence in favor of the null hypothesis in a straightforward manner. A Bayes factor of 3 means that the data are three times more likely under the null than under the alternative hypothesis, while a Bayes factor of 1/3 means that the data are three times more likely under the alternative hypothesis.

In the current work, computing frequentist statistics was important for comparison purposes with previous work. Adding Bayesian statistics was important given the possibility of observing null effects, whose likelihood is to be assessed against alternative models. Hence, we report both types of analysis.

Bayesian statistics were performed with the software JASP (JASP Team, 2015, see Love et al., 2015) using the default priors implemented in the software (for technical details on Bayesian ANOVAs, see Rouder, Morey, Speckman, and Province, 2012). Following Jeffreys (1961), Bayes factors greater than three, ten, and thirty are considered respectively as “some evidence”, “strong evidence” and “very strong evidence” for a hypothesis over the other.

***Results***

Two participants used the wrong determiner in more than 50% of cases, their data were disregarded. The percentage of correct responses for the remaining participants was between 72 and 99%. These participants made 571 errors out of 4608 trials (12%), of which 37 % concerned the noun, 34% were dysfluencies, 16% concerned the determiner, 11% were no response and 2% mispronunciations. We further removed the five responses with yawns and two responses for which it was difficult to set the onset of the vocal response. The distribution of the naming latencies for the remaining 4030 data points was checked for extreme values and the 59 data points above 1800 were removed (1.5%). The

3971 remaining data points were included in the statistical analyses. Figure 2 represents the descriptive statistics for Experiments 1a and 1b (see Appendix B for details).

- Insert Figure 2 about here –

The data of Experiments 1a and 1b were first analysed separately, this way we could include different sets of predictors and interactions for each determiner. For the definite determiner, we tested the effect of Gender congruency and its interaction with the onset of the noun. For the indefinite determiner, we tested the effect of Gender congruency and its interaction with Gender. The two experiments were then analysed together to test for the interaction between Gender congruency and Determiner.

*Statistical analysis of Experiment 1a:* A first statistical model was conducted to check whether the experiment succeeded in replicating the lexical interference effect, i.e., longer naming latencies when the picture is presented with a superimposed written word than with a superimposed line of Xs (baseline condition). The naming latencies in the baseline condition were faster than in the congruent (β = 45.9, SE = 10.2, $t$ = 4.5, $p$ = 0.00014) and incongruent conditions (β = 48.8, SE = 12.6, $t$ = 3.9, $p$ = 0.00065, model without a by Item random slope for the contrast between incongruent and baseline). The faster naming latencies in the baseline condition show that the distractors were processed to some substantial extent.

The data points from the baseline condition were removed from further analysis. The next model examined the effects of Gender congruency, Noun onset and their interaction. The interaction between the two factors was not significant (β = 26.5, SE = 20.2, $t$ = 1.3, $p$ = 0.19) and was removed from the model. The model without the interaction revealed no main effect of Gender congruency (β = 4.5, SE = 0.13.6, $t$ = 0.33, $p$ = 0.74, model without a by Item random slope for Gender congruency).

A Bayesian repeated measures ANOVA was conducted, which compared a model with Noun onset and Participant (null model) to a model with Gender congruency and to a model with Gender congruency and its interaction with Noun onset. The results of this analysis are presented in Table 1.

Following Wagenmakers (2017) the table should be interpreted as follows. **P(M)** indicates the distribution of prior model probabilities across all the models tested, and **P(M|data)** indicates the posterior model probabilities. The column **$BF_{01}$** indicates, for each model, how much more likely the data are under the null model compared to this model. The column **$BF_M$** provides the Bayes factors that compares one model against all others (i.e., difference between prior model odds and posterior model odds). The **$BF_M$** value here indicates that the model which receives the most support is the null model. The **$BF_{01}$** of 3.178 in the Model comparison table indicates that the data are about three times less likely under a model with Gender congruency than under the null model. Adding the interaction increases the degree of support for the null model by a factor of 4.206/3.178 = 1.323. Note that the **$BF_{01}$** of 4.2 does not allow us to conclude much on the usefulness of the interaction, as this value compares a model with both the main effect and the interaction to the null model. The analysis of effects table allows us to compare a model with an effect (or interaction) to all other models without this term. The **$BF_{Inclusion}$** value provides us with this information. Here the data are 3.178 (1/0.315) times more likely under a model without Gender congruency than under all models with this effect. Moreover, the data are 1.32 (1/0.756) times more likely under a model without the Gender congruency * Noun onset interaction than under all models with this term (note that in more complex models, the values in the two tables will differ). To summarize, this analysis provides some support for the absence of a gender congruency effect but is inconclusive with regard to the presence of the interaction between Gender congruency and Noun onset.

-Insert Table 1 about here-

*Statistical analysis of Experiment 1b.* A first statistical model was conducted to check whether the experiment succeeded in replicating the lexical interference effect. The naming latencies in the baseline condition were faster than in the congruent (β = 28.7, SE = 10.1, $t$ = 2.8, $p$ = 0.008) and incongruent conditions (β = 54.6, SE = 10.1, $t$ = 5.4, $p < 0.0001$). The faster naming latencies in the baseline condition show again that the distractors were processed to some substantial extent. The data points from the baseline condition were removed from further analysis. The next model

examined the effect of Gender congruency and its interaction with Gender. The interaction between the two factors was not significant (β = 18.1, SE = 21.3, *t* = 0.85, *p* = 0.39, model with no by Item or by Participant random slope for the gender congruency effect) and was removed from the model. The model without the interaction revealed a main effect of Gender congruency, with longer naming latencies for incongruent trials (β = 21.6, SE = 9.6, *t* = 2.3, *p* = 0.034) but no effect of Gender (β = 21.9, SE = 16.2, *t* = 1.4, *p* = 0.18, model with no by Item or by Participant random slope for the gender congruency effect[3]).

A Bayesian repeated-measures ANOVA was conducted, which compared a null model including Gender and Participant to a model with only a main effect of Gender congruency and a model with Gender congruency and its interaction with Gender. The results of this analysis are presented in Table 2. The **BF** $_{M\ in}$ Table 2.A shows that the model with Gender congruency receives the most support but the Bayes Factor is too low (1.67) to be conclusive. The data are 3.37 (1/0.297) less likely under a model with the interaction than under a model with only Gender congruency. This analysis is inconclusive with regard to the main effect of Gender congruency and favors the hypothesis that there is no interaction between Gender congruency and Gender.

-Insert Table 2 about here-

*Joint analysis of Experiments 1a-1b*. The model with Gender congruency, Determiner, and the interaction between these two variables revealed no main effect of Determiner (β = 11.5, SE = 20.8, *t* = 0.55, *p* = 0.59) or Gender congruency (β = 11.7, SE = 7.8, *t* = 1.5, *p* = 0.15). The model with the interaction revealed no significant interaction (β = 22.5, SE = 12.9, *t* = 1.7, *p* = 0.083).

The corresponding Bayesian repeated measures ANOVA is presented in Table 3. Participant, Determiner, and Gender were included in the null model. The null Model receives the most support.

[3] These random terms were removed because their variance had a value of zero. The results do not differ when these random slopes are kept in the model.

This analysis provides no support in favour or against the inclusion of Gender congruency in the model, but suggests that for each interaction term, the data are less likely under a model with the interaction than under models without it by a factor of at least 3.

-Insert Table 3 about here-

***Discussion***

To summarize, the frequentist analysis provides some evidence in favour of a gender congruency effect for the indefinite determiners *un-une* while the Bayes factors regarding this effect are inconclusive, and further suggest that Gender congruency does not interact with Gender. For the determiners *le-la*, there is no gender congruency effect in the frequentist analysis, and the Bayes factor favours the null hypothesis. The analysis is inconclusive with regard to the interaction between Gender congruency and Noun onset.

The gender congruency effect for the indefinite determiner clearly awaits additional empirical evidence. The absence of gender congruency effect for *le-la* contrasts with the findings of Foucart et al. (2010). It is possible that we failed to find evidence in favour of the gender congruency effect with *le-la* because of specific properties of our materials. There is evidence that some effects in the picture-word interference paradigm are modulated by the properties of the distractors. La Heij et al. (1998) report for instance that the gender congruency effect decreases with less familiar distractors. Miozzo and Caramazza (2003, see also Scaltritti, Navarrete, & Peressotti, 2015) compared the naming latencies in trials without a distractor (baseline) and trials with an unrelated distractor and found more interference for low than for high frequency distractor words. In an attempt to maximize the chances of observing a gender congruency effect at 200 SOA for *le-la*, we conducted Experiment 2, in which we used the very materials and procedure used in Foucart et al. (2010). Like these authors, we manipulated the SOA, with distractors at 0 and 200 SOA, to examine further the hypothesis that gender congruency effects for *le-la* only occur at late SOAs. Finally, as in Experiment 1, we also tested

the indefinite determiner *un-une*. Different groups of participants performed the experiments with *le-la* (Experiment 2a) and *un-une* (Experiment 2b).

**Experiment 2a**

***Participants***

Nineteen participants from the same pool as Experiment 1 (4 males) took part in the experiment. None of them had participated in Experiment 1. They received monetary compensation for their participation.

***Materials***

We used the same stimuli used in Foucart et al., Experiment 1 [4] (see Appendix C). These consisted in 48 pictures of inanimate objects, 24 masculine and 24 feminine. The number of occurrences per million (as given for films by the database *Lexique*, Matos et al., 2001) for these words ranged from 1.4 to 389, with a mean of 27.6. All the picture names started with a consonant. Thirty-two words were used as distractors (frequency per million around 300, mean number of letters = 5). Each picture was paired with a gender congruent, a gender incongruent distractor, and a line of Xs. Distractors in the congruent and incongruent conditions had a similar value of lexical written frequency (as given for books by the database *Lexique*, 304 vs. 290 occurrences per million for congruent and incongruent distractors, respectively), number of letters (5.1 vs. 4.9 for congruent and incongruent distractors, respectively), and number of phonemes (3.8 vs. 3.4 for congruent vs. incongruent distractors). Note that three distractor-target word pairs are semantically related, and two share the first phoneme or syllable.

In addition, thirty-two pictures representing consonant-initial nouns were selected as fillers, and only presented with a line of Xs. Each participant was presented with 16 experimental trials in each

[4] We thank Alice Foucart for sharing this material with us.

condition, and only saw a given picture once. The order of the pictures was randomized for each participant.

***Procedure***

The procedure for this experiment was identical to that used in Foucart et al. (2010, Experiment 1), except that the software DMDX was used. Participants first performed a familiarization task in which they saw all the pictures and corresponding nouns, and a picture naming task. In the latter, each trial started with a fixation point for 500 ms, then, after a 500 ms blank screen interval the picture was presented. In the 0 SOA condition the distractor appeared with the picture, in the 200 SOA condition it appeared 200 ms after the picture. The two SOA conditions were presented in separate blocks. Half the participants started with an SOA of 0, the other half with an SOA of 200 ms. The picture and distractor remained on the screen until a response was provided, for a maximal duration of 2000 ms.

**Experiment 2b**

***Participants***

Twenty participants from the same pool as for Experiment 1 (3 males) took part in the experiment. None of them had participated in Experiment 1 or 2a.

***Materials and Procedure***

The Materials and Procedure were the same as for Experiment 2a, except that the participants had to use the determiner *un* or *une* instead of *le or la*.

***Results***

One participant had less than 50% of correct responses in Experiment 2a and her data were disregarded. The remaining participants had between 73% and 98% of correct responses. These participants made 350 errors out of 3648 trials (10%), of which 33 % concerned the noun, 33% were dysfluencies, 28% concerned the determiner, 4% were no response and 3% mispronunciations. We

further removed the two responses with yawns and two responses for which it was difficult to set the onset of the vocal response.

The distribution of the naming latencies for the 3294 remaining correct responses were checked for extreme values and the 27 data points above 1800 ms, together with two negative values were removed. The analyses were based on the 3265 remaining data points. Figure 3 represents the mean naming latencies broken down by gender congruency and SOA for Experiment 2a and 2b. See Appendix D for details.

- Insert Figure 3 about here –

*Statistical analysis of Experiment 2a:* A first statistical model was conducted to check whether the experiment succeeded in replicating the lexical interference effect. The naming latencies in the baseline condition were faster than in the congruent (β = 45.7, SE = 8.04, $t$ = 5.7, $p$ < 0.0001) and incongruent conditions (β = 58.5, SE = 10.7, $t$ = 5.5, $p$ < 0.0001). There was no effect of SOA (β = 38.6, SE = 19.44, $t$ = 1.99, $p$ = 0.062, model without a by Item random slope for the contrast between congruent and baseline). The faster naming latencies in the baseline condition show that the distractors were processed to some substantial extent. The data points from the baseline condition were removed from further analysis. The next model examined the effects of Gender congruency and SOA. The interaction between the two factors was not significant (β = 15.9, SE = 15.6, $t$ = 1.02, $p$ = 0.31) and was removed from the model. There was no significant effect of Gender congruency (β = 13.6, SE = 13.5, $t$ = 1.0, $p$ = 0.34) or SOA (β = 45.2, SE = 23.1, $t$ = 1.96, $p$ = 0.064).

The corresponding Bayesian repeated measures ANOVA is presented in Table 4. Participant and SOA were included in the null model. The null model receives the most support (**$BF_M$**= 6.06). The data are four times more likely under a model without Gender congruency than under models with this effect, and they are about 3 times less likely under a model with the interaction between SOA and Gender Congruency than under a model without it. This analysis favours the hypothesis that there is no gender congruency effect and no interaction between Gender congruency and SOA.

-Insert Table 4 about here-

*Statistical analysis of Experiment 2b:* A first statistical model was conducted to check whether the experiment succeeded in replicating the lexical interference effect. The naming latencies in the baseline condition were faster than in the congruent (β = 81.0, SE = 12.9, $t$ = 6.3, $p$ < 0.0001) and incongruent conditions (β = 92.8, SE = 13.3, $t$ = 7.0, $p$ < 0.0001). The effect of SOA was not significant (β = 15.5, SE = 18.1, $t$ = 0.86, $p$ = 0.40). The two contrasts interacted with SOA (β = 38.6, SE = 15.3, $t$ = 2.5, $p$ = 0.012, and β = 59.4, SE = 15.2, $t$ = 3.9, $p$ < 0.0001) showing that the lexical interference effect was stronger at 200 SOA. The faster naming latencies in the baseline condition show that the distractors were processed to some substantial extent. The data points from the baseline condition were removed from further analysis. The next model examined the effects of Gender congruency, Gender, SOA, and their interactions. The three-way interaction was not significant and was removed from the model (β = 55.6, SE = 33.0, $t$ = 1.69, $p$ = 0.092, model without a random slope for Gender congruency by Participant). The model including all the two-way interactions revealed they were not significant (all $p$ > 0.16, model without a by Participant random slope for Gender congruency) and thus these interaction terms were removed as well. In the final model there was no main effect of Gender congruency (β = 0.53, SE = 9.0, $t$ = 0.06, $p$ = 0.95). There was a main effect of SOA (β = 65.7, SE = 19.4, $t$ = 3.4, $p$ = 0.0025), with faster latencies at 200 SOA.

The corresponding Bayesian repeated measures ANOVA is presented in Table 5. The null model includes Participant, Gender, and SOA. The null model is again the model with the most support ($\mathbf{BF_M}$ = 17.2). The data are about 5.5 less likely under a model with Gender congruency than under models without it, they are more than 3.5 less likely under models with one of the 2-way interactions than under all models without. The evidence with regard to the three-way interaction is inconclusive. In short this analysis favors the hypothesis that there is no gender congruency effect and no interaction between Gender congruency and SOA.

-Insert Table 5 about here-

*Joint analysis of Experiments 2a-2b.* The details of the mixed-models and repeated measures Bayesian ANOVA for the two experiments are presented in Appendix E. There is no evidence that Gender congruency influences naming latencies or that this influence is modulated by determiner, SOA, or Gender.

***Discussion***

The frequentist analyses show no evidence that responses to gender congruent and gender incongruent picture-word pairs differ at 200 or 0 SOA, for the determiner *le-la* or *un-une*. The Bayesian analysis suggests that the data are more likely under the null hypothesis.

To further assess this absence of effect, we acknowledge that the finding by Foucart et al. (2010) that the gender of the distractor matters at 200 but not at 0 SOA together with the interpretation of this finding, suggests that the precise timing at which the distractor is processed is crucial for the gender congruency effect in French to manifest itself. Similarly, the time course of presentation/processing of the distractor also influences whether a gender congruency effect is found in Germanic languages. For instance, in Schriefers (1993), the gender congruency effect is present at SOAs of 0 and 200 but not at an SOA of 450. Using auditory distractors, Schriefers and Teruel (2000) found a gender congruency effect in German at an SOA of 150, but not at SOAs of -150 and 0 in a first experiment, and reliable effects at an SOA of 75, but not at SOAs of -300, -225, -150, -75, 0, 150, 225, or 300 in a second experiment.

Under the assumption that the timing of distractor processing modulates the gender congruency effect, it can be hypothesized that differences in distractor processing speed across participants may modulate the gender congruency effect. Even though we used the same materials as in Foucart et al. (2010), it is possible (and likely) that our participants had different processing times for the distractors and this may have influenced the SOA at which the effect would have surfaced. In a recent study, Bürki (2017) found evidence that the phonological facilitation effect (shorter naming latencies when picture and distractor share phonemes/letters) interacts with the difference in

processing times for target (as assessed in the picture-word interference naming task) and distractor (as assessed in a word reading task). In Experiment 3, we examine whether the gender congruency effect also depends on distractor processing times or differences in processing times, between distractor and target word.

**Experiment 3**

***Participants***

Twenty-nine students from the same pool as for Experiment 1 and 2 (4 men) took part in this experiment. They were paid or given course credit for their participation.

***Materials and Procedure***

The materials and procedure for this experiment were identical to those of Experiment 2. In addition, the reading performance of the participants was assessed in several tasks. A first series of off-line tasks were taken from the ECLA-16+ (« Evaluation des Compétences de Lecture chez l'Adulte de plus de 16 ans, Gola-Asmussen, C., Lequette, G., Pouget, C., Rouyet-Nicolas, C., et Zorman, M., 2010). Participants were asked to read a text as quickly as possible and were stopped after one minute, their score was computed by subtracting the number of errors from the number of words read. They were further asked to read a list of 20 regular words, a list of 20 irregular words, and a list of 20 pseudowords. Total reading time and number of errors were computed for each list. Participants further performed an on-line reading task during which they were asked to read words that appeared on the screen one by one as quickly as possible. Among these words were the 32 distractor words and 32 additional words, of similar length/frequency (see Appendix F for details). Participants were asked to read this word list twice and their reading latencies were recorded. Results in the ECLA-16+ revealed that participants were good readers. There was little variation in the participants' scores for these tasks.

***Results***

Two participants were excluded because they used a wrong determiner or no determiner at all. The remaining participants had between 76% and 98% of correct responses. These participants made 312 errors out of 2592 (12%), of which 40% were dysfluencies, 29% concerned the noun (wrong noun provided), 25% concerned the determiner (no or wrong determiner), 4% were no response and 2% mispronunciations. The visual inspection of the data further led us to remove the 50 data points below 180 ms or above 1450 ms. The analyses were performed on the 2230 remaining data points. Figure 4 represents the mean naming latencies broken down by gender congruency for each SOA. Details are provided in Appendix G.

- Insert Figure 4 about here -

A first statistical model was conducted to check whether the experiment succeeded in replicating the lexical interference effect. The naming latencies in the baseline condition were faster than in the congruent (β = 50.5, SE = 8.5, $t$ =5.9, $p$ < 0.0001) and incongruent conditions (β = 59.7, SE = 9.3, $t$ =6.4, $p$ < 0.001). There was no effect of SOA (β = 18.7, SE = 12.9, $t$ =1.5, $p$ = 0.16), and no significant interaction between SOA and Gender congruency (*ps* > 0.19).

The data points from the baseline condition were removed from further analysis. The model comparing the congruent and incongruent conditions at each SOA revealed no effect of Gender congruency (β = 9.04, SE = 8.9, $t$ = 1.02, $p$ = 0.32) or SOA (β = 26.5, SE = 17.0, $t$ = 1.6, $p$ = 0.13, model without a by Item random slope for SOA). The model with the interaction between Gender congruency and SOA revealed no such interaction (β = 2.7, SE = 14.6, $t$ = 0.18, $p$ = 0.86, model without a by Item random slope for SOA).

A Bayesian repeated measures ANOVA was conducted. SOA and Participant were included in the null model. The output is given in Table 6 and shows that the null model receives the most support ($\mathbf{BF_M}$ = 6.14). The data are 3.9 less likely under a model with Gender congruency than under a model without it, and 3.7 times less likely under a model with the interaction between Gender congruency

and SOA than under a model without it. This analysis provides support for the hypothesis that there is no gender congruency effect or interaction between Gender congruency and SOA.

-Insert Table 6 about here-

*Interaction between processing times for distractor and gender congruency effect*

We examined whether the gender congruency effect interacted with (1) participants' mean reading latencies for all words in the word reading task or (2) differences in processing times for the distractor and target word. Based on a visual inspection of the distribution of response latencies in the word reading task we identified the 17 data points above 700 ms as outliers and removed them from the dataset. We then computed a difference measure of processing time for the distractor and target word by subtracting the distractor reading latency from the target picture naming latency, for each picture-word combination. This measure was then centered around the mean. Figure 5 presents the distribution of the two measures. The naming latencies and reading latencies for the corresponding distractors were moderately correlated ($r = 0.20$, $t(1465) = 7.7$, $p < 0.001$).

- Insert Figure 5 about here -

The first statistical model included as fixed effects SOA, Gender congruency, the participants' mean reading latencies (linear and quadratic terms) as well as the interactions between each continuous measure (linear and quadratic term) and the two factors (the model had no by Item random slope for the linear term, and for the effect of SOA). The full model revealed no three-way interaction with the linear term ($p = 0.66$), or the quadratic term ($p = 0.068$). These interactions were removed. The model with two-way interactions revealed no such interaction (all $p$s > 0.52, model with no by Item random slope for the linear term, and for the effect of SOA). In the model without interactions, there was no main effect of SOA (β = 27.8, SE = 16.9, $t = 1.6$, $p = 0.11$), no main effect of Gender congruency (β = 7.3, SE = 9.0, $t = 0.81$, $p = 0.14$), no effect of the quadratic term (β = 10.8, SE = 7.1, $t = 1.5$, $p = 0.14$). There was a main effect of the linear term, with increasing naming latencies for participants with higher mean reading latencies (β = 0.99, SE = 0.35, $t = 2.8$, $p = 0.00911$).

The second statistical model included as fixed effects SOA, Gender congruency, a linear and quadratic term for the difference measure (Naming latency minus reading latency for each picture-word association), as well as the interactions between these terms and the two factors (model without by Participant or by Item random slopes for SOA). The full model revealed no three-way interaction with the difference measure ($p$ = 0.79), or with the quadratic term ($p$ = 0.74). These were removed. The model with two-way interactions revealed no such interaction (all $p$s > 0.46). In the model without interactions, there were effects of SOA, with shorter naming latencies at 200 SOA (β = 180, SE = 2.7, $t$ = 67.3, $p$ < 0.0001), of both the linear and quadratic term accounting for the difference measure, with increasing naming latencies for greater difference measures (β = 0.93, SE = 0.0064, $t$ = 143.8, $p$ < 0.0001 and β = 1.49, SE = 0.5, $t$ = 2.96, $p$ = 0.0032). In this model, the effect of Gender congruency was significant and in the expected direction (β = 13.1, SE = 5.3, $t$ = 2.49, $p$ = 0.016).

***Discussion***

Experiment 3 revealed no effect of Gender congruency at any of the tested SOA when reading times were not considered. Taking into account the variance accounted by differences in processing times between picture and distractor in the model resulted in significant effects of SOA and Gender congruency. As a final reliability test, Experiment 4 is the replication of Experiment 3 with a different group of participants.

**Experiment 4**

***Participants***

Thirty-one participants, all between 18 and 35 years took part in Experiment 4 (seven men). They had no reported hearing or language impairment. They were paid for their participation.

***Material and Procedure***

Experiment 4 was an exact replication of Experiment 3 with a different group of participants.

***Results***

One participant was excluded because he systematically used the indefinite determiner instead of the definite determiner. The remaining participants had between 80% and 99% of correct responses. These participants made 274 errors out of 2880 (10%) trials, of which 42 % were dysfluencies, 33% concerned the noun (wrong noun provided), 21% concerned the determiner (no or wrong determiner), 4% were no response and 1% mispronunciations. The visual inspection of the data further led us to remove the 63 data below 180 ms or above 1450 ms. The analyses were performed on the 2543 remaining data points. Figure 6 represents the mean naming latencies broken down ba distractor type for each SOA. Details are provided in Appendix G.

- Insert Figure 6 about here -

The first statistical model was conducted to check whether the experiment succeeded in replicating the lexical interference effect. The naming latencies in the baseline condition were faster than in the congruent (β = 40.4, SE = 7.6, $t$ = 5.3, $p$ < 0.0001) and incongruent conditions (β = 49.5, SE = 8.4, $t$ = 5.9, $p$ < 0.0001). The effect of SOA was not significant (β = 24.1, SE = 15.5, $t$ = 1.6, $p$ = 0.13). There was no significant interaction between SOA and Gender congruency (*ps* > 0.1). The data points from the baseline condition were removed from further analysis. The model comparing the congruent and incongruent conditions at each SOA revealed no effect of SOA (β = 30.5, SE = 17.5, $t$ = 1.7, $p$ < 0.09), and no effect of Gender congruency (β = 6.8, SE = 7.3, $t$ = 0.94, $p$ = 0.35, model without a by Item random slope for the effect of Gender congruency). The model with the interaction revealed no such interaction ($p$ = 0.65, model without a by Item random slope for the effect of Gender congruency).

The results of the corresponding repeated measures Bayesian ANOVA are presented in Table 7. The null model included SOA and Participant, and is the model that receives the most support (**$BF_M$** = 8.06). The data are about five times less likely under a model with Gender congruency than under models without it, and they are about 3.7 times less likely under the model with the interaction between Gender congruency and SOA than under models without this interaction. This analysis thus

favors the hypothesis that the naming latencies are not influenced by Gender congruency and that Gender congruency does not interact with SOA.

-Insert Table 7 about here-

*Interaction between processing times for distractor and gender congruency effect*

- Insert Figure 7 about here -

The first statistical model included as fixed effects SOA, Gender congruency, a linear and a quadratic term for the participants' mean reading latencies as well as the interactions between these terms and the two factors (the model did not converge with a by Item random slope for the linear term). The full model revealed no three-way interaction with mean reading latencies ($p$ = 0.17), or with the quadratic term ($p$ = 0.65). These were removed. The model with two-way interactions revealed no such interactions (all $p$s > 0.16, model with no by Item random slopes for Gender congruency and the linear term). In the model without interactions (model with no by Item random slopes for Gender congruency and the linear term) there was no effect of Gender congruency (β = 7.7, SE = 7.5, $t$ = 1.04, $p$ = 0.31) or SOA (β = 29.4, SE = 17.4, $t$ = 1.7, $p$ = 0.099). There was an effect of both the linear and quadratic term for the mean reading latencies, with increasing naming latencies for participants with higher reading latencies (β = 1.09, SE = 0.36, $t$ = 3.1, $p$ = 0.0047; β = 15.1, SE = 7.0, $t$ = 2.2, $p$ = 0.040).

The second statistical model included as fixed effects SOA, Gender congruency, a linear and quadratic term for the difference measure, as well as the interactions between these terms and the two factors. The full model revealed no three-way interaction with the difference measure ($p$ = 0.58), or with the quadratic term ($p$ = 0.15, model with no by Participant or by Item random slope for the effect of SOA). These were removed. The model with two-way interactions revealed no interaction between the difference measure and Gender congruency ($p$ = 0.0504) or Gender congruency and the quadratic term ($p$ = 0.40, model with no random slopes for the effect of SOA). In the model without interactions, there were main effects of SOA, with shorter naming latencies at 200 SOA (β = 180, SE = 2.3, $t$ = 77.4, $p$ < 0.0001), of the linear and quadratic term for the difference measure (with increasing

naming latencies for greater difference measures, β = 0.93, SE = 0.0057, *t* = 161.2, *p* < 0.0001 and β = 0.90, SE = 0.42, *t* = 2.1, *p* = 0.035). In this model, the effect of Gender congruency was significant and in the expected direction (β = 10.9, SE = 5.1, *t* = 2.14, *p* = 0.037).

***Discussion***

The results of Experiment 4 are almost identical to those of Experiment 3, and suggest once again that there is no effect of Gender congruency at 0 or 200 SOA when reading times are not considered. Taking into account the variance accounted by differences in processing times between picture and distractor in the model results in significant effects of SOA and Gender congruency.

**General discussion**

In the present study, we set out to replicate the gender congruency effect reported in two previous studies (Foucart et al., 2010; Schriefers & Teruel, 1999) with the French definite determiners *le-la*, to resolve discrepancies in the interaction of the effect with the timing of distractor presentation (SOA) and the onset of the following word, and to extend the investigation to the indefinite determiners *un-une*. The outcome of four picture-word interference experiments can be summarized as follows. Gender congruency effects in French are not as robust as previously argued but may surface under certain conditions. These conditions seem to be related to the participants' processing times. Importantly, when gender congruency effects are observed, they are not restricted to late SOAs and there is no evidence that they are modulated by the onset of the next word. Gender congruency effects in French can be found for the definite determiners *le-la*. There is some evidence that they can also be found for other determiners (i.e., indefinite determiners *un-une*) but this latter conclusion awaits further empirical confirmation.

In the remainder of this General discussion, we discuss these results in three steps. We first discuss the lack of robustness of the gender congruency effect in French as well as existing inconsistencies across studies and experiments. We then discuss the implications of the present study's findings for

current accounts of determiner selection across languages. Finally, we discuss the benefits (and limitations) of replications and Bayes factors.

*Inconsistency of gender congruency effects in French*

At least three published studies examined gender congruency effects for the French definite determiners *le-la*. Schriefers and Teruel (1999) reported effects at 300 SOA for the reduced form *l'* of the determiner, and at 0 SOA for the full forms *le-la*. Alario & Caramazza (2002) reported null effects at 0 SOA for the latter. Foucart et al. (2010) reported gender congruency effects at 200 SOA in three separate studies testing only the full forms *le-la*, and no effect at 0 SOA in one study. Notably, none of these studies provided an explicit statistical test of whether the gender congruency effect in French depends on the moment in time at which the distractor is presented (SOA; i.e. a significant interaction). In sum, existing studies suggest that gender congruency effects in French show inconsistencies and provide no robust evidence of whether and how they interact with the SOA. The present study's findings offer a similar picture. A gender congruency effect was found in three out of six experiments. The picture is more homogenous regarding the interaction between Gender congruency and SOA. None of the experiments in which the interaction between SOA and Gender congruency was tested favors the hypothesis that the two interact.

At least two hypotheses can be considered to explain the instability of the gender congruency effect in French. According to a first hypothesis, and as postulated for other Romance languages, there is no such effect in French, and the positive evidence reported in the literature results from type I errors. In order to gain additional information on this issue, we performed a random effects meta-analysis. This analysis allows deriving an estimate of the effect, taking all the available evidence into account (Alario & Caramazza, 2002; Foucart, 2010; and Schriefers & Teruel, 1999; data of the present study). SOA was treated as a predictor in the analysis. The results of this analysis are plotted in Figure 8. This analysis targets two issues. First, the model tests whether overall, the available evidence provides evidence in favor of a gender congruency effect in French and provides an estimate of this effect.

Under the assumptions that (1) estimations of missing information are correct[5] and that (2) the studies included in the meta-analysis involve all the studies ever performed on the gender congruency effect in French, this meta-analysis suggests that overall, the evidence in favor of gender congruency effects in this language is positive. Secondly, the model provides information on whether the evidence differs depending on the SOA. The effect of SOA is not significant ($p$ = 0.15, value adjusted following Knapp and Hartung, 2003's method). Note however that this latter result must be taken with care and the analysis be considered exploratory, especially because some SOAs are only tested in one study. This meta-analysis supports the hypothesis that gender congruent trials tend to be named more quickly than gender incongruent trials in French. The amount of residual heterogenity possibly indicates that variables not included in the analysis further influence the gender congruency effect in French.

-Insert Figure 8 about here-

[5] The meta-analysis requires information on the mean difference between conditions (i.e., gender congruent and gender incongruent), the variance and number of data points in each condition. This information was not available for all experiments. For the Experiments in Foucart et al. (2010) we had to estimate the number of data points in each condition. Foucart et al. report a global error rate of 1.2% for their first experiment. We therefore removed 1.2% of the original number of trials in each condition for each of their experiments. In Schriefers and Teruel (1999), no information is provided on the variance or the number of errors per condition. We used the mean variance of all other studies included in the meta-analysis as an estimation of the variance and the original number of trials in each condition in the study as an estimation of the number of data points.

According to an alternative hypothesis, whether gender congruency effects surface depends on how the processing streams for the target picture and the distractor word interact in time. The corollary is that the gender congruency effect depends on participants processing times across the materials. More specifically, the effect would surface when the activation of alternative determiner forms resulting from the processing of the distractor occurs in a time window where this information is relevant for the encoding of the to-be-produced determiner. This may, or may not, be reflected in an interaction between Gender congruency and SOA. Whether Gender Congruency interacts with SOA under this hypothesis depends on participants processing times in naming the pictures and reading the distractors. Participants vary in their processing times for both the encoding of the target word and the processing of the distractor. For instance, in Experiment 4, the mean naming times ranged between 584 ms and 1162 ms at an SOA of 0 across participants, and between 479 ms and 1050 ms at an SOA of 200. The mean reading times ranged from 337 ms to 538 ms. If one computes the mean difference between the mean naming and mean reading times (averaging across SOAs), our participants ranged between 14 ms and 407 ms. As can be seen, fixed differences of 200 SOA are not sufficient to capture this variability. Under the hypothesis that the gender congruency effect depends on participants' processing times for both the target word and distractor, we would expect an interaction between the difference measure and Gender Congruency in Experiments 3 and 4. No such interaction was found. However, simply accounting in the model for the variance in the difference measure resulted in significant gender congruency effects. This shows that gender congruency effects may be hidden by an important variability in processing time differences between picture naming and distractor processing. It is possible that with more participants, the interaction would be significant.

In the next section, we assume that gender congruency effects surface in French for the definite and indefinite determiners and discuss possible accounts of these effects.

*Possible accounts of determiner selection in French (and other Romance languages)*

Previous studies on determiner selection revealed systematic gender congruency effects in Germanic languages, but not in Romance languages. The *late selection hypothesis* was designed to account for this discrepancy and assumes that in Romance languages, determiners are selected later. The data of the present study, together with that of two other studies in French, challenge this hypothesis by showing that gender congruency effects can surface in French. How can this effect be explained and reconciled with the absence of gender congruency effects in other Romance languages? We consider two accounts.

According to a first account, gender congruency effects are restricted to the definite (and possibly the indefinite) French determiners because the time course of determiner selection for these determiners differs from that of other determiners in French and other Romance languages. This first account follows Foucart et al.'s proposal that some determiners require an adjustment mechanism, but departs from their proposal in that it does not involve a checking mechanism. The French definite and indefinite determiners can easily be accounted for by a general phonological rule, and many authors in the field of phonology provide rule-based accounts of vowel/schwa deletion or the liaison process. In Foucart et al.'s proposal (unlike in classical phonological accounts) the late adjustment comes with a checking mechanism. This checking mechanism is postulated to explain why a gender congruency effect was only found at later SOAs. Having provided some evidence that the gender congruency effect does not interact with SOA, we consider that this mechanism is not necessary. As a consequence, the definite and indefinite determiners can be selected early, as soon as the gender of the noun is available. A gender congruency effect is predicted for these determiners at SOAs similar to the SOAs at which such effects occur in Germanic languages (provided that the time window of distractor processing overlaps with the time window of target word encoding). This account does not follow the *maximum consistency principle* (Caramazza et al., 2001), which would assume a late selection for all French determiners. A further test of this account could be performed by comparing the definite and indefinite determiners to other French determiners, whose pronunciation is also constrained by the gender of the noun and the phonological properties of the following word, but for

which the alternation between the pronunciations to be used before consonants vs. vowels can less easily be described as a general rule or adjustment process (see again Alario & Caramazza, 2002). No gender congruency effects are therefore predicted for these determiners. For instance, before consonant-initial words, the masculine possessive singular determiners are realised and spelt as *mon* [mɔ̃], *son* [sɔ̃], and *ton* [tɔ̃] (e.g., *mon chat [mɔ̃ʃa]* 'my cat') and the feminine determiners as *ma*, *ta*, and *sa* (e.g., *ma fille [mafij]* 'my daughter'). Before vowel-initial words, the two genders share a single spelling (*mon*, *ton* and *son*), and pronunciation ([tɔ̃n], [mɔ̃n], and [sɔ̃n] with possible desanalisation of the vowel, e.g., *mon âne [mɔ̃nan]* 'my donkey'). Importantly, the feminine determiners are the only words in the French language whose ending alternates between [a] and [ɔ̃n]. Similarly, the masculine demonstrative singular determiner is realized (and spelt) *ce* before consonant-initial words (e.g., *ce chat* 'this cat') and *cet* before vowel-initial words (e.g., *cet âne* 'this donkey'), while with feminine nouns, the determiner is *cette* irrespective of the phonological onset of the next word (e.g., *cette chaise* 'this chair', *cette armoire* 'this cupboard') note that *cet* and *cette* have the same pronunciation, [sEt]). No other word in the French language ends in "e" or "et" depending on the phonological context, consequently, and as for the possessive determiners, this alternation can hardly be described with a general phonological rule. The production of these determiners, unlike that of *le-la* and *un-une*, would therefore not involve an adjustment process.

According to an alternative account, the representation and selection of French definite and indefinite determiners is similar to that of other determiners in French and other Romance languages. These determiners have a phonological representation for each pronunciation, and in cases where the selection requires the phonological context, the selection is delayed. Note that the hypothesis that forms such as l' and *un* with liaison each have a separate phonological representation has also been discussed in the phonological literature (e.g., Klausenburger 1984). Under this second account, gender congruency effects can surface in Romance languages for all determiners, provided that the information activated by the processing of the distractor is active when the to-be-produced

determiner is selected. This account is not in line with the existing data on Romance languages other than French, unless we assume that when determiners are selected later, gender congruency effects are harder to observe, possibly because the number of trials for which the determiners activated by the distractor can influence the selection of the to-be-produced determiner is smaller. An important aim for further studies should be to examine whether gender congruency effects surface in these languages when processing times for the distractor and target words are taken into account. Note that the *late selection hypothesis* predicts gender congruency effects in these languages if the information by the distractor is active at the time of determiner selection. So far however, no such effects have been found. On a related note, accounting for variability in processing times for target and distractor words in picture-word interference experiments may help resolve discrepancies in other studies (see for instance Cubelli, Lotto, Paolieri, Girelli, & Job, 2005; Paolieri et al., 2010; versus Finocchiaro et al., 2011 on gender congruency effects in bare noun naming).

*Note on replications, report of null effects, Bayes Factors, and meta-analyses as complementary tools*

As experimental psychologists, we rely on experimental effects to model the language processing system. In order for this process to be efficient, it is necessary to assess the robustness of experimental effects, as well as the conditions under which these effects surface or do not surface. Current practice in experimental psychology does not provide this information. Only significant effects usually make their way into journals. Moreover, while most researchers (e.g., Martin & Clarke, 2017) agree that replications are part of this process (as lack of replications are, recall that by setting alpha to 0.05, we agree to reject the null in 5% of the cases when this null is in fact correct) they are not always given the importance they deserve. Novel, surprising effects are much trendier, and make their way much more easily in top journals. As a consequence, the literature does not always provide information on the robustness of effects or on their boundary conditions. There is no unique remedy to this state of affair. The present study highlights the potential complementarity of Bayes factors, report of null effects, replications, and meta-analyses. It is worth noting here that Bayes Factors offer useful complimentary information, but also have limitations. They can favour the

null hypothesis because the data are too noisy, as evidenced in the present study, or because the alternative hypothesis is ill-selected.

**Conclusion**

The present study focussed on the representation and selection of determiners, a crucial issue to understand language production beyond isolated words. The data of this study, taken together with previous findings, point to positive evidence for a gender congruency effect in French. In addition, they reveal that the effect can be difficult to observe. We introduced and tested a plausible explanation, which relies on the alignment in processing times for the distractor and target word. We showed that effects of gender congruency surface when such differences are taken into account. This suggests that the variability in processing times for the target word and distractor should be taken into account in studies using the picture word interference paradigm to test hypotheses about the time course of encoding processes. Perhaps the most important conclusion of the present study is that the time course of determiner selection across languages is far from settled. The divide between Germanic and Romance languages is challenged and further cross-linguistic research is needed to understand how language-specific and universal constraints (e.g., maximum consistency) shape the time course of determiner selection.

**Acknowledgements**

We acknowledge funding from the University of Geneva, from the European Research Council under the European Community's Seventh Framework Program (FP7/2007-2013 Grant agreement no. 263575). This research was also partly supported by the *grants ANR-16-CONV-0002 (ILCB), ANR-11-LABX-0036 (BLRI) and the Excellence Initiative of Aix-Marseille University (A*MIDEX),* as well as the Deutsche Forschungsgemeinschaft (DFG), Collaborative Research Centre SFB 1287, Project B05. We thank Irene Rotondi for help in collecting the data for Experiment 1, Cassandre Mabut and Capucine Du peloux for help in collecting and pre-processsing the vocal responses for Experiment 2.

**References**


Alario, F.-X., & Caramazza, A. (2002). The production of determiners: Evidence from French. *Cognition*, 82(3), 179–223. https://doi.org/10.1016/S0010-0277(01)00158-5

Anderson, S. (1982). The analysis of French schwa: Or, how to get something for nothing. *Language*, 58, 534-573.

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

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for Participants and items. *Journal of Memory and Language*, 59(4), 390–412. https://doi.org/10.1016/j.jml.2007.12.005

Brown-Schmidt, S., & Konopka, A. E. (2008). Little houses and casas pequeñas: Message formulation and syntactic form in unscripted speech with speakers of English and Spanish. *Cognition*, 109(2), 274–280. https://doi.org/10.1016/j.cognition.2008.07.011

Bürki, A. (2017). Differences in processing times for distractors and pictures modulate the influence of distractors in picture-word interference tasks. *Language, Cognition, & Neuroscience*, 32, 709-723. http://dx.doi.org/10.1080/23273798.2016.1267783

Bürki, A., Frauenfelder, U. H., & Alario, F.-X. (2015). On the resolution of phonological constraints in spoken production: Acoustic and response time evidence. *The Journal of the Acoustical Society of America*, 138(4), EL429-434. https://doi.org/10.1121/1.4934179

Bürki, A., Sadat, J., Dubarry, A.-S., & Alario, F.-X. (2016). Sequential processing during noun phrase production. *Cognition*, 146, 90–99. https://doi.org/10.1016/j.cognition.2015.09.002

Bybee, J. (2001). Frequency effects on French liaison. In J. Bybee and P. Hopper (Eds.), *Frequency and the Emergence of Linguistic Structure*, John Benjamins, Amsterdam (2001), pp. 337-359.

Caramazza, A., Miozzo, M., Costa, A., Schiller, N., & Alario, F.-X. (2001). A crosslinguistic investigation of determiner production. *Language, Brain, and Cognitive Development: Essays in honor of Jacques Mehler*, 209–226. Cambridge, MA : MIT Press.

Costa, A., Alario, F. X., & Sebastián-Gallés, N. (2007). Cross-linguistic research on language production. In M. G. Gaskell (Ed.), *The Oxford Handbook of Psycholinguistics*, Oxford: Oxford University Press (pp. 531–546).

Costa, A. Sebastian-Galles, N., Miozzo, M. & Caramazza, A. (1999). The gender congruity effect: Evidence from Spanish and Catalan. *Language and Cognitive Processes*, 14, 381–391. http://dx.doi.org/10.1080/016909699386275

Côté, M.-H. (2005). Le statut lexical des consonnes de liaison. *Langages*, 158. 66-78.

Côté, M., & Morrison, G. S. (2007). The nature of the schwa/zero alternation in French clitics: Experimental and non-experimental evidence. *Journal of French Language Studies*, 17, 159-186. http://dx.doi.org/10.1017/S0959269507002827

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(1), 42–59. https://doi.org/10.1016/j.jml.2005.02.007

Cutler, A., Mehler, J., Norris, D., & Segui, J. (1983). A language-specific comprehension strategy. *Nature*, *304*, 159-160. http://dx.doi.org/10.1038/304159a0

Cutler, A., Sebastián-Gallés, N., Soler-Vilageliu, O., & van Ooijen, B. (2000). Constraints of vowels and consonants on lexical selection: Cross-linguistic comparisons. *Memory & Cognition*, 28(5), 746–755. http://dx.doi.org/10.3758/BF03198409

Dell, F. (1985). *Les règles et les sons*. Paris: Hermann.

DerSimonian, R., & Laird, N. (1986). Meta-analysis in clinical trials. Controlled Clinical Trials, 7, 177–188.

Dumay, N., & Damian, M. F. (2011). A word-order constraint in single-word production?: Failure to replicate Janssen, Alario, and Caramazza (2008). *Psychological Science*, 22(4), 559–561. https://doi.org/10.1177/0956797611401754

Encrevé, P. (1988). *La liaison avec et sans enchaînement. Phonologie tridimensionnelle et usages du français.* Paris: Seuil

Finocchiaro, G., Alario, F. X., Schiller, N. O., Costa, A., Miozzo, M., Caramazza, A., (2011). Gender congruency goes Europe: A cross-linguistic study of the gender congruency effect in Romance and Germanic languages*. Italian Journal of Linguistics / Rivista di Linguistica,* 23, 161-198.

Forster, K. I. & Forster, J. C. (2003). DMDX: A windows display program with millisecond accuracy. *Behavior Research Methods, Instruments, & Computers*, 35, 116-124.

Foucart, A., Branigan, H. P., & Bard, E. G. (2010). Determiner selection in Romance languages: Evidence from French. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, 36(6), 1414–1421. https://doi.org/10.1037/a0020432

Ganushchak, L. Y., Konopka, A. E., & Chen, Y. (2014). What the eyes say about planning of focused referents during sentence formulation: A cross-linguistic investigation. *Frontiers in Psychology*, 5. https://doi.org/10.3389/fpsyg.2014.01124

Garrett, M. F. (1980). Levels of processing in sentence production. In B. Butterworth (Ed*.), Language production* (Vol. 1, pp. 177-220). London: Academic Press.

Garrett, M. F. (1984). The organization of processing structure for language production: Applications to aphasic speech. In D. Caplan, A. R. Lecours, & A. Smith (Eds*.), Biological Perspectives on Language* (pp. 172-193). Cambridge, MA: MIT Press.

Gola-Asmussen, C., Lequette, C., Pouget, G., Rouyer, C., & Zorman, M. (2010). Outil d'évaluation de compétences de lecture chez l'adulte de plus de 16 ans. Université de Provence Aix-Marseille 1 et Cognisciences LES Université Pierre Mendès.

Goldstein, H. (1987*). Multilevel models in educational and social research*. London: Griffin

Hawkins, S. (2007). Processing typology and why psychologists need to know about it. *New ideas in Psychology*, 25, 87-107. https://doi.org/10.1016/j.newideapsych.2007.02.003

Knapp, G., & Hartung, J. (2003). Knapp, G., & Hartung, J. (2003). Improved tests for a random effects meta-regression with a single covariate. *Statistics in Medicine*, *22*(17), 2693–2710. https://doi.org/10.1002/sim.1482

La Heij, W., Mak, P., Sander, J., & Willeboordse, E. (1998). The gender-congruency effect in picture-word tasks. *Psychological Research*, 61(3), 209–219. https://doi.org/10.1007/s004260050026

Heim, S., Friederici, A.D, Schiller, N. O., Rüschemeyer, S. A., & Amunts, K. (2009). The determiner congruency effect in language production investigated with functional MRI. *Human Brain Mapping*, 30, 928-940. https://doi.org 10.1002/hbm.20556.

Jaeger, T. F., & Norcliffe, E. J. (2009). The cross-linguistic study of sentence production. *Language and Linguistics Compass,* 3(4), 866–887. https://doi.org/10.1111/j.1749-818X.2009.00147.x

Janssen, N., Schiller, N. O., & Alario, F.-X. (2014). The selection of closed-class elements during language production: A reassessment of the evidence and a new look on new data. *Language, Cognition and Neuroscience*, *29*(6), 695–708DOI: 10.1080/01690965.2012.693617

Jarosz, A., & Wiley, J. (2014). What are the Odds? A practical guide to computing and reporting Bayes Factors. *The Journal of Problem Solving*, *7*(1). https://doi.org/10.7771/1932-6246.1167

Jeffreys, H. (1961*). Theory of probability* (3rd edition). Oxford: Oxford University Press.

Klausenburger, J. (1984). *French liaison and linguistic theory*. Wiesbaden: F. Steiner.

Kuznetsova, A. (2014). Package « lmerTest »: Tests for random and fixed effects for linear mixed effect models (lmer objects of lme4 package).

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

Levelt, W. J. M., Roelofs, A., & Meyer, A. S. (1999). A theory of lexical access in speech production. *The Behavioral and Brain Sciences*, 22, 1-38. DOi: 10.1017/S0140525X99001776

Love, J., Selker, R., Verhagen, J., Marsman, M., Gronau, Q. F., Jamil, T., Smira, M., Epskamp, S., Wild, A., Morey, R., Rouder, J. & Wagenmakers, E. J. (2015). JASP (Version 0.6)[Computer software].

Martin, G. N., & Clarke, R. M. (2017). Are psychology journals anti-replication? A snapshot of editorial practices. *Frontiers in Psychology*, 8. https://doi.org/10.3389/fpsyg.2017.00523

Matos, R., Ferrand, L., Pallier, C., & New, B. (2001). Une base de données lexicales du français contemporain sur internet : LEXIQUE. *L'année psychologique*, 101(3), 447–462. https://doi.org/10.3406/psy.2001.1341

Mehler J, Dupoux E, Pallier C, & Dehaene-Lambertz G (1994). Cross-linguistic approaches to speech processing. *Current Opinion in Neurobiology*. 4: 171-6. https://doi.org/10.1016/0959-4388(94)90068-X

Meyer, A. S. (1990). The time course of phonological encoding in language production: The encoding of successive syllables of a word. *Journal of Memory and Language*, *29*(5), 524–545. https://doi.org/10.1016/0749-596X(90)90050-A

Meyer, A. S. (1991). The time course of phonological encoding in language production: Phonological encoding inside a syllable. *Journal of Memory and Language*, *30*(1), 69–89. https://doi.org/10.1016/0749-596X(91)90011-8

Miozzo, M., & Caramazza, A. (1999). The selection of determiners in noun phrase production. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 25(4), 907–922. https://doi.org/10.1037/0278-7393.25.4.907

Miozzo, M., & Caramazza, A. (2003). When more is less: A counterintuitive effect of distractor frequency in the picture-word interference paradigm. *Journal of Experimental Psychology: General*, 132(2), 228–252. https://doi.org/10.1037/0096-3445.132.2.228

Miozzo, M., Costa, A., & Caramazza, A. (2002). The absence of a gender congruency effect in Romance languages: A matter of stimulus onset asynchrony? *Journal of Experimental Psychology. Learning, Memory, and Cognition*, 28(2), 388–391.

Morey, R. D. (2008). Confidence intervals from normalized data: A correction to Cousineau (2005). *Tutorials in Quantitative Methods for Psychology*, 4(2), 61–64.

Norcliffe, E., Konopka, A. E., Brown, P., & Levinson, S. C. (2015). Word order affects the time course of sentence formulation in Tzeltal. *Language, Cognition and Neuroscience*, 30(9), 1187–1208. https://doi.org/10.1080/23273798.2015.1006238

Paolieri, D., Lotto, L., Morales, L., Bajo, T., Cubelli, R., & Job, R. (2010). Grammatical gender processing in Romance languages: Evidence from bare noun production in Italian and Spanish. *European Journal of Cognitive Psychology*, 22(3), 335–347. https://doi.org/10.1080/09541440902916803

R Development Core Team (2017). R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from http://www.R-project.org

Rouder, J. N., Morey, R. D., Speckman, P. L., & Province, J. M. (2012). Default Bayes factors for ANOVA designs. *Journal of Mathematical Psychology*, 56(5), 356–374. https://doi.org/10.1016/j.jmp.2012.08.001

Rouder, J. N., Speckman, P. L., Sun, D., Morey, R. D., & Iverson, G. (2009). Bayesian t tests for accepting and rejecting the null hypothesis. *Psychonomic bulletin & review*, 16(2), 225–237.

Scaltritti, M., Navarrete, E., & Peressotti, F. (2015). Distributional analyses in the picture–word interference paradigm: Exploring the semantic interference and the distractor frequency effects. *The Quarterly Journal of Experimental Psychology*, 68(7), 1348–1369. https://doi.org/10.1080/17470218.2014.981196

Schiller, N. O., & Caramazza, A. (2003). Grammatical feature selection in noun phrase production: Evidence from German and Dutch. *Journal of Memory and Language*, 48(1), 169–194. https://doi.org/10.1016/S0749-596X(02)00508-9

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

Schriefers, H., & Teruel, E. (2000). Grammatical gender in noun phrase production: The gender interference effect in German. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, 26(6), 1368–1377.

Schriefers, H., and Teruel, E. (1999). The production of noun phrases: A cross-linguistic comparison of French and German. *Proceedings of the Twenty-First Annual Conference of the Cognitive Science Society*, 637–642, Mahwah, NJ: Lawrence Erlbaum.

Schwarzer, G. (2007). meta: An R Package for Meta-Analysis

Selkirk, E. (1978). Comments on Morin's paper: The French foot: On the status of "mute e". *Studies in French Linguistics*, 1, 24-28.

Tranel, B. (1981). *Concreteness in Generative Phonology: Evidence from French*. University of California Press.

Tranel, B. (2000). Aspects de la phonologie du français et la théorie de l'optimalité. *Langue française*, 126, 39-72.

van Berkum, J. J. A. (1997). Syntactic processes in speech production: The retrieval of grammatical gender. *Cognition*, 64(2), 115–152. https://doi.org/10.1016/S0010-0277(97)00026-7

Vasishth, S., & Nicenboim, B. (2016). Statistical Methods for Linguistic Research: Foundational Ideas – Part I. *Language and Linguistics Compass*, *10*(8), 349–369. https://doi.org/10.1111/lnc3.12201

Wagenmakers, E.-J., Love, J., Marsman, M., Jamil, T., Ly, A., Verhagen, J., … Morey, R. D. (2017). Bayesian inference for psychology. Part II: Example applications with JASP. *Psychonomic Bulletin & Review*, 1–19. https://doi.org/10.3758/s13423-017-1323-7

## Tables

**Table 1.** Bayesian repeated measures ANOVA, Experiment 1a

**A.Model Comparison**

| Models | P(M) | P(M\|data) | $BF_M$ | $BF_{01}$ | error % |
|---|---|---|---|---|---|
| Null model (incl. Noun onset, Participant) | 0.333 | 0.644 | 3.620 | 1.000 | |
| Gender Congruency | 0.333 | 0.203 | 0.508 | 3.178 | 0.840 |
| Gender Congruency + Gender Congruency ✻ Noun onset | 0.333 | 0.153 | 0.362 | 4.206 | 8.650 |

*Note.* All models include Noun onset, Participant. **P(M):** distribution of prior model probabilities across all the models tested. P**(M|data):** posterior model probabilities. $BF_{01}$: Bayes factor that compares one model against the null model. $BF_M$: Bayes factor that compares one model against all others (i.e., difference between prior model odds and posterior model odds). For details, see Wagenmakers, 2017.

**B.Analysis of Effects**

| Effects | P(incl) | P(incl\|data) | $BF_{Inclusion}$ |
|---|---|---|---|
| Gender Congruency | 0.333 | 0.203 | 0.315 |
| Gender Congruency ✻ Noun onset | 0.333 | 0.153 | 0.756 |

*Note.* Compares models that contain the effect to equivalent models stripped of the effect. Higher-order interactions are excluded. Analysis suggested by Sebastiaan Mathôt.

**Table 2.** Bayesian repeated measures ANOVA, Experiment 1b

**A.Model Comparison**

| Models | P(M) | P(M\|data) | $BF_M$ | $BF_{01}$ | error % |
|---|---|---|---|---|---|
| Null model (incl. Gender, Participant) | 0.333 | 0.409 | 1.386 | 1.000 | |
| Gender congruency | 0.333 | 0.455 | 1.672 | 0.899 | 1.125 |
| Gender congruency + Gender congruency ✻ Gender | 0.333 | 0.135 | 0.313 | 3.023 | 0.963 |

*Note.* All models include Gender, Participant. **P(M):** distribution of prior model probabilities across all the models tested. P**(M|data):** posterior model probabilities. $BF_{01}$: Bayes factor that compares one model against the null model. $BF_M$: Bayes factor that compares one model against all others (i.e., difference between prior model odds and posterior model odds). For details, see Wagenmakers, 2017.

**B.Analysis of Effects**

| Effects | P(incl) | P(incl\|data) | $BF_{Inclusion}$ |
|---|---|---|---|
| Gender congruency | 0.333 | 0.455 | 1.113 |
| Gender congruency ✻ Gender | 0.333 | 0.135 | 0.297 |

*Note.* Compares models that contain the effect to equivalent models stripped of the effect. Higher-order interactions are excluded. Analysis suggested by Sebastiaan Mathôt.

**Table 3.** Bayesian repeated measures ANOVA, joint analysis of Experiments 1a and 1b

**A.Model Comparison**

| Models | P(M) | P(M\|data) | $BF_M$ | $BF_{01}$ | error % |
|---|---|---|---|---|---|
| Null model (incl. Determiner, Gender, Determiner ✻ Gender, Participant) | 0.167 | 0.490 | 4.810 | 1.000 | |
| Gender congruency | 0.167 | 0.331 | 2.478 | 1.480 | 3.994 |
| Gender congruency + Gender congruency ✻ Determiner | 0.167 | 0.089 | 0.486 | 5.531 | 3.140 |
| Gender congruency + Gender congruency ✻ Gender | 0.167 | 0.066 | 0.354 | 7.410 | 2.999 |
| Gender congruency + Gender congruency ✻ Determiner + Gender congruency ✻ Gender | 0.167 | 0.018 | 0.094 | 26.601 | 3.609 |
| Gender congruency + Gender congruency ✻ Determiner + Gender congruency ✻ Gender + Gender congruency ✻ Determiner ✻ Gender | 0.167 | 0.005 | 0.026 | 96.211 | 3.888 |

*Note.* All models include Determiner, Gender, Determiner ✻ Gender, Participant. **P(M):** distribution of prior model probabilities across all the models tested. P**(M|data):** posterior model probabilities. $BF_{01}$: Bayes factor that compares one model against the null model. $BF_M$: Bayes factor that compares one model against all others (i.e., difference between prior model odds and posterior model odds). For details, see Wagenmakers, 2017.

**B.Analysis of Effects**

| Effects | P(incl) | P(incl\|data) | $BF_{Inclusion}$ |
|---|---|---|---|
| Gender congruency | 0.167 | 0.331 | 0.676 |
| Determiner ✻ Gender congruency | 0.333 | 0.107 | 0.269 |
| Gender ✻ Gender congruency | 0.333 | 0.085 | 0.201 |
| Determiner ✻ Gender ✻ Gender congruency | 0.167 | 0.005 | 0.276 |

*Note.* Compares models that contain the effect to equivalent models stripped of the effect. Higher-order interactions are excluded. Analysis suggested by Sebastiaan Mathôt.

**Table 4.** Bayesian repeated measures ANOVA, Experiment 2a

**A.Model Comparison**

| Models | P(M) | P(M\|data) | $BF_M$ | $BF_{01}$ | error % |
|---|---|---|---|---|---|
| Null model (incl. SOA, Participant) | 0.333 | 0.752 | 6.062 | 1.000 | |
| Gender Congruency | 0.333 | 0.188 | 0.463 | 4.003 | 2.001 |
| Gender Congruency + Gender Congruency ✻ SOA | 0.333 | 0.060 | 0.128 | 12.483 | 1.194 |

*Note.* All models include SOA, Participant. **P(M):** distribution of prior model probabilities across all the models tested. P**(M|data):** posterior model probabilities. $BF_{01}$: Bayes factor that compares one model against the null model. $BF_M$: Bayes factor that compares one model against all others (i.e., difference between prior model odds and posterior model odds). For details, see Wagenmakers, 2017.

**B.Analysis of Effects**

| Effects | P(incl) | P(incl\|data) | $BF_{Inclusion}$ |
|---|---|---|---|
| Gender Congruency | 0.333 | 0.188 | 0.250 |
| SOA ✻ Gender Congruency | 0.333 | 0.060 | 0.321 |

*Note.* Compares models that contain the effect to equivalent models stripped of the effect. Higher-order interactions are excluded. Analysis suggested by Sebastiaan Mathôt.

**Table 5.** Bayesian repeated measures ANOVA, Experiment 2b

**A.Model Comparison**

| Models | P(M) | P(M\|data) | $BF_M$ | $BF_{01}$ | error % |
|---|---|---|---|---|---|
| Null model (incl. Gender, SOA, Participant) | 0.091 | 0.633 | 17.255 | 1.000 | |
| Gender Congruency | 0.091 | 0.113 | 1.279 | 5.582 | 5.115 |
| Gender Congruency + Gender Congruency ✻ Gender | 0.091 | 0.032 | 0.332 | 19.723 | 7.225 |
| Gender Congruency + Gender Congruency ✻ SOA | 0.091 | 0.027 | 0.274 | 23.728 | 3.476 |
| Gender Congruency + Gender Congruency ✻ Gender + Gender Congruency ✻ SOA | 0.091 | 0.007 | 0.074 | 86.713 | 6.963 |
| Gender ✻ SOA | 0.091 | 0.144 | 1.687 | 4.387 | 2.430 |
| Gender Congruency + Gender ✻ SOA | 0.091 | 0.027 | 0.278 | 23.435 | 7.809 |
| Gender Congruency + Gender Congruency ✻ Gender + Gender ✻ SOA | 0.091 | 0.008 | 0.077 | 82.927 | 4.236 |
| Gender Congruency + Gender Congruency ✻ SOA + Gender ✻ SOA | 0.091 | 0.006 | 0.059 | 108.519 | 4.181 |
| Gender Congruency + Gender Congruency ✻ Gender + Gender Congruency ✻ SOA + Gender ✻ SOA | 0.091 | 0.002 | 0.017 | 378.094 | 4.882 |
| Gender Congruency + Gender Congruency ✻ Gender + Gender Congruency ✻ SOA + Gender ✻ SOA + Gender Congruency ✻ Gender ✻ SOA | 0.091 | 9.449e -4 | 0.009 | 670.029 | 4.132 |

*Note.* All models include Gender, SOA, Participant. **P(M):** distribution of prior model probabilities across all the models tested. P**(M|data):** posterior model probabilities. **$BF_{01}$**: Bayes factor that compares one model against the null model. **$BF_M$** : Bayes factor that compares one model against all others (i.e., difference between prior model odds and posterior model odds). For details, see Wagenmakers, 2017.

**B.Analysis of Effects**

| Effects | P(incl) | P(incl\|data) | $BF_{Inclusion}$ |
|---|---|---|---|
| Gender Congruency | 0.182 | 0.140 | 0.181 |
| Gender ✻ SOA | 0.455 | 0.186 | 0.229 |
| Gender ✻ Gender Congruency | 0.364 | 0.049 | 0.282 |
| SOA ✻ Gender Congruency | 0.364 | 0.041 | 0.230 |
| Gender ✻ SOA ✻ Gender Congruency | 0.091 | 9.449e -4 | 0.564 |

*Note.* Compares models that contain the effect to equivalent models stripped of the effect. Higher-order interactions are excluded. Analysis suggested by Sebastiaan Mathôt.

**Table 6**. Bayesian repeated measures ANOVA, Experiment 3

**A.Model Comparison**

| Models | P(M) | P(M\|data) | $BF_M$ | $BF_{01}$ | error % |
|---|---|---|---|---|---|
| Null model (incl. SOA, Participant) | 0.333 | 0.754 | 6.143 | 1.000 | |
| Gender Congruency | 0.333 | 0.193 | 0.480 | 3.899 | 1.287 |
| Gender Congruency + Gender Congruency ✻ SOA | 0.333 | 0.052 | 0.110 | 14.471 | 2.146 |

*Note.* All models include SOA, Participant. **P(M):** distribution of prior model probabilities across all the models tested. P**(M|data):** posterior model probabilities. **$BF_{01}$**: Bayes factor that compares one model against the null model. **$BF_M$** : Bayes factor that compares one model against all others (i.e., difference between prior model odds and posterior model odds). For details, see Wagenmakers, 2017.

**B.Analysis of Effects**

| Effects | P(incl) | P(incl\|data) | $BF_{Inclusion}$ |
|---|---|---|---|
| Gender Congruency | 0.333 | 0.193 | 0.256 |
| SOA ✻ Gender Congruency | 0.333 | 0.052 | 0.269 |

*Note.* Compares models that contain the effect to equivalent models stripped of the effect. Higher-order interactions are excluded. Analysis suggested by Sebastiaan Mathôt.

**Table 7.** Bayesian repeated measures ANOVA, Experiment 4

**A.Model Comparison**

| Models | P(M) | P(M\|data) | $BF_M$ | $BF_{01}$ | error % |
|---|---|---|---|---|---|
| Null model (incl. SOA, Participant) | 0.333 | 0.801 | 8.061 | 1.000 | |
| Gender Congruency | 0.333 | 0.156 | 0.371 | 5.123 | 1.306 |
| Gender Congruency + Gender Congruency ✻ SOA | 0.333 | 0.042 | 0.089 | 18.904 | 4.294 |

*Note.* All models include SOA, Participant. **P(M):** distribution of prior model probabilities across all the models tested. P**(M|data):** posterior model probabilities. **$BF_{01}$**: Bayes factor that compares one model against the null model. **$BF_M$** : Bayes factor that compares one model against all others (i.e., difference between prior model odds and posterior model odds). For details, see Wagenmakers, 2017.

**B.Analysis of Effects**

| Effects | P(incl) | P(incl\|data) | $BF_{Inclusion}$ |
|---|---|---|---|
| Gender Congruency | 0.333 | 0.156 | 0.195 |
| SOA ✻ Gender Congruency | 0.333 | 0.042 | 0.271 |

*Note.* Compares models that contain the effect to equivalent models stripped of the effect. Higher-order interactions are excluded. Analysis suggested by Sebastiaan Mathôt.

## Figures and figure captions

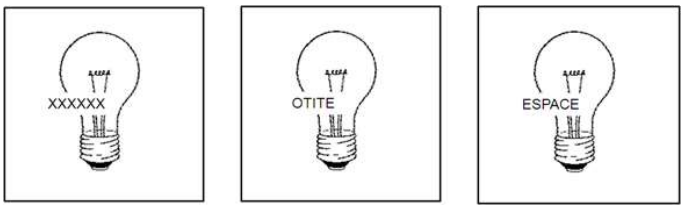



*Figure 1.* Example of stimuli in, from left to right, the baseline, gender congruent and gender incongruent conditions in Experiment 1. Participants are expected to produce *l'ampoule* 'the light bulb' in Experiment 1a, and *une ampoule* 'a light bulb' in Experiment 1b.

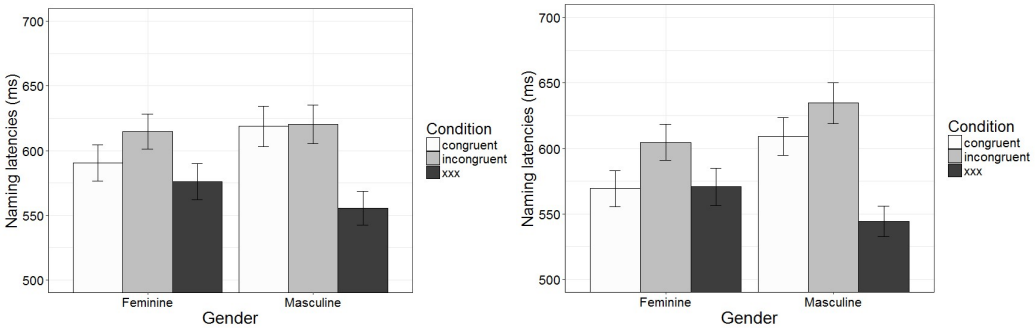



*Figure 2*. Observed mean naming latencies for gender congruent, gender incongruent, and baseline broken down by gender for the definite determiner (Experiment 1a, left panel) and the indefinite determiner (Experiment 1b, right panel). The errors bars represent the standard error of the means (values are adjusted for within-Participant designs following (Morey, 2008).

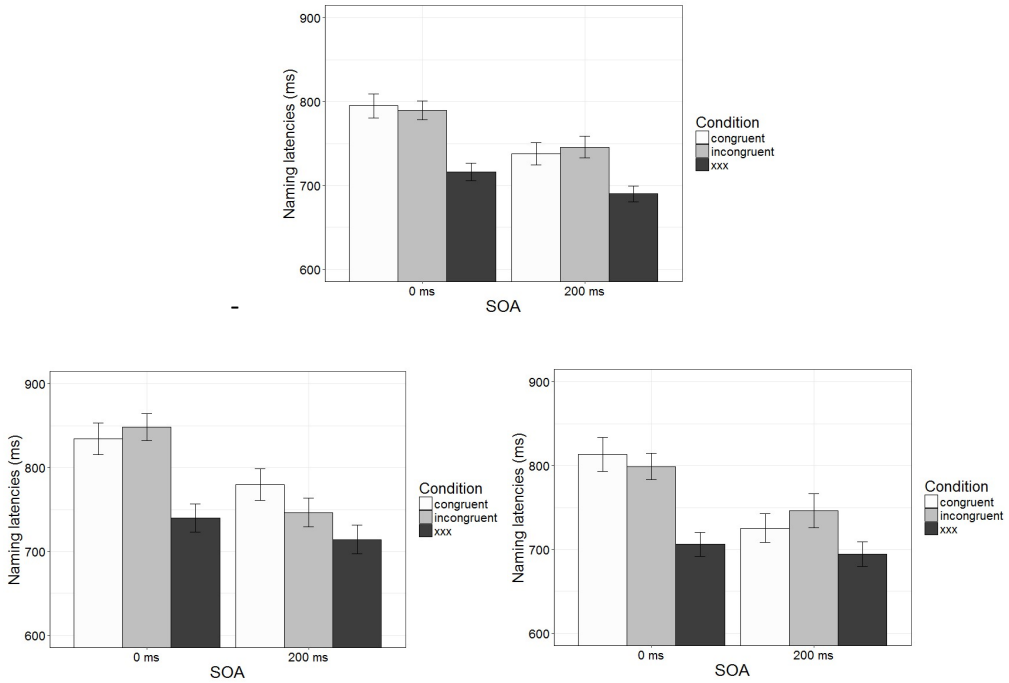



*Figure 3*. Observed mean naming latencies for gender congruent, gender incongruent, and baseline broken down by SOA for the definite determiner (Experiment 2a, upper panel); the masculine indefinite determiner (Experiment 2b, lower left panel); and the feminine indefinite determiner (Experiment 2b lower right panel). The errors bars represent the standard error of the means (values are adjusted for within-Participant designs following (Morey, 2008).

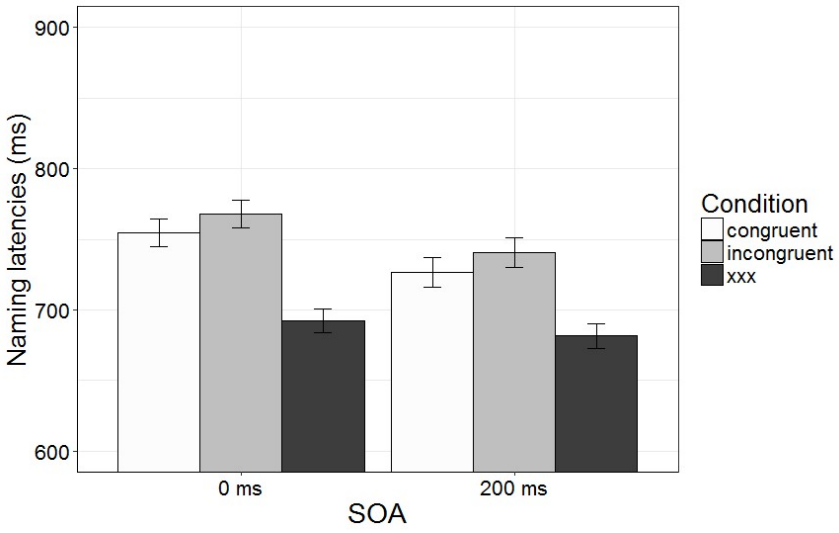



*Figure 4.* Observed mean naming latencies and standard errors for gender congruent, gender incongruent and baseline broken down by SOA in Experiment 3. The errors bars represent the

standard error of the means (values are adjusted for within-Participant designs following (Morey, 2008)

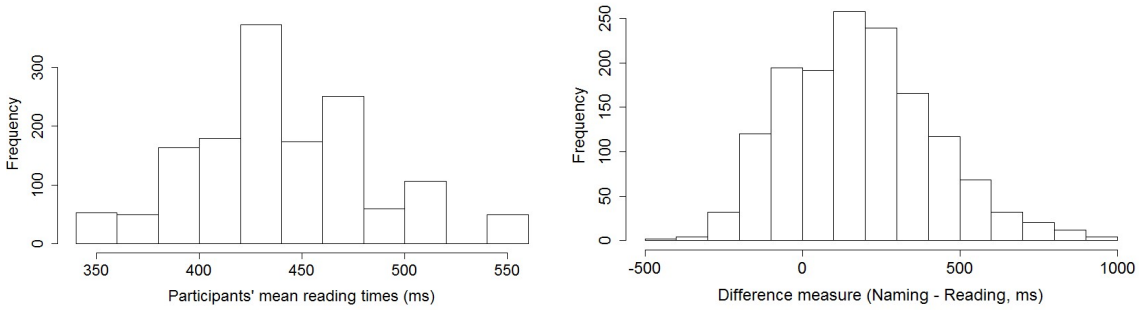



*Figure 5.* Participants' mean reading latencies in the word reading task (left panel) and difference between picture naming and word reading latencies for each picture-word association (right panel) in Experiment 3.

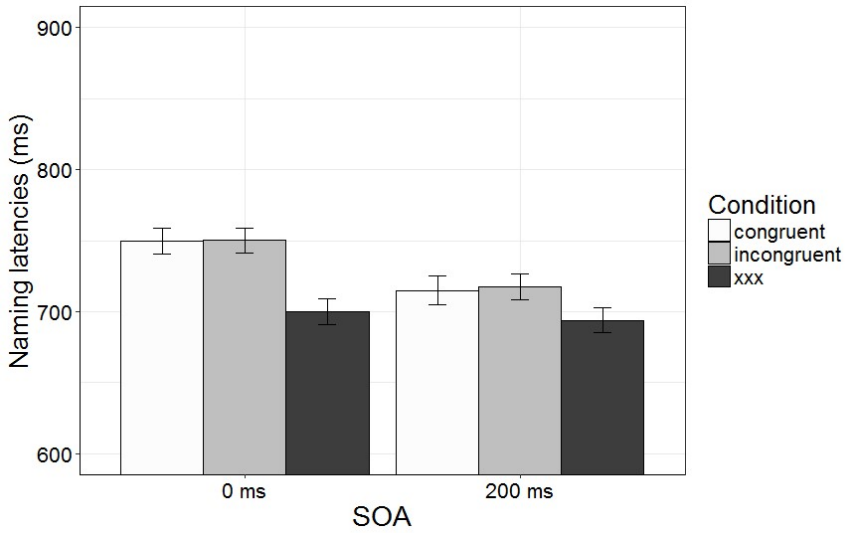



*Figure 6*. Observed mean naming latencies and standard errors (values are adjusted for within-Participant designs following Morey, 2008) for each distractor type and SOA in Experiment 4.

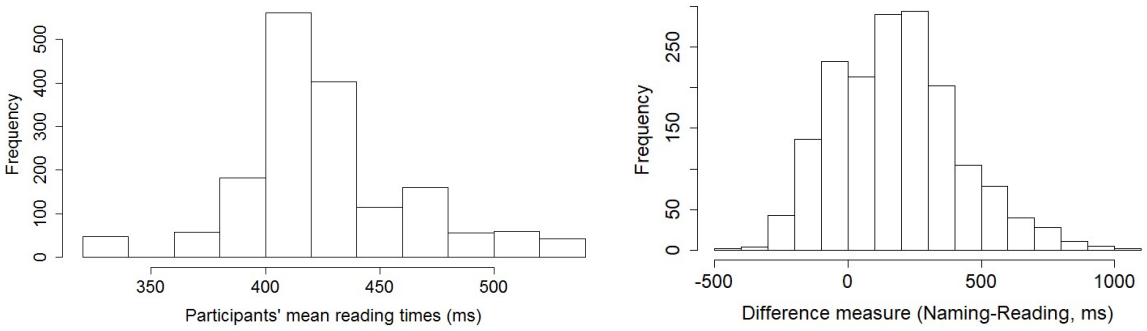



*Figure 7.* Participants' mean reading latencies in the word reading task (left panel) and difference between picture naming and word reading latencies for each picture-word association (right panel) in Experiment 4.

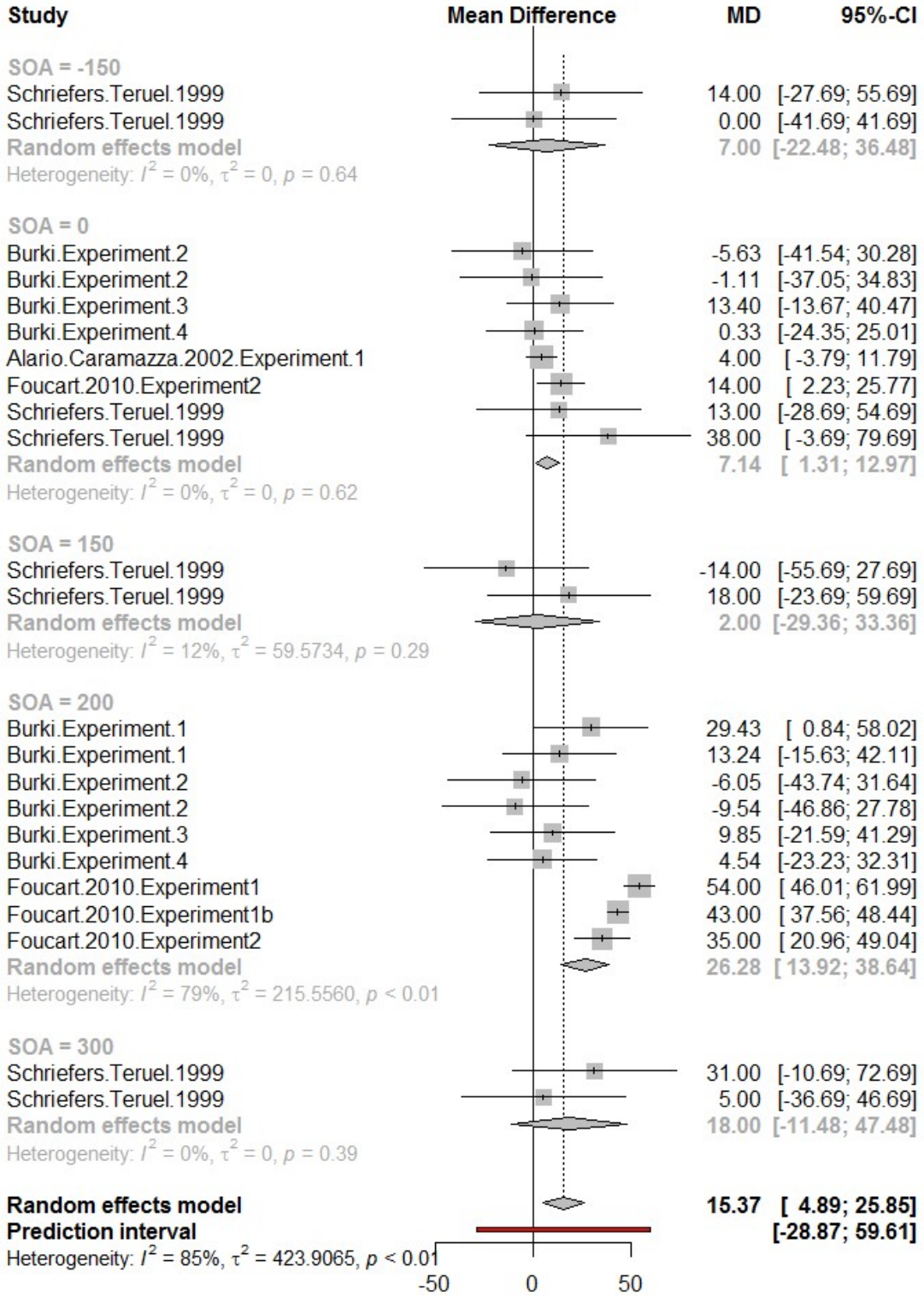



*Figure 8*. Results of Random effects meta-analysis. The analysis was performed with the R package *meta* (Schwarzer, 2007). Between study variance $\tau^2$ estimated using DerSimonian-Laird estimator (DerSimonian & Laird 1986), inverse variance weighting is used for pooling. For each study, the grey square highlights the mean difference between gender congruent and gender incongruent trials (= MD), the line extending either side of

the square represent the 95% confidence interval. Grey diamonds illustrate the results of the meta-analysis, with the treatment estimate in the center, and right and left sides corresponding to lower and upper confidence limits.

**Appendix A.** Materials used in Experiments 1a and 1b

| Gender | Target word | Congruent distractor | Incongruent distractor |
|---|---|---|---|
| Masculine | Accident (accident) | Écouteur (earphone) | Offre (offer) |
| | Agenda (diary) | Épisode (episode) | Étincelle (spark) |
| | Ail (garlic) | Orage (storm) | Énigme (mystery) |
| | Aimant (magnet) | Empire (empire) | Ordure (rubbish) |
| | Ananas (pineapple) | Épi (ear) | Étincelle (spark) |
| | Arbre (tree) | Œuf (egg) | Ordure (rubbish) |
| | Arc (bow) | Œuf (egg) | Olive (olive) |
| | Archet (bow) | Impôt (tax) | Épaule (shoulder) |
| | Arrosoir (watering can) | Exemple (example) | Idée (idea) |
| | Artichaut (artichoke) | Épisode (episode) | Écaille (scale) |
| | Avion (airplane) | Outil (tool) | Écaille (scale) |
| | Avocat (avocado) | Œil (eye) | Issue (exit) |
| | Balai (broom) | Village (village) | Fenêtre (window) |
| | Bouclier (shield) | Visage (face) | Cabane (shed) |
| | Bureau (desk) | Cœur (heart) | Corde (rope) |
| | Cahier (notebook) | Bijou (jewel) | Fleur (flower) |
| | Chapeau (hat) | Tunnel (tunnel) | Tondeuse (lawnmower) |
| | Château (castle) | Ciel (sky) | Pipe (pipe) |
| | Citron (lemon) | Fouet (whisk) | Gomme (eraser) |

| Gender | Target word | Congruent distractor | Incongruent distractor |
|---|---|---|---|
| | Collier (necklace) | Doigt (finger) | Punaise (tack) |
| | Couteau (knife) | Papier (paper) | Terre (soil) |
| | Drapeau (flag) | Pied (foot) | Lumière (light) |
| | Écran (screen) | Amour (love) | Urgence (emergency) |
| | Écrou (nut) | Album (album) | Armée (army) |
| | Entonnoir (funnel) | Accueil (welcome) | Industrie (industry) |
| | Escalier (stairs) | Outil (tool) | Urgence (emergency) |
| | Éventail (fan) | Appétit (appetite) | Industrie (industry) |
| | Évier (sink) | Achat (purchase) | Absence (absence) |
| | Fauteuil (armchair) | Chemin (way) | Tête (head) |
| | Filet (net) | Cactus (cactus) | Couleur (color) |
| | Igloo (igloo) | Achat (purchase) | École (school) |
| | Livre (book) | Feu (fire) | Maison (house) |
| | Oignon (onion) | Hiver (winter) | Épaule (shoulder) |
| | Ongle (nail) | Amour (love) | Idée (idea) |
| | Orgue (organ) | Emploi (job) | Erreur (error) |
| | Orteil (toe) | Impôt (tax) | Issue (exit) |
| | Os (bone) | Art (art) | Image (picture) |
| | Panier (basket) | Nuage (cloud) | Fougère (fern) |
| | Peigne (comb) | Train (train) | Main (hand) |
| | Pinceau (paintbrush) | Bras (arm) | Nuit (night) |
| | Pouce (thumb) | Vent (wind) | Liste (list) |
| | Sapin (fir tree) | Duvet (sleeping bag) | Clef (key) |
| | Sifflet (whistle) | Phare (light) | Cuisine (kitchen) |
| | Sourcil (eyebrow) | Jardin (garden) | Lampe (lamp) |
| | Tableau (painting) | Pont (bridge) | Jambe (leg) |
| | Tapis (rug) | Raisin (grape) | Route (road) |

| Gender | Target word | Congruent distractor | Incongruent distractor |
|---|---|---|---|
| | Tiroir (drawer) | Soleil (sun) | Lettre (letter) |
| | Tonneau (barrel) | Bagage (luggage) | Roue (wheel) |
| *Feminine* | Agrafeuse (stapler) | Erreur (error) | Épi (ear) |
| | Aiguille (needle) | Omelette (omelette) | Objet (object) |
| | Aile (wing) | Œuvre (work) | Intérêt (interest) |
| | Allumette (match) | Overdose (overdose) | Emploi (job) |
| | Ambulance (ambulance) | Image (picture) | Examen (examination) |
| | Ampoule (light bulb) | Otite (ear infection) | Espace (space) |
| | Ancre (anchor) | Olive (olive) | Œil (eye) |
| | Anse (handle) | École (school) | Uniforme (uniform) |
| | Antenne (antenna) | Ombre (shadow) | Opéra (opera) |
| | Arche (arch) | Œuvre (work) | Objet (object) |
| | Asperge (aspargus) | Option (option) | Uniforme (uniform) |
| | Assiette (plate) | Offre (offer) | Orage (storm) |
| | Baignoire (bathtube) | Médaille (medal) | Dragon (dragon) |
| | Bouée (rubber ring) | Dent (tooth) | Jeu (game) |
| | Bouteille (bottle) | Porte (door) | Vélo (bike) |
| | Carotte (carrot) | Piscine (swimming pool) | Rêve (dream) |
| | Casserole (pan) | Trompette (trumpet) | Piano (piano) |
| | Chaise (chair) | Rue (street) | Stylo (pen) |
| | Chaussette (sock) | Fumée (smoke) | Journal (newspaper) |
| | Chaussure (shoe) | Bêtise (stupidity) | Gâteau (cake) |
| | Cloche (bell) | Bombe (bomb) | Nœud (knot) |
| | Couronne (crown) | Glace (ice cream) | Pneu (tyre) |
| | Échelle (ladder) | Otite (ear infection) | Accueil (welcome) |
| | Église (church) | Absence (absence) | Aliment (food) |
| | Empreinte (fingerprint) | Affiche (poster) | Organe (organ) |

| Gender | Target word | Congruent distractor | Incongruent distractor |
|---|---|---|---|
| | Enveloppe (envelope) | Aubergine (aubergine) | Intérêt (interest) |
| | Épée (sword) | Aubergine (aubergine) | Album (album) |
| | Éprouvette (test tube) | Avalanche (avalanche) | Appétit (appetite) |
| | Étiquette (label) | Avalanche (avalanche) | Opéra (opera) |
| | Étoile (star) | Affiche (poster) | Organe (organ) |
| | Guitare (guitar) | Framboise (raspberry) | Pays (country) |
| | Ile (island) | Ombre (shadow) | Art (art) |
| | Lune (moon) | Neige (snow) | Cintre (hanger) |
| | Montre (watch) | Noix (walnut) | Bus (bus) |
| | Niche (dog house) | Brosse (brush) | Gland (acorn) |
| | Orange (orange) | Aventure (adventure) | Exemple (example) |
| | Oreille (ear) | Énigme (mystery) | Espace (space) |
| | Pelle (shovel) | Boîte (box) | Ski (ski) |
| | Plume (feather) | Moto (motorcycle) | Bol (bowl) |
| | Poire (pear) | Feuille (leaf) | Banc (bench) |
| | Pomme (apple) | Ville (city) | Clou (nail) |
| | Poubelle (bin) | Cravate (tie) | Cadeau (present) |
| | Robe (dress) | Cuillère (spoon) | Cerveau (brain) |
| | Saucisse (sausage) | Raquette (racket) | Disque (record) |
| | Scie (saw) | Langue (tongue) | Palmier (palm tree) |
| | Usine (factory) | Option (option) | Écouteur (earphone) |
| | Valise (suitcase) | Montagne (mountain) | Crayon (pencil) |
| | Voiture (car) | Bougie (candle) | Coussin (cushion) |

**Appendix B**. Mean naming latencies, standard-deviations, and standard errors (values are adjusted for within-Participant designs following Morey, 2008) in Experiment 1a and 1b for each Gender and Distractor type.

| Determiner | Gender | Distractor Type | N | Naming latencies | Standard Deviation | Standard Error |
|---|---|---|---|---|---|---|
| **Definite** | **Feminine** | **Congruent** | 333 | 780.8 | 227.7 | 12.5 |
| | | **Incongruent** | 353 | 805.4 | 228.1 | 12.1 |
| | | **Baseline** | 325 | 769.6 | 229.1 | 12.7 |
| | **Masculine** | **Congruent** | 305 | 811.5 | 251.1 | 14.4 |
| | | **Incongruent** | 305 | 813.0 | 244.3 | 14.0 |
| | | **Baseline** | 349 | 752.6 | 227.6 | 12.2 |
| **Indefinite** | **Feminine** | **Congruent** | 327 | 762.8 | 225.9 | 12.5 |
| | | **Incongruent** | 361 | 795.8 | 235.6 | 12.4 |
| | | **Baseline** | 320 | 766.9 | 235.7 | 13.2 |
| | **Masculine** | **Congruent** | 323 | 809.2 | 253.8 | 14.1 |
| | | **Incongruent** | 311 | 824.7 | 249.1 | 14.1 |
| | | **Baseline** | 359 | 744.3 | 213.1 | 11.2 |

**Appendix C.** Materials used in Experiments 2 to 4 (taken from Foucart et al., 2010)

| Target word Gender | Target word | Congruent distractor | Incongruent distractor |
|---|---|---|---|
| **Masculine** | **Balai** (broom) | Village (village) | Fenêtre (window) |
| | **Bouclier** (shield) | Voyage (travel) | Bouche (mouth) |
| | **Chapeau** (hat) | Voyage (travel) | Bouche (mouth) |
| | **Chou** (cabbage) | Train (train) | Main (hand) |
| | **Collier** (necklace) | Doigt (finger) | Ville (town) |
| | **Couteau** (knife) | Doigt (finger) | Terre (earth) |
| | **Drapeau** (flag) | Pied (foot) | Lumière (light) |
| | **Fauteuil** (armchair) | Chemin (path) | Tête (head) |
| | **Gant** (glove) | Ciel (sky) | Terre (earth) |
| | **Genou** (knee) | Train (train) | Main (hand) |
| | **Haricot** (bean) | Village (village) | Fenêtre (window) |
| | **Nez** (nose) | Pied (foot) | Main (hand) |
| | **Nœud** (node) | Soleil (sun) | Lettre (letter) |
| | **Nuage** (cloud) | Feu (fire) | Maison (house) |
| | **Panier** (basket) | Soleil (sun) | Lettre (letter) |
| | **Pinceau** (paintbrush) | Bras (arm) | Maison (house) |
| | **Pont** (bridge) | Feu (fire) | Jambe (leg) |
| | **Pouce** (thumb) | Vent (wind) | Ligne (line) |
| | **Sapin** (pine tree) | Papier (paper) | Jambe (leg) |
| | **Seau** (bucket) | Papier (paper) | Lettre (letter) |

| | | | |
|---|---|---|---|
| | **Sourcil** (eyebrow) | Jardin (garden) | Fleur (flower) |
| | **Stylo** (pen) | Cœur (heart) | Nuit (night) |
| | **Tableau** (board) | Soleil (sun) | Jambe (leg) |
| | **Tapis** (carpet) | Cœur (heart) | Nuit (night) |
| **Feminine** | **Baignoire** (bathtub) | Ligne (line) | Vent (wind) |
| | **Bouteille** (bottle) | Porte (door) | Doigt (finger) |
| | **Brosse** (brush) | Table (table) | Rêve (dream) |
| | **Casserole** (saucepan) | Fleur (flower) | Jardin (garden) |
| | **Chaise** (chair) | Rue (street) | Bras (arm) |
| | **Chaussette** (sock) | Table (table) | Rêve (dream) |
| | **Chaussure** (shoe) | Table (table) | Rêve (dream) |
| | **Clef** (key) | Rue (street) | Bras (arm) |
| | **Cloche** (bell) | Rue (street) | Feu (fire) |
| | **Couronne** (crown) | Ligne (line) | Vent (wind) |
| | **Dent** (tooth) | Nuit (night) | Cœur (heart) |
| | **Fourchette** (fork) | Ville (town) | Chemin (path) |
| | **Fumée** (smoke) | Terre (earth) | Chemin (path) |
| | **Glace** (ice-cream) | Porte (door) | Jeu (game) |
| | **Montre** (watch) | Lumière (light) | Train (train) |
| | **Neige** (snow) | Bouche (mouth) | Voyage (travel) |
| | **Niche** (doghouse) | Tête (head) | Ciel (sky) |
| | **Pelle** (shovel) | Fleur (flower) | Jardin (garden) |
| | **Poire** (pear) | Ville (town) | Jeu (game) |

| | | |
|---|---|---|
| **Pomme** (apple) | Tête (head) | Ciel (sky) |
| **Poubelle** (bin) | Maison (house) | Papier (paper) |
| **Robe** (dress) | Fenêtre (window) | Pied (foot) |
| **Saucisse** (sausage) | Lumière (light) | Village (village) |
| **Voiture** (car) | Porte (door) | Jeu (game) |

**Appendix D**. Mean naming latencies, standard-deviations, and standard errors (values are adjusted for within-Participant designs following Morey, 2008) in Experiments 2a and 2b for each SOA, Gender, and Distractor type.

| Determiner | SOA | Gender | Distractor Type | N | Naming latencies | Standard Deviation | Standard Error |
|---|---|---|---|---|---|---|---|
| **Definite** | **0ms** | **Feminine** | **Congruent** | 135 | 767.0 | 199.7 | 17.2 |
| | | | **Incongruent** | 127 | 783.1 | 167.1 | 14.8 |
| | | | **Baseline** | 138 | 710.8 | 152.4 | 13.0 |
| | | **Masculine** | **Congruent** | 125 | 825.2 | 234.4 | 21.0 |
| | | | **Incongruent** | 123 | 795.8 | 161.4 | 14.6 |
| | | | **Baseline** | 122 | 722.0 | 174.0 | 15.8 |
| | **200ms** | **Feminine** | **Congruent** | 133 | 729.9 | 184.4 | 16.0 |
| | | | **Incongruent** | 124 | 739.5 | 201.1 | 18.1 |
| | | | **Baseline** | 135 | 683.1 | 146.0 | 12.6 |
| | | **Masculine** | **Congruent** | 113 | 747.1 | 202.4 | 19.0 |
| | | | **Incongruent** | 119 | 752.1 | 174.8 | 16.0 |
| | | | **Baseline** | 125 | 697.2 | 130.0 | 11.6 |
| **Indefinite** | **0ms** | **Feminine** | **Congruent** | 140 | 813.3 | 226.5 | 19.1 |
| | | | **Incongruent** | 155 | 798.9 | 183.6 | 14.7 |
| | | | **Baseline** | 150 | 706.0 | 170.3 | 13.9 |
| | | **Masculine** | **Congruent** | 140 | 834.4 | 207.5 | 17.5 |
| | | | **Incongruent** | 143 | 848.6 | 182.1 | 15.2 |
| | | | **Baseline** | 149 | 740.1 | 193.1 | 15.8 |
| | **200ms** | **Feminine** | **Congruent** | 146 | 725.5 | 194.6 | 16.1 |
| | | | **Incongruent** | 142 | 746.5 | 227.2 | 19.1 |
| | | | **Baseline** | 148 | 694.1 | 165.8 | 13.6 |
| | | **Masculine** | **Congruent** | 140 | 780.1 | 210.3 | 17.8 |
| | | | **Incongruent** | 145 | 746.4 | 193.4 | 16.1 |
| | | | **Baseline** | 148 | 714.3 | 192.5 | 15.8 |

**Appendix E**. Results of joint analysis of Experiments 2a and 2b

**Mixed-effects model, main effects only**

| | β | S.E | *t* | *p* |
|---|---|---|---|---|
| **Intercept** | 777.6 | 30.5 | 25.5 | < 0.0001 |
| **Position trial in Experiment** | 0.71 | 0.14 | 5.3 | < 0.0001 |
| **Gender Congruency** | 6.6 | 8.3 | 0.79 | 0.44 |
| **Determiner** | 18.4 | 34.4 | 0.53 | 0.60 |
| **SOA** | 55.7 | 14.8 | 3.8 | 0.00048 |
| **Gender** | 44.8 | 24.2 | 1.85 | 0.071 |

β = Estimate, S.E. = Standard error of the estimate

**Mixed-effects model, with two-way interactions and three-way interaction between Gender congruency, Determiner, and SOA.**

| | β | S.E | *t* | *p* |
|---|---|---|---|---|
| **Intercept** | 779.1 | 31.9 | 24.5 | < 0.0001 |
| **Position trial in Experiment** | 0.73 | 0.14 | 5.3 | < 0.0001 |
| **Gender Congruency** | 7.7 | 15.8 | 0.49 | 0.63 |
| **Determiner** | 16.3 | 36.7 | 0.45 | 0.66 |
| **SOA** | 42.8 | 24.2 | 1.77 | 0.082 |
| **Gender** | 52.5 | 26.4 | 1.99 | 0.0516 |
| **Determiner ✻ SOA** | 9.1 | 30.1 | 0.3 | 0.76 |
| **SOA ✻ Gender** | 3.1 | 16.5 | 0.19 | 0.85 |
| **Gender Congruency ✻ Determiner** | 2.6 | 18.9 | 0.14 | 0.89 |
| **Gender Congruency ✻ SOA** | 8.8 | 16.9 | 0.52 | 0.60 |
| **Determiner ✻ Gender** | 0.71 | 14.2 | 0.05 | 0.96 |
| **Gender Congruency ✻ Gender** | 15.9 | 13.6 | 1.17 | 0.25 |
| **Gender Congruency ✻ Determiner ✻ SOA** | 16.0 | 23.1 | 0.69 | 0.49 |

β = Estimate, S.E. = Standard error of the estimate

**Repeated measures Bayesian ANOVA**

**A. Model Comparison**

| Models | P(M) | P(M\|data) | $BF_M$ | $BF_{01}$ | error % |
|---|---|---|---|---|---|
| **Null model (incl. Determiner, SOA, Participant)** | 0.091 | 0.613 | 15.836 | 1.000 | |
| **Gender Congruency** | 0.091 | 0.103 | 1.150 | 5.942 | 3.934 |
| **Gender Congruency + Gender Congruency ✻ Determiner** | 0.091 | 0.025 | 0.258 | 24.376 | 4.429 |
| **Gender Congruency + Gender Congruency ✻ SOA** | 0.091 | 0.024 | 0.245 | 25.668 | 4.430 |
| **Gender Congruency + Gender Congruency ✻ Determiner +** | 0.091 | 0.006 | 0.062 | 98.811 | 5.326 |

**A. Model Comparison**

| Models | P(M) | P(M\|data) | $BF_M$ | $BF_{01}$ | error % |
|---|---|---|---|---|---|
| **Gender Congruency ✻ SOA** | | | | | |
| **Determiner ✻ SOA** | 0.091 | 0.179 | 2.178 | 3.428 | 3.806 |
| **Gender Congruency + Determiner ✻ SOA** | 0.091 | 0.032 | 0.330 | 19.185 | 4.218 |
| **Gender Congruency + Gender Congruency ✻ Determiner + Determiner ✻ SOA** | 0.091 | 0.008 | 0.078 | 79.639 | 5.279 |
| **Gender Congruency + Gender Congruency ✻ SOA + Determiner ✻ SOA** | 0.091 | 0.008 | 0.077 | 80.358 | 5.214 |
| **Gender Congruency + Gender Congruency ✻ Determiner + Gender Congruency ✻ SOA + Determiner ✻ SOA** | 0.091 | 0.002 | 0.020 | 311.594 | 5.668 |
| **Gender Congruency + Gender Congruency ✻ Determiner + Gender Congruency ✻ SOA + Determiner ✻ SOA + Gender Congruency ✻ Determiner ✻ SOA** | 0.091 | 5.985e -4 | 0.006 | 1024.178 | 6.917 |

*Note.* All models include Determiner, SOA, Participant. **P(M):** distribution of prior model probabilities across all the models tested. P**(M|data):** posterior model probabilities. **$BF_{01}$**: Bayes factor that compares one model against the null model. **$BF_M$** : Bayes factor that compares one model against all others (i.e., difference between prior model odds and posterior model odds). For details, see Wagenmakers, 2017.

**B.Analysis of Effects**

| Effects | P(incl) | P(incl\|data) | $BF_{Inclusion}$ |
|---|---|---|---|
| **Gender Congruency** | 0.182 | 0.135 | 0.171 |
| **Determiner ✻ SOA** | 0.455 | 0.228 | 0.296 |
| **Determiner ✻ Gender Congruency** | 0.364 | 0.041 | 0.246 |
| **SOA ✻ Gender Congruency** | 0.364 | 0.040 | 0.236 |
| **Determiner ✻ SOA ✻ Gender Congruency** | 0.091 | 5.985e -4 | 0.304 |

*Note.* Compares models that contain the effect to equivalent models stripped of the effect. Higher-order interactions are excluded. Analysis suggested by Sebastiaan Mathôt.

**Appendix F.** Additional 32 words used in word reading task (Experiments 3 and 4).

| Word | Gender | Written Frequency (nb of occurences per million) | Number of letters |
|---|---|---|---|
| type | M | 184.2 | 4 |
| trou | M | 108.4 | 4 |
| soir | M | 562.6 | 4 |
| semaine | F | 197.5 | 7 |
| salut | M | 61.8 | 5 |
| sac | M | 174.3 | 3 |
| raison | F | 308.8 | 6 |
| problème | M | 95.0 | 8 |
| prix | M | 107.5 | 4 |
| pluie | F | 122.6 | 5 |
| pierre | F | 189.9 | 6 |
| peur | F | 311.7 | 4 |
| nord | M | 72.3 | 4 |
| nom | M | 395.0 | 3 |
| neige | F | 80.9 | 5 |
| monde | M | 741.4 | 5 |
| minute | F | 201.4 | 6 |
| matin | M | 396.8 | 5 |
| langue | F | 126.3 | 6 |
| lampe | F | 93.1 | 5 |

| | | | |
|---|---|---|---|
| jour | M | 1341.8 | 4 |
| joie | F | 150.2 | 4 |
| homme | M | 1398.9 | 5 |
| forêt | F | 113.3 | 5 |
| fille | F | 592.2 | 5 |
| famille | F | 274.4 | 7 |
| faim | F | 76.0 | 4 |
| droit | M | 163.7 | 5 |
| cuisine | F | 135.4 | 7 |
| coin | M | 199.3 | 4 |
| carte | F | 111.4 | 5 |
| bord | M | 228.1 | 4 |
| **Mean** | | **291** | **4.9** |
| **Min** | | **62** | **3** |
| **Max** | | **1399** | **8** |

**Appendix G**. Mean naming latencies, standard-deviations, and standard errors (values are adjusted for within-Participant designs following Morey, 2008) in Experiments 3 and 4 for each SOA and Distractor type.

| Experiment 3 | | | | | |
|---|---|---|---|---|---|
| **SOA** | **Distractor Type** | **N** | **Naming latencies** | **Standard Deviation** | **Standard Error** |
| **0ms** | **Congruent** | 368 | 754.4 | 190.1 | 9.9 |
| | **Incongruent** | 361 | 767.8 | 182.3 | 9.6 |
| | **Baseline** | 389 | 692.3 | 166.8 | 8.5 |
| **200ms** | **Congruent** | 361 | 726.5 | 203.3 | 10.7 |
| | **Incongruent** | 369 | 740.6 | 206.3 | 10.7 |
| | **Baseline** | 382 | 681.5 | 170.2 | 8.7 |

| Experiment 4 | | | | | |
|---|---|---|---|---|---|
| **SOA** | **Distractor Type** | N | **Naming latencies** | **Standard Deviation** | **Standard Error** |
| **0ms** | **Congruent** | 429 | 749.9 | 186.6 | 9.0 |
| | **Incongruent** | 419 | 750.3 | 178.4 | 8.7 |
| | **Baseline** | 430 | 700.2 | 191.2 | 9.2 |
| **200ms** | **Congruent** | 425 | 714.9 | 206.4 | 10.0 |
| | **Incongruent** | 413 | 717.7 | 186.8 | 9.2 |
| | **Baseline** | 427 | 694.0 | 185.4 | 9.0 |