ARTICLE PUBLISHED IN PSYCHONOMIC BULLETIN & REVIEW

The Gender Congruency Effect across languages in bilinguals: A meta-analysis

Ana Rita Sá-Leite[1], Karlos Luna[2], Isabel Fraga[1], & Montserrat Comesaña[3, 4]

[1]Cognitive Processes & Behaviour Research Group, Department of Social Psychology, Basic

Psychology & Methodology, University of Santiago de Compostela, Spain

[2]Human Memory Research group, CIPsi, School of Psychology, University of Minho, Portugal

[3]Psycholinguistics Research group, CIPsi, School of Psychology, University of Minho, Portugal

[4]Centro de Enseñanza Superior Cardenal Cisneros, Universidad Complutense de Madrid, Spain

Corresponding author:

Ana Rita Sá Leite

Cognitive Processes & Behaviour Research Group

Department of Social Psychology, Basic Psychology & Methodology

University of Santiago de Compostela

15782, Santiago de Compostela, Spain

E-mail: anarita.saleite.dias@usc.es

Abstract

In the study of gender representation and processing in bilinguals, two contrasting perspectives exist: integrated vs. the autonomous (Costa, Kovacic, Fedorenko, & Caramazza, 2003). In the former, cross-linguistic interactions during the selection of grammatical gender values are expected; in the latter, they are not. To address this issue, authors have typically explored the cross-linguistic Gender Congruency Effect (GCE: a facilitation on the naming or translation of second language [L2] nouns when their first language [L1] translations are of the same gender, in comparison to those of a different gender). However, the literature suggests that this effect is sometimes difficult to observe and might vary as a function of variables such as the syntactic structure produced to translate or name the target (bare nouns vs. noun phrases), the phonological gender transparency of both languages (whether or not they have phonological gender cues associated with the ending letter [e.g., "–a" for feminine words and "–o" for masculine words in Romance languages]), the degree of L2 proficiency, and task requirements (naming vs. translation). The aim of the present quantitative meta-analysis is to examine the robustness of the cross-linguistic GCE obtained during language production. It involves 25 experiments from 11 studies. The results support a bilingual gender-integrated view, in that they show a small but significant GC effect regardless of the variables mentioned above.

Key words: bilingualism, gender representation and processing, gender congruency effect.

Introduction

Grammatical gender is an abstract lexical-syntactic property of nouns present in every language that has a gender system (Schriefers & Jescheniak, 1999). A gender system classifies words according to different categories known as gender values (Corbett, 1991). For instance, in Spanish there are two gender values: feminine (e.g., '*mesa*' [table]) and masculine (e.g., '*abrigo*' [coat]). The main function of grammatical gender is to determine the agreement relations that take place between the gender value of a given noun and the other elements in speech (such as determiners and adjectives). To fulfil the requirements of agreement, these elements have to change morphologically or to modify their form. For example, in Spanish, because '*abrigo*' (coat) is masculine, the phrase 'the expensive coat' would be translated as '*El abrigo caro*', but conversely, if the referent were feminine, such as '*mesa*' (table), then the phrase would be translated as '*La mesa cara*'. As can be seen in these examples, the noun ('*abrigo*' or '*mesa*') is the word that genuinely carries the gender information that determines the form of other elements. For this reason, gender is defined as an inherent characteristic of every noun. This has important implications for lexical processing, and therefore many authors have focused on how gender is represented in the lexicon and how it is selected during the processing of nouns from native and non-native languages (e.g., Cubelli, Lotto, Paolieri, Girelli, & Job, 2005; Miozzo, Costa, & Caramazza, 2002; Sá-Leite, Fraga, & Comesaña, 2019; Schiller & Caramazza, 2003).

In the present study we explore the representation and processing of grammatical gender during the production of nouns in bilinguals. To this end, we first address the models that have explicitly considered this lexical-syntactic feature in their formulations. We then introduce the main hypotheses about the organization of the bilingual grammatical gender system (the integrated vs. the autonomous view) and we survey the recent and current literature concerning the most prominent effect used to test them: the cross-linguistic gender

congruency effect (GCE). Based on the most notable theories and unanswered questions about grammatical gender representation, plus experimental results obtained with bilingual populations, we identify a group of variables that may have an impact on the effect, and hence also on the conclusions we can draw about the organization of the bilingual's grammatical gender system. We therefore conduct a meta-analysis to explore the robustness of this effect as well as the role of such variables on its size. Results shed some light on the still controversial issues regarding the nature of grammatical gender representation, and more specifically on the peculiarities of bilingual gender representation and processing.

Theoretical issues of grammatical gender representation and processing

Only a few models of language production have overtly assessed the representation of grammatical gender. These have based their predictions on data obtained with presumably "monolingual" populations (since no information was given regarding the linguistic knowledge of participants concerning second or foreign languages). Two models constitute the main references in the area: WEAVER++ (Levelt, Roelofs, & Meyer, 1999) and the Independent Network model (IN, Caramazza, 1997; Caramazza & Miozzo, 1997). WEAVER++ states that grammatical features are represented in a stratum called "lemma", which includes the abstract syntactic representations of words and is located between the conceptual and the lexeme strata (the latter being where we can find the ortho-phonological form of words; see Figure 1). According to this model, lexical access initiates with the activation of a concept in the conceptual stratum and the subsequent spreading of a proportional part of this activation to all associated lemmas (e.g., the concept of TABLE would activate lemmas such as table, desk, bench). Lemmas compete, so that the lemma that reaches the highest degree of activation is selected. Consequently, every related grammatical feature, including grammatical gender, which is represented by a specific node (e.g., in Spanish, there would be a node for feminine gender value and a node for masculine gender

value) is also activated. In the case of Spanish, if '*mesa*' (table) is the lemma to be processed, the feminine gender node will be activated. Critically, activation is not the same as selection, and the authors state that gender access only occurs if it is required for the task, i.e., only if grammatical gender is explicitly asked to be taken into consideration (e.g., pressing different buttons depending on the gender value of the target nouns) or if agreement must be fulfilled (Levelt, 2001). Hence, experimentally, this model predicts no gender effects during the production of bare nouns (BNs), when no other words need to be completely or morphologically adapted to the gender of the target noun at issue. For instance, in the previous example, feminine gender value would have been selected only if the speaker, instead of '*mesa*', had produced a phrase such as '*la mesa cara*' (the expensive table). On the other hand, the IN model assumes that phonological processing can be completely independent of syntactic processing, since the latter is sometimes dispensable for word production. According to this model, activation spreads outwards through three networks: a lexical-semantic network, a lexical-syntactic network, and two lexeme networks, the phonological one and the orthographic one (See Figure 2). During the production of a word, a selected lexical-semantic representation propagates activation in parallel towards the lexical-syntactic and the Phonological (P) and Orthographic (O) lexeme networks. Importantly, because of this parallel activation, the selection of the lexeme node does not depend on the prior selection of its associated syntactic features (including grammatical gender). Lexical-semantic to phonological processing can occur without grammatical gender being accessed. Similarly to the WEAVER++ model, this allows the rationale that gender will not be selected if agreement is not necessary.

<INSERT FIGURE 1 HERE>

<INSERT FIGURE 2 HERE>

Experimental studies conducted with a picture-word interference paradigm (PWIP) and native speakers of Germanic languages support this deeply syntactical view of grammatical gender representation. Within this paradigm, participants are asked to name a picture (the target) as fast as possible while ignoring a superimposed written noun (the distractor). They are instructed to use a noun phrase, usually formed by a definite determiner and a noun (NP; e.g., in Dutch, '*de stoel*' [the chair]). Target and distractor can both be of the same gender value (e.g., in Dutch, '*stoel*' [chair] and '*kat*' [cat] are both of common gender) or of different gender values (e.g., common noun '*stoel*' and neuter noun '*huis*' [house]). What is systematically found with "monolingual" or native speakers of Germanic languages is a GCE by which naming latencies are shorter when the gender of the distractor and target nouns is congruent in comparison to incongruent pairs (e.g., Costa, Kovacic, Fedorenko, & Caramazza, 2003; Miozzo & Caramazza, 1999; Schiller & Caramazza, 2003; Schriefers & Teruel, 2000; van Berkum, 1997). However, when participants are asked to name the pictures using BNs instead of NPs (e.g., '*stoel*' rather than '*de stoel*') the authors fail to observe any signs of a GCE (e.g., Caramazza, Miozzo, Costa, Schiller, & Alario, 2001; La Heij, Mak, Sander, & Willeboorsde, 1998; Starreveld & La Heij, 2004). The absence of the effect with BNs has been interpreted not only as evidence in support of the assumptions of the abovementioned models regarding the need for an agreement context in order that a noun be selected, but also as evidence supporting a more extreme view. Specifically, Caramazza et al. (2001) argue that gender is not the subject of competitive processes here, in that it is an automatic consequence of lexical selection itself, and hence the observed effect is the result of competition at the level of the selection of definite articles (e.g., Costa, Kovacic, Fedorenko, & Caramazza, 2003; Schriefers, 1993).

Nevertheless, studies conducted with native speakers of Romance languages producing BNs have called the universality of these amendments into question, highlighting not the

syntactic but the lexical nature of grammatical gender. This is because gender effects in the form of a gender-incongruent effect (the opposite direction of a GCE, i.e., shorter naming times for incongruent pairs) have been found with Italian and Spanish participants (Cubelli et al., 2005; Paolieri et al., 2010; Paolieri, Lotto, Leoncini, Cubelli, and Job, 2011; see however evidence for a classical GCE with European Portuguese native speakers; Sá-Leite, Oliveira, Soares, Carreiras, & Comesaña, 2017). The reason for these contrasting results with BNs between certain Germanic and Romance languages is not clear, but the degree of phonological gender transparency has been noted as a possible factor. This variable is related to the proportion of transparent nouns in a language. Transparent nouns are those in which we can find regularities in the ending letters in relation to gender (e.g., Harris, 1991; De Martino, Bracco, Postiglione, & Laudanna, 2017; Sá-Leite et al., 2017). For instance, in Spanish, Italian and Portuguese, the majority of masculine nouns tend to end in '-o' and the majority of feminine nouns in '-a' (e.g., in Spanish, '*abrigo*' [masculine] and '*mesa*' [feminine]). When masculine and feminine nouns end in other vowels or consonants that do not generally relate to a specific gender value, such as the feminine Spanish noun '*torre*' (tower) or the masculine Spanish noun '*enjambre*' (swarm), we call these opaque nouns. On the basis of the contrasting results between linguistic families, Cubelli et al. (2005) created a model that sought to link grammatical gender selection to nominal endings (i.e., to gender transparency), the Double-Selection Model. On this view, the architecture of our linguistic system is universal, but the mechanisms that operate on it are constrained by the peculiarities of each language. Critically, syntactic information is an intrinsic part of the lexical representation of nouns. Therefore, at the lemma level a word is specified both semantically and syntactically with its own gender. For a language like Italian, it is assumed that both semantic and syntactic information have to be selected to access the morpho-phonological form of a noun. This is because of the complex morphological structure of Italian nouns: in this language grammatical gender and number

interact, since the morphological mark for plural depends on the grammatical gender value itself (e.g., '*luna*' [moon] is feminine, and takes the morpheme '-e' for plural, '*lune*', but '*libro*' [book] is masculine and so takes the morpheme '-i' for plural, '*libri*'). Thus, it is argued, semantic information allows the stem selection of nouns, and grammatical properties are needed to access the correct nominal ending and the associated declensional class (i.e., without number and gender, speakers do not know the ending of a noun). When target and distractor appear on the screen, competition occurs between semantically similar words but also between words with similar syntactic characteristics (i.e., nouns of the same gender, originating a gender-incongruency effect). According to Cubelli et al. (2005), in the case of Germanic languages such as Dutch, because word form is not related to gender values (it is not a transparent language) there is no need for gender to be processed to access the morpho-phonological form of the word, and thus no gender effects are observed. This model challenges the classic interaction proposed between the syntactic and the morpho-phonological strata, but in our opinion it seems somewhat limited in its predictions. First, there is a vast body of evidence supporting the *singular-as-default hypothesis* (Bock, 1995; Schriefers, Jescheniak, & Hantsch, 2002), that is, showing that singular acts as the default value for number. In other words, singular features are firstly activated and plural features are activated only when this is overtly required. According to Cubelli's model, however, number and grammatical gender interact to select, for instance, the feminine singular vs. feminine plural nominal ending ('-a' vs. '-e'), even when a speaker's intention is to produce a word in the singular. If the *singular-as-default hypothesis* is right, this is debatable, as plural would not be taken into account in the first place. Second, Spanish and Portuguese have a much simpler gender structure in which transparency is not associated with morphemes but with phonological endings statistically related to gender (in Spanish, '-a' is not a morpheme in '*casa*'). Also, number morphemes do not vary depending on the gender value of the noun, yet gender effects are observed. In this

sense, even for opaque nouns (in which nominal endings are not phonologically related to gender) gender effects have been found (Paolieri et al., 2011; Sá-Leite et al., 2017). Finally, the model fails to explain the gender congruent (and not incongruent) effect found with speakers of European Portuguese in that it argues for competition by similarity. In this sense, Duràn and Pillon (2011) recently formulated an activation-based model in which grammatical features, including gender nodes, would be located at the lemma level, but separately from the abstract lexical entries. These grammatical features would have direct connections with both the lexical entries (bidirectional connections) and with the morphological processes (forward-unidirectional connections). Once again, however, according to this model morphological processes only receive activation from grammatical features, and thus gender values are necessary to select the morphological endings of nouns, which is highly debatable as '-a' in '*casa*' (house) in Spanish and Portuguese is not a morpheme but merely a phonological cue for gender. Importantly, the gender-incongruency effect would not be predicted here, since competition does not occur by similarity.

In sum, the true nature of gender as one of the most central elements in linguistic analysis is far from clear. Contrasting results between linguistic families continue to challenge universal views on grammatical gender processing during language production, and many essential questions, such as the competitive vs. automatic selection of grammatical gender, agreement-as-a-condition for gender processing, the abstract gender feature vs. determiner form selection at the root of the GCE, and the role of phonological gender transparency during gender access, have yet to be understood.

Grammatical gender representation and processing in bilingual populations

In the bilingual domain, the study of gender as a lexical property provides an excellent opportunity not only to provide evidence for the still unclear debate about the nature of its

representation, but also to examine the organization of grammatical features across languages. Although types of cross-language interference have been observed at different levels of processing (see Comesaña et al., 2015, 2018; Dijkstra, Miwa, Brummelhuis, Sappelli, & Baayen, 2010; Sá-Leite, Fraga, & Comesaña, 2019 for more detail; also Soares, Oliveira, Comesaña, & Costa, 2018; Soares et al., 2019, for evidence of interactions between lexical-syntactic levels), research at the lemma level, where gender is thought to be stored, is scarcer and more controversial than at the lexeme level (see Kroll & Tokowicz, 2005). Hence, over the last decade a number of studies have explored cross-language interactions at the grammatical level through the processing of grammatical gender, testing two conflicting perspectives on the structure of bilingual gender representation: the integrative vs. the autonomous view (Costa, Kovacic, Franck, & Caramazza, 2003; Klassen, 2016). Importantly, both views make their predictions on two assumptions: (1) that grammatical gender is an abstract feature subject to competition, not accessed automatically, as some authors have argued (Caramazza et al., 2001), and (2) that in line with one of the most influential models of bilingual language organization, the Revised Hierarchical Model (Kroll & Stewart, 1994), there is a direct connection between the two languages at the level of lexical representation (lemma). However, they differ in the organization of gender values across languages in the bilingual mind. According to the integrative view, bilinguals share the representation of grammatical gender values across languages in a unique integrated-gender system. Thus, the activation of a word and its gender value in a specific language affects the gender activation of its translation in the other language (see Figure 3). On the contrary, the autonomous view posits that each language has its own particular gender system. Hence, whether or not two translations share the same gender value is irrelevant for the organization of second language (L2) grammatical knowledge (see Figure 4). Simply put, because gender values are taken to

be independent across languages, the gender of a target word does not compete with the gender of its translation equivalent.

<INSERT FIGURE 3 HERE>

<INSERT FIGURE 4 HERE>

To test the tenets of these two proposals, many authors have based their research on the so-called cross-linguistic GCE. This effect is usually obtained within a simple picture-naming task in which participants have to name aloud a picture in their L2. The automatic activation of the first language (L1) translation equivalents of the L2 nouns that depict the pictures generates the classical gender-congruent and incongruent conditions depending on whether or not they share the same gender value. Therefore, conditions are created through the careful selection of homogeneric nouns (i.e., nouns that have the same gender values in both languages) and heterogeneric ones (i.e., nouns that have one gender value in one language and another gender in the other language). The GCE consists of faster responses to homogeneric target nouns than to heterogeneric ones. This effect has also been studied in forward translation tasks (i.e., participants are presented with an L1 noun on a screen and they have to translate it into the L2 as fast and accurately as possible) and even in comprehension through lexical decision tasks (i.e., participants are asked to decide whether or not a chain of letters is a real word in a given language; see Lemhöfer, Spalek, & Schriefers, 2008).

The first study to examine the cross-linguistic GCE was Costa et al. (2003). The authors ran five L2 picture-naming experiments with highly proficient and native-like bilinguals, including different language pairs (i.e., Croatian-Italian, Spanish-Catalan, and Italian-French). Participants were asked to name pictures using an NP formed by a gender-marked definite determiner and a noun (except for the third experiment, in which an adjective

following the noun was also included). In their first three experiments, they tested Croatian-Italian bilinguals. Thus, pictures were named using Italian (L2) nouns whose translation could be homogeneric (e.g., '*mela*' [feminine.] in Italian is '*jabuka*' [feminine] in Croatian, 'apple' in English) or heterogeneric (e.g., '*pomodoro*' [masculine] in Italian is '*rajcica*' [feminine] in Croatian, 'tomato' in English). In their fourth and fifth experiments, they tested whether the dissimilarity between gender systems modulated cross-linguistic effects. Thus, they replicated their first experiments with native-like Spanish-Catalan and Italian-French bilinguals, two language pairs that are far more similar than the ones featured in their previous experiments (e.g., whereas Italian, Spanish, Catalan, and French have feminine and masculine gender values, Croatian has a third: the neuter gender value). In all the experiments, monolingual speakers of L2s (Italian, Catalan, or French) were recruited as control groups. The results revealed that whereas no gender effects of GC were found with Croatian-Italian speakers, both the Spanish-Catalan and Italian-French bilinguals showed the effect. However, since the effect was also found with the monolingual control groups, its interpretation as evidence in support of a gender-integrated system was not possible (note that monolinguals should not show the GC effect because they did not activate any translation).

Costa and colleagues explained their findings as either evidence against the integration of both gender systems or as support for the idea that grammatical gender is a feature that cannot be experimentally assessed, since it is not subject to competition. More specifically, if a unique gender system is shared across languages, their experiments should have shown gender effects according to activation-dependant models such as WEAVER++ or IN. First, because according to them the selection of a gender value is achieved by the build-up of activation, which allows competition to take place, and second, because NPs were required to name the pictures and so agreement had to be fulfilled. Conversely, null effects were indeed expected according to an automatic gender-access perspective by which the

selection of a noun's lexical node entails the automatic availability of the appropriate gender value for processing (e.g., Caramazza et al., 2001; Schiller and Caramazza, 2003). According to this perspective, gender effects are a consequence of competition when the determiner is selected. Because Croatian has no definite articles, the GC effect was absent.

In any case, although Costa et al. (2003) considered their findings to be possible evidence for an autonomous view of bilingual gender representation or for an automatic selection of grammatical gender, many criticisms have been raised that have weakened such interpretations. On the one hand, some variables were not properly controlled, such as the phonological gender transparency of the L2 targets (i.e., the number of target nouns that had the typical gender ending letters) or the cognateness status of the translations (i.e., to what extent L1 translations were not only semantically but also phonologically and orthographically similar to the target L2 nouns; for example, 'lemon': *limón* [Spanish] and *llimona* [Catalonian] vs. 'bed': *cama* [Spanish] and *llit* [Catalonian]). On the other hand, the number of participants was too small in the first three experiments (10 per experiment), resulting in a possible lack of statistical power. Besides, the results might have been influenced by the considerable differences between the gender systems of Italian and Croatian, as well as by the native-like L2 proficiency of participants, a point we will return to below (see Lemhöfer et al., 2008, and Sá-Leite, Fraga, & Comesaña, 2019, for an overview).

Several authors have replicated Costa et al.'s (2003) study by improving the experimental control and using different sets of stimuli, different pairs of languages, and late-learner bilinguals of multiple degrees of proficiency. Although a GCE has been systematically observed in the majority of naming studies (Bordag, 2004; Bordag & Pechmann, 2007; Klassen, 2016; Morales, Paolieri, & Bajo, 2011; Lemhöfer et al., 2008; Manolescu & Jarema, 2015; Paolieri et al., 2010) and also forward translation tasks (from L1 to L2: Manolescu & Jarema, 2015; Paolieri et al., 2010; Paolieri, Padilla, Koreneva, Morales,

& Macizo, 2019; Salamoura & Williams, 2007), there are likewise studies in which the effect has been harder to observe. On the one hand, significant differences between congruent and incongruent conditions were absent in the analysis of variance by items in many cases (e.g., Klassen, 2016; Lemhöfer et al., 2008; Paolieri et al., 2010, 2019). On the other hand, there is unless an interesting and well conducted study which completely failed to obtain the effect (Bordag and Pechmann; 2008). The authors conducted three forward translation experiments with Czech-German bilinguals of upper-intermediate to advanced proficiency. In their first two experiments, a condition featuring short (BNs) and long (NP: adjective SMALL or BIG + noun) responses was included along with the GC conditions (gender-congruent and gender-incongruent translations). In the second experiment, a new set of materials was used in which phonological gender transparency was manipulated as a third condition (1/3 of words were transparent, 1/3 opaque, and 1/3 irregular). No GCE was obtained in either of these experiments. However, in the second experiment transparent nouns were faster and more accurately translated than opaque and irregular nouns, although only in the long condition (NPs). The third experiment was conducted because in both the long and short conditions participants only had to translate the noun (i.e., the adjective they had to produce was determined by the size of a "dot" that appeared in front of the word). Consequently, the authors wanted to increase the probability of L1 gender retrieval including BNs but also complex NPs to translate (gender marked adjective + noun). Again, no effects were obtained except for the effect of phonological gender transparency for NPs in RTs. The authors interpreted this as evidence that gender in an L2 is computed each time it is needed through different pieces of relevant information, in this case through the phonological form of the word. We will assess this finding in more detail in the Discussion. See Table 1 section for more detail on these studies.

< INSERT TABLE 1 >

Relevant factors for bilingual gender processing

As stated in a theoretical review by Sá-Leite, Fraga, & Comesaña (2019), the cross-linguistic GCE might be under the influence of certain relevant variables that may be increasing its slipperiness. Taking into consideration the bilingual studies reported above, the following factors should be highlighted: (1) the naming instruction (BNs vs. NPs), (2) the phonological gender transparency of the target language, (3) the L2 proficiency, and (4) the type of task.

A careful review of the literature led us to believe that the first two points are closely related. Assuming that gender access is a competitive process in which gender nodes are involved, it may be the case that during the production of an NP, grammatical gender is indeed activated to a higher degree as a function of the agreement status between the definite article and the noun, in comparison to the production of a BN. Since it has to fulfil its main function of agreement, it is expected that gender will be activated to a higher level than when agreement is not necessary. The higher the activation of grammatical gender features, the higher the competition between translations and, as a consequence, the size of the GCE. Indeed, in the third experiment of Lemhöfer et al.'s study (2008) featuring a naming task with L2 Dutch participants, a tendency was observed by which the GCE was greater for NPs than for BNs. Similarly, Salamoura and Williams (2007), in a forward-translation task with L2 German participants, failed to obtain the effect with BNs, but obtained it with NPs. In this sense, the variable "phrase type" might play a primary role in gender processing. Furthermore, the phonological gender transparency of the target language may influence this role. More specifically, the level of activation may also be higher the higher the phonologically gender transparency of a language, especially when it comes to BNs. Although insufficient studies are currently available to establish a well-grounded continuum from the most to the least transparent languages, it seems clear that in general, Romance

languages would fall on the more transparent side and Germanic languages would be closer to the opaque side (the picture is not so clear for Slavic languages, and further research is needed). For instance, in Italian, a number of studies have indicated that more than 90% of nouns ending in '-o' and '-a' are masculine and feminine, respectively (e.g., De Martino et al., 2017; Harris, 1991). In Spanish, almost 100% of nouns ending in '-o' are masculine and 96% of nouns ending in '-a' are feminine (Harris, 1991). In European Portuguese, 77% of nouns ending in '-o' are masculine, while approximately 95% of nouns ending in '-a' are feminine, as can be seen in the Procura-PALavras lexical database (P-PAL, Soares et al., 2014). Importantly, these simple phonological cues ('-o' and '-a') are present in all kinds of inanimate nouns and also coincide with the morphological rules used for natural gender[1] (e.g., in Spanish, '*chico*' [boy] and '*chica*' [girl]). One Romance language that would probably fall nearer to the opaque side of the continuum is French, since phonological gender cues here are far less straightforward than in the case of the Romance languages described above (see Karmiloff-Smith, 1979). Indeed, due to specific phonological regularities, it has been systematically shown that native-speaking children of Italian, Portuguese, and Spanish rely heavily on the phonological form of nouns ('-o' and '-a') to acquire grammatical gender (e.g., Bates, Devescovi, Pizzamiglio, D'Amico, & Hernandez, 1995; Corrêa, Augusto, & Castro, 2011; Pérez-Pereira, 1991). As expected, recent evidence is not so clear for French children (e.g., Boloh, Escudier, Royer, & Ibernon, 2011; Boloh & Ibernon, 2010; Boloh & Ibernon, 2013; Carroll, 1999). Transparency, thus seems to have a primary role in gender acquisition, which probably does not remove but partially diminishes the role of determiners

---

[1] Natural gender specifies not only the grammatical gender of a noun but also the biological sex of the referent. It is semantic-related and its location and processing in the lexical system differ from that of the abstract grammatical gender.

and adjective inflections in gender awareness and assignation during childhood (De Martino et al., 2017). Hence, in general, nouns themselves prompt gender activation, and so it is highly probable that gender processing is more frequent in Romance languages because encounters with NPs (and morphological gender inflections in other kind of words such as adjectives) are not the only core mechanism for gender acquisition. However, some Germanic languages have been shown to be quite opaque. For instance, in German there are regularities on which gender assignation is based, yet there are many such regularities and these are restricted largely to derived nouns, building a complex system that does not seem to facilitate gender acquisition and processing in the same way that simple and extended regularities like nominal endings '-a' and '-o' do. The rules in fact involve forty-four semantic, phonological, pseudosuffix and suffix gender regularities on German nouns (Köpcke, 1982; Köpcke & Zubin, 1983; Zubin & Köpcke, 1984). Dutch is even more opaque than German, in that regularities are almost non-existent, especially for non-derived nouns (Unsworth, 2013). Interestingly, phonological gender transparency seems to have an impact on the age of acquisition of grammatical gender, since in the Romance languages mentioned above (but excluding French) it occurs at between 2 and 4 years of age (Kupisch, Müller, & Cantone, 2002; Pérez-Pereira, 1991) whereas in German and Dutch it occurs between 4 and 6 years (Jansen, 2009; Van der Velde, 2003). Because of this, German and Dutch children rely strongly on the articles that precede and agree with nouns (most frequently, definite articles) as a means of being aware and acquiring grammatical gender (see Arnon & Ramscar, 2012). As actually shown in the "monolingual" literature, in which the classical GCE is only obtained with NPs in the case of Germanic languages' speakers, we propose that this may contribute to a tendency for retrieving gender every time a definite article is present (see Sá-Leite, Tomaz, Hernández-Cabrera, Fraga, & Comesaña, 2019). Because BNs have not usually been the main element used to retrieve a grammatical gender value, they might not

activate gender to the same extent as a NP does, and thus they would not elicit GCE (see La Heij et al., 1998). Consequently, although a theoretical consideration of the papers cited here show that the cross-linguistic GCE seems to be obtained regardless of the phrase type used to name the picture (see Table 1), we must at least contemplate the possibility of a general tendency in the direction of a greater GCE effect for bilinguals of Romance languages than for bilinguals of the Germanic ones when BNs are required.

Regarding the possible modulations of the GCE due to participants' L2 proficiency, it seems that the higher the L2 proficiency, the shorter the effect. Thus, for instance, Bordag and Pechmann (2007) explain a decrease in the size of the GCE in their third experiment as a function of the higher L2 proficiency of the participants. Strikingly, in the first three experiments from Costa et al.'s (2003) study, conducted with native-like bilinguals (who were also probably simultaneous bilinguals), no GCEs were found. Sá-Leite et al. (2019) address this issue through a consideration of several theoretical perspectives (e.g., Dell, Chang, & Griffin 1999; MacWhinney, 1997). The proposal that most satisfyingly fits the results from all the reviewed studies is the developmental Bilingual Interactive Activation model (BIA-d, Grainger, Midgley, & Holcomb, 2010). The BIA-d is an extension of the connectionist but also localist BIA model (Grainger & Dijkstra, 1992) that aims to explain L2 lexical acquisition and processing. According to this model, both languages of a bilingual are always active to some degree, at least when it comes to late-learners. Thus, representations from the non-target language will be active along with the representations of the target language, and thus interaction between both languages will occur. However, the influence of the L1 (non-target language) on the L2 (target language) will be higher when the proficiency of the bilingual is lower, since during L2 acquisition direct form links between the L1 and their L2 translations are generated. Therefore, bilinguals will rely on the lemma features of the L1 to access L2 lexical entries until they are proficient enough to create their own L2

independent lemmas. Similarly, according to the most recent connectionist model of bilingual language production and comprehension, the Multilink model (Dijkstra et al., 2018), L2 words representations in our lexicon have frequency-dependent resting levels of activation. More specifically, proficiency is closely related to the time of usage of a language, measured through both the age of acquisition of the L2 and its active use (e.g., daily use). Hence, different levels of L2 proficiency are related to different levels of frequency of L2 use and exposition to L2 vocabulary. Although no direct allusions are made to grammatical gender in the Multilink model, following its rationale, it is plausible to think that the strength of a link between an L2 noun and its grammatical gender value depends on the frequency of use of this noun along with the retrieval of its own gender value.

Finally, the claim that there might be differences between tasks (naming vs. translation) was first made in Bordag and Pechmann (2008). They pointed out that task requirements might have been responsible for their null results. According to their explanation, the time course of the activation of the gender features in L1 and L2 in translation and picture-naming tasks differs. In naming tasks, the activation spreads in parallel, from the concept in common to both L1 and L2 lemmas, to the level of grammatical encoding. Therefore, at the same time as the L1 gender nodes are activated, those of L2 are activated too, and may compete for selection. In forward translation tasks, however, it is the L1 word form and its lemmas that are activated first; the activation then spreads to the lemma of the L2 translation equivalent (or to the concept, and later to the L2 lemma). Consequently, the L2 gender node becomes active after the activation of the L1 gender node. In any case, the GCE has been successfully observed in other studies featuring forward translation tasks, such as Manolescu and Jarema's (2015) second experiment, Paolieri et al.'s (2010) third experiment, and Paolieri et al. (2019). Hence, the notion of the task being a modulatory variable of the GCE effect is highly debatable.

Purpose of the study

Taking into consideration the literature reviewed thus far, many essential questions about the nature of grammatical gender representation and the extent to which the cross-linguistic GCE is reliable remain unclear. Indeed, in spite of the null results observed in the Costa et al.´s study (2003), other recent studies have found evidence for the cross-linguistic GCE, giving support to the idea of a gender-integrated system (e.g., Klassen, 2016, Paolieri et al., 2010). However, the effect in these studies has sometimes not reached statistical significance in analyses of variance by items. Besides, there are many variables, such as task requirements (using BNs or NPs), the phonological gender transparency of the target language, L2 proficiency, and the type of task (naming or translation), that might modulate the effect. Importantly, observations made thus far in research on the GCE with bilinguals have been defined according to a qualitative and theoretical analysis and not to a quantitative one. This was precisely the aim of the present study, i.e., to examine the robustness of the GCE and to analyse quantitatively the role of those variables in the stability of the effect. This would shed light on still unsolved questions as to the nature of grammatical gender representation, as well as the organization of the bilingual grammatical gender system.

Considering the findings of the reviewed studies, we expected the following results: (1) a small to moderately sized GCE that supports a gender-integrated perspective; (2) greater GCEs for bilinguals of transparent L2s in comparison to bilinguals of opaque L2s, especially for BNs; (3) a decrease in the size of the GCE as L2 proficiency rises, and (4) a greater GCE for naming tasks in comparison to translation tasks.

Method

Literature Search

A literature search was conducted on ERIC (Educational Resources Information Center), LLBA (Linguistics and Language Behaviour Abstracts), Psychology Database, PsycINFO, and Google Scholar. Dissertations were not considered. The search included the keywords 'gender congruency' and 'bilingualism' and yielded 362 results. After removing duplicates with RefWorks®, we individually screened all the articles and applied the following criteria for inclusion: (1) papers are on language production, in that the GC effect in comprehension is restricted to only three papers (Lemhöfer et al., 2008 – Experiment 1; Morales et al., 2016; Weber & Paris, 2004) and the cognitive processes involved in production and comprehension, although closely linked, are distinct (see Meyer, 2016); (2) studies explore the influence of the L1 on the L2 and not vice-versa – this because the study of how the acquisition of an L2 modifies the processing of the L1 (L2-L1) is based on different mechanisms and principles from those that analyse the influence in the opposite direction (L1-L2, see for instance, Lim & Christianson, 2013); (3) because many studies have found cross-linguistic interaction between an L3 and an L2 during language processing (e.g., Fung & Murphy, 2016; Williams & Hammarberg, 1998), the presence of any L3 might modulate the congruency status of the target, and thus the papers reviewed here test participants that are bilingual speakers of two gendered languages, with no information reported about a third language.

After the application of the inclusion criteria, ten papers were retained as relevant (Bordag & Pechmann, 2007; Bordag & Pechmann, 2008; Costa et al., 2003; Klassen, 2016; Lemhöfer et al., 2008; Manolescu & Jarema, 2015; Morales et al., 2011; Paolieri et al., 2010; Paolieri et al., 2019; Salamoura & Williams, 2007). Citations from these studies were

inspected as an additional search, and thus one further study was included (Bordag, 2004). In total, 11 papers were carefully analysed. They most frequently featured participants that were late-learners of an L2 (except Costa et al., 2003), thus the conclusions here established must be restricted to bilingual populations that acquired their L2 after approximately the age of 10. Twenty-five experiments were considered and classified according to four different relevant variables, as a means of ensuring comparability in the analysis of the GC effect. First, we considered the naming instruction given to the participants (BNs vs. NPs). Second, we considered the phonological gender transparency of the target language (the L2). As specified in the previous section, we classified Italian and Spanish as transparent, and Dutch, French, and German as opaque. Czech was also classified as opaque, because although we can find some gender cues in this language, they are not as simple as '-o' and '-a' for both grammatical and natural gender. For example, inanimate nouns ending in a consonant tend to be masculine (but they can be feminine if their final consonant is soft or ambivalent), those ending in '-a' and '-e' tend to be feminine, and those ending in '-o' or '-í' are neuter. However, masculine animate nouns with natural gender can also end in '-a'. In addition, there are not only soft and ambivalent consonants to consider (see Naughton, 2006) but also seven cases in which the endings of nouns vary. We are aware that this classification is far from precise, and indeed we will return to this issue in the Discussion section. Finally, we divided the experiments according to the L2 proficiency of participants as reported by the authors (intermediate, intermediate to high proficient, high proficient to native-like) and according to the task (naming vs. forward translation tasks). In Table 1 we identify and describe these experiments.

Meta-analytic Approach

From the 25 experiments, we identified 41 comparisons of interest between gender-congruent and gender-incongruent trials. Hedges' *g* was computed as measure of effect size

for each comparison for which enough statistical information was available (see Section 1 of the Supplemental Materials for technical details on how *g* was calculated). When published information was not enough, we contacted the authors and requested the original data[2]. When we received no response, or when the original data from the authors were not available, we made a variety of different decisions, these described in the Supplemental Materials. Thus, in total we obtained 35 effect sizes from 22 experiments reported in 10 articles. We computed the effect size from the response time of gender incongruent minus gender congruent cases, that is, we obtained positive *g*'s when response was slower to gender-incongruent trials and negative *g*'s when the response was slower to gender-congruent trials.

Of note, effect sizes in a meta-analysis have to be independent, but several experiments provided two estimates from the same sample, one for BNs and another for NPs, and thus some of the effect sizes were correlated. To test the effect of the non-independence of the observations in our results, we conducted the meta-analyses twice, first with a standard approach that considers effect sizes as independent, and then with the robust variance estimation method, developed to take into account the correlation between observations (Hedges, Tipton, & Johnson, 2010). A brief description of how we computed both meta-analyses is presented in Section 2 of the Supplemental Materials. The two methods yielded similar results, with only trivial differences that did not affect interpretation. Thus, we report here the results of the meta-analyses with independent observations, and for completeness we also provide a full report of the results with the correlated observations approach, in Section 3 of the Supplemental Materials.

---

[2] We thank Albert Costa[†], Kristin Lemhöfer, and Daniela Paolieri for sharing their data. † deceased

Results

All the analyses were done in R (R Team, 2018). A series of meta-analyses to test the effect of gender congruency on response time and whether it was moderated by the proposed variables were conducted. For the full sample, the effect of gender congruency was small but significant, $g = 0.24$, $SE = 0.04$, 95% CI [0.16, 0.32], $z = 6.24$, $p < .001$ (see Figure 3). Heterogeneity was high, $Q_E(34) = 190.27$, $p < .001$, showing a large variability between effect sizes. Several moderators were tested to try to explain this variability, but none showed any difference.

< INSERT FIGURE 5 >

First, we tested the effect of the naming instructions, BNs or NPs. Two experiments from Manolescu and Jarema (2015) were lost from our analysis because they reported the results of both conditions in a collapsed form (see the Supplemental Materials for details). Results showed no differences in the GC effect between BN, $g = 0.23$, $SE = 0.05$, 95% CI [0.13, 0.33] and NP, $g = 0.21$, $SE = 0.05$, 95% CI [0.11, 0.31], $Q_M(1) = 0.10$, $p = .75$. We also tested the effect of language transparency depending on whether the L2 of the participants was transparent or opaque. We found no differences between transparent languages, $g = 0.20$, $SE = 0.06$, 95% CI [0.08, 0.32], and opaque languages, $g = 0.27$, $SE = 0.05$, 95% CI [0.17, 0.36], $Q_M(1) = 0.66$, $p = .42$

Linguistic proficiency did not affect the results either, $Q_M (2) = 5.04$, $p = .08$, although effect sizes were numerically larger with the lowest proficiency level: for intermediate, $g = 0.43$, $SE = 0.11$, 95% CI [0.21, 0.64], for intermediate to high proficiency, $g = 0.17$, $SE = 0.05$, 95% CI [0.07, 0.27], and for high proficiency to native-like, $g = 0.28$, $SE = 0.06$, 95% CI [0.16, 0.40]. Finally, with respect to the task that participants completed, naming or translation, no significant differences were found between naming, $g = 0.28$, $SE = 0.05$, 95%

CI [0.18, 0.38], and translation, $g = 0.19$, *SE* $= 0.06$, 95% *CI* [0.08, 0.31], $Q_M (1) = 1.19$, $p =$

.28.

Discussion

The main aim of the present meta-analysis was to examine the nature of grammatical gender representation and processing in bilinguals. To this end, we examined the robustness of the cross-linguistic GCE and we analysed quantitatively the role of certain variables observed in the literature as apparently relevant in the stability of this effect. As expected, our meta-analysis provided systematic evidence of the cross-linguistic GCE. The effect, though, is small ($g = 0.24$) and shows considerable variability ($Q_E(34) = 190.27$, $p < .001$), which may explain why it is sometimes hard to observe and also why statistical significance is not reached in the analyses of variance by items in many cases (e.g., Klassen, 2016; Lemhöfer et al., 2008; Paolieri et al., 2010, 2019). Surprisingly, this variability does not seem to be consequence of the variables that we included in the analysis. The results, hence, go against the automatic gender-selection perspective (Caramazza et al., 2001) and support a competitive view between abstract grammatical gender. This because the GCE does exist and is found to be significant, even in the condition featuring BNs. As the variable "naming instructions" (BNs vs. NPs) did not interact with the gender congruency conditions (congruent vs. incongruent), we can conclude that it is found regardless of the production of a BN or an NP in bilinguals. In other words, gender effects are indeed observable through experimental assessment and, contrary to the main tenets of the WEAVER++ and IN models (Caramazza & Miozzo, 1997; Levelt et al., 1999), an agreement context is not mandatory to recover grammatical gender during bilingual lexical access. Thus, competition between different gender values at the lemma selection of BNs does indeed occur. In any case, this competition is not led by similarity, as the Double-Selection Model (Cubelli et al., 2005) suggests. The gender effects reported here were of congruency and not incongruency, even for studies with Italian or Spanish as L2. Having established this, the existence of a cross-linguistic GCE supports an integrated view of the bilingual gender system, at least regarding

bilinguals who are late-learners of an L2. It seems, then, that both gender systems are merged

into one and that the abstract gender values of both languages interact during the selection of

L2 nouns (see Sá-Leite et al., 2019 for more detail).

In this sense, contrary to our predictions, not only the naming instruction (BN vs. NP)

but also the phonological gender transparency of the L2 and the interaction between both

variables failed to modulate the effect. The peculiarities of L2 acquisition could be behind

this result. More specifically, such peculiarities could effectively be changing the

mechanisms by which grammatical gender is normally retrieved and processed when only the

linguistic system comprises just one language. This is the case because, in comparison to

monolingual speakers of opaque languages, gender might be a more salient characteristic in

bilingual populations, being directly associated with nouns in opaque L2s. In particular,

learning an L2 usually involves doing so in a classroom setting (which is definitely true for

our participants, who generally learned their L2s at school or in language academies). In

relation to this issue, it has been shown that language lessons tend to emphasize single

vocabulary items and their characteristics (Doughty & Williams, 1998). Thus, with the

exception of simultaneous bilinguals (i.e., those who acquire both languages at the same time,

having two L1s rather than an L1 and an L2), when people learn an L2 not only will their

focus be on the single vocabulary units themselves, but they will also be given a direct

explanation about gender categories. Then, a conscious effort is made to associate single

nouns to gender, despite the articles that precede them (see Arnon & Ramscar, 2012). This

means that it is not surprising that the GCE is found with opaque languages in bilinguals

producing BNs, in that they do not need to rely naturally on definite articles to be aware of

gender values and acquire grammatical gender. Consequently, what would be expected

because of the nature of the languages at issue (i.e., the GCE to be absent or weaker when

bilinguals of opaque language/s produce BNs) is in fact not observed. Thus, if bilinguals of

opaque L2s are similarly attentive to nouns as native speakers of transparent languages are, they must be especially sensitive to phonological gender cues in the L2. In this sense, although Bordag and Pechmann (2008) failed to obtain the GCE with bilinguals of L2 German in their forward translation tasks, they did find effects related to phonological gender transparency. More specifically, they found a faster retrieval of L2 nouns with a gender transparent ending, which indicates a computation of the gender value in which gender transparency is playing a key role. This means that (1) bilinguals are sensitive to those phonological cues in their L2; (2) gender is somehow being processed, but perhaps the degree of activation is not enough to reflect this competition through a GCE; and (3) there is a need to reconsider the interaction between the lexical and phonological stratums, looking in greater depth at similar proposals such as those of Cubelli et al. (2005) and Duràn and Pillon (2011). Unfortunately, the phonological gender transparency of L2 nouns was only included in the experimental design of Bordag and Pechmann's studies (2007, 2008) and, hence, it should be taken into account in future research.

A consideration of the results of the reviewed studies from a qualitative perspective suggests that the higher the proficiency the lower the size of the GCE (e.g., Costa et al.'s [2003] null results have been explained through nativelikeness). Modulations on the interaction between languages caused by proficiency are observed widely in bilingual studies (e.g., Grainger et al., 2010; Jackson, 2008). Following the tenets of the Multilink model (Dijkstra et al., 2018), it is plausible to think that modulations on the GCE due to L2 proficiency would be expected in one way or another as a consequence of variations in the frequency of the use of nouns and their gender values. This would affect the strength of the links between nouns and their grammatical characteristics, and thus would have an impact on the size of the GCE. The absence of modulations of L2 proficiency in the GCE might be related to the fact that studies differed a great deal in terms of how they described

proficiency. In fact, this was especially obvious when it came to age of acquisition, as well as the years that participants had spent studying the L2 or the time they had been actively using the L2, variables that directly influence the frequency of use of nouns, and hence "proficiency" as it is conceived of in the Multilink model. For example, whereas in Bordag and Pechmann's study (2008) intermediate to high-proficiency participants were characterised as those who started to learn the L2 after the age of 9 (maximum at the age of 14), with an average of 10 years of acquisition, in Paolieri et al.'s (2010) study, high-proficiency participants were described as those who had started to learn the L2 after the age of 18 and who had been using it for 3.3 years on average. Many seemingly contradictory definitions of high-proficiency in different studies can be found (e.g., Paolieri et al., 2019 vs. Salamoura & Williams, 2007). Also, there is an overall lack of objective tests to verify the L2 proficiency of bilingual participants; the self-reported tests seen in many studies also differed in terms of the scale used (e.g., in Paolieri et al. 2010, a 10-point scale; Lemhöfer's et al., 2008, a 7-point scale). In sum, we should not rule out the possibility of proficiency influencing the GCE, as an imprecise classification of participants in terms of their comprehension and production skills in the L2 might explain the absence of significant interactions between the GCE and this variable.

Finally, despite a qualitative appraisal of the literature leading to the tentative suggestion that the type of task (forward translation task vs. naming task) affected the GCE, our analysis failed to confirm this. A tentative explanation of these results can be made based on the variability found in the size of the effect. We might bear in mind here that Bordag and Pechmann (2008) were the first to suggest this. In their study, the GCE was absent in three forward-translation experiments and they explained these nulls results in terms of differences in the time-course of gender selection between naming and translation tasks, as we noted in the Introduction. However, other researchers have indeed found GCEs using forward-

translation tasks (e.g., Manolescu & Jarema, 2015; Paolieri et al., 2010, 2019; Salamoura & Williams, 2007). Thus, the absence of GCEs in Bordag and Pechmann's experiments might simply be the consequence of the slippery nature of the effect, which has a small size and a high variability that leads to difficulties in its observation (at least, behavioural observation), even in naming tasks (e.g., Costa et al., 2003; Klassen, 2016; Lemhöfer et al., 2008; Paolieri et al., 2010).

We recognise, however, there are variables other than those considered in the present study which may be modulating the GCE, such as the state of cognateness between translations (see Lemhöfer et al., 2008) or the concreteness of the target nouns (see Paolieri et al., 2019). Further work is required here. Also, studies looking in greater detail at the phonological gender transparency of languages would make the development of a well-grounded continuum possible, thus allowing for a more accurate characterisation (and hence division) of languages. Indeed, in the present study we lacked sufficient evidence on the proportion and characteristics of transparent nouns in languages such as French and Czech, and thus an unequivocal categorization was not possible. Although we opted to classify them as opaque, we think of transparency more as a continuum, and hence systematic research on the phonological gender transparency of languages would be of great help here.

Conclusions

This meta-analysis has presented systematically collected evidence of a small-sized and highly-variable cross-linguistic GCE, which confirms its existence and also explains the difficulties in observing it. Evidence analysed here supports an integrated view of bilingual grammatical gender representation. In contrast to the tenets of WEAVER++ and the IN models (Caramazza & Miozzo, 1997; Levelt et al., 1999), the GCE is significant even when the target noun is not embedded in an agreement context, and thus abstract gender values are being selected and competing with each other during the lexical access of BNs in bilinguals. This competition, however, is not led by similarity as the Double-Selection model by Cubelli et al. (2005) suggests. No differences in the effect are obtained as a result of different naming instructions (BN vs. NP), phonological gender transparency of the L2, or task type (naming vs translation). The absence of the moderation of L2 proficiency in the GCE might not be reliable, in that this factor was extremely heterogeneous and sometimes imprecisely measured across studies. A better control of L2 proficiency and objective L2 testing is recommended. We encourage subsequent work on the GCE to focus on longitudinal studies or on studies involving different levels of L2 proficiency. Finally, because of the high variability and small size of the GCE observed in behavioural studies, electrophysiological techniques such as event related potentials might usefully be used.

Table 1 Description of the experiments conducted on the GC effect with bilingual population

	Task	Participants (n)	L1	L2	L1-L2 Linguistic Family	L2 Transparency	L2 Age of Acquisition (AoA)	L2 Proficiency	GC Effect (BNs)	GC Effect (NPs)
Costa et al. (2003) – E1, E2, E3	N	10, 10, 10	Croatian	Italian	NR-R	T	Early learners	3	-	Null
Costa et al. (2003) – E4	N	48	Spanish	Catalan	R-R	T	Early learners	3	-	Found
Costa et al. (2003) – E5	N	10	Italian	French	R-R	O	Early learners	3	-	Found
Bordag (2004) – E1	N	18	German	Czech	NR-NR	O	20yo	1	Found	Found
Bordag (2004) – E2	N	18	Czech	German	NR-NR	O	14yo	2	Found	Found
Bordag & Pechmann (2007) – E1	N	18	Czech	German	NR-NR	O	13.2yo	2	Found	Found
Bordag & Pechmann (2007) – E2	N	18	Czech	German	NR-NR	O	12.4yo	2	Found	Found
Bordag & Pechmann (2007) – E3	N	18	Czech	German	NR-NR	O	10.7yo	2	Found	Found
Lemhöfer et al. (2008) – E2	N	16	German	Dutch	NR-NR	O	Late learners	3	Found(-)	Foundd(-)
Lemhöfer et al. (2008) – E3	N	22	German	Dutch	NR-NR	O	Late learners	1	Found	Found
Paolieri et al. (2010) – E1	N	12	Italian	Spanish	R-R	T	21.8yo	2	Found	Found
Paolieri et al. (2010) – E2	N	12	Italian	Spanish	R-R	T	22.2yo	2	Found(-)	-
Morales et al. (2011) – E1	N	32	Italian	Spanish	R-R	T	Late learners	2	Found	-
Manolescu & Jarema (2015) – E1	N	27	Romanian	French	R-R	O	Early learners	3	Found	Found
Klassen (2016)	N	17	Spanish	German	R-NR	O	Late learners	1	Found(-)	KFound(-)
Salamoura & Williams (2007)	T	18	Greek	German	NR-NR	O	10yo	3	Null	Found
Bordag & Pechmann (2008) – E1	T	18	Czech	German	NR-NR	O	12.4yo	2	Null	Null
Bordag & Pechmann (2008) – E2	T	18	Czech	German	NR-NR	O	11.8yo	2	Null	Null
Bordag & Pechmann (2008) – E3	T	15	Czech	German	NR-NR	O	12.7yo	2	Null	Null
Paolieri et al. (2010) – E3	T	12	Italian	Spanish	R-R	T	24.7yo	2	Found	Found
Manolescu & Jarema (2015) – E3	T	27	Romanian	French	R-R	O	Early learners	3	Found	Found
Paolieri et al. (2019) – E1	T	54	Russian	Spanish	NR-R	T	22yo	3	Found(-)	Found(-)
Paolieri et al. (2019) – E2	T	32	Italian	Spanish	R-R	T	28.7yo	3	Found	Found

Note. Task: N = Picture naming task, T = Forward translation task; L1-L2 Linguistic Family: R = Romance language; NR = Non-Romance

language; L2 Transparency = T = Transparent; O = Opaque; L2 Age of Acquisition: yo = years old; When age is not specified the authors did not

inform of the exact age of acquisition; L2 Proficiency: 1 = Low-Intermediate to upper intermediate; 2 = Intermediate to high proficient, 3 =

Advanced/high proficient to nativelike; BNs = Bare nouns; NPs = Noun phrases; GC Effect (BNs) and GC Effect (NPs): Null = GC effect not

obtained, Found = GC Effect obtained, Found(-) = GC effect obtained only in the analysis of variance by subjects, - = condition not tested.

Figure 1. Lexical network proposed by the WEAVER++ model consisting of nodes for lexical concepts, syntactic words and their properties, and morphemes, phonemes, and syllables. Grammatical Gender would be located at the Lemma stratum, along with other grammatical properties such as "lexical category" in the form of a node per grammatical gender value. Figure taken from Levelt, Roelofs, and Meyer (1999).

Figure 2. Lexical network proposed by the Independent Network model. The scheme represents the Italian words '*tavolo*' (table), '*sedia*' (chair), and '*tigre*' (tiger). Information flows in parallel from the semantic to the lexeme and syntactic networks, and then onto segmental information. Information also flows from the lexeme to the syntactic network. The figure is simplified and the O-Lexeme network is not presented here. Dotted lines represent weaker connections. Links within a network are inhibitory. Grammatical gender is located at the syntactic network. Since Italian is depicted here, two gender nodes are included: M for masculine and F for feminine. Other abbreviations are N, noun; V, verb; Adj, adjective; Cn, count noun, and Ms, mass noun. Figure taken from Caramazza and Miozzo (1997).

Figure 3. Representation of the gender-integrated hypothesis for words of the same gender across languages (panel A) and those of different gender (panel B). '*Jabuka*' and '*mela*' mean "apple" in Croatian and Italian, respectively. '*Rajcica*' and 'p*omodoro*' mean "tomato". Gender features (feminine [fem] and masculine [masc]) are shared across languages. Figure reproduced from Costa, Kovacic, Franck, and Caramazza (2003) and taken from Sá-Leite, Fraga, and Comesaña (2019).

Figure 4. Representation of the gender-autonomous hypothesis for words of the same gender across languages (panel A) and those of different gender (panel B). '*Jabuka*' and '*mela*' mean "apple" in Croatian and Italian, respectively. '*Rajcica*' and '*pomodoro*' mean "tomato". Gender features (feminine [fem] and masculine [masc]) are independent for both languages.

Figure reproduced from Costa, Kovacic, Franck, and Caramazza (2003) and taken from Sá-Leite, Fraga, and Comesaña (2019).

Figure 5. Forest plot showing the effect sizes of gender congruency on response times for all the studies and the meta-analytic estimate. BN = Bare Nouns; NP = Noun Phrases.

References

References marked with an asterisk indicate studies included in the meta-analysis.

Arnon, I., & Ramscar, M. (2012). Granularity and the acquisition of grammatical gender: How order-of-acquisition affects what gets learned. *Cognition*, *122*, 292–305. doi:10.1016/j.cognition.2011.10.009

Bates, E., Devescovi, A., Pizzamiglio, L., D'Amico, S., & Hernandez, A. (1995). Gender and lexical access in Italian. *Perception & Psychophysics*, *57*, 847–847. doi:10.3758/BF03206800

Bock, K. (1995). Producing Agreement. *Current Directions in Psychological Science*, *4*, 56–61. doi:10.1111/1467-8721.ep10771165

Boloh, Y., Escudier, F., Royer, S., & Ibernon, L. (2011). Determiners and noun endings in French children's gender attribution choices. *First Language*, *32*, 398–410. doi:10.1177/0142723711422621

Boloh, Y., & Ibernon, L. (2010). Gender attribution and gender agreement in 4- to 10-year-old French children. *Cognitive Development*, *25*, 1–25. doi:10.1016/j.cogdev.2009.09.011

Boloh, Y., & Ibernon, L. (2013). Natural gender, phonological cues and the default grammatical gender in French children. *First Language*, *33*, 449–468. doi:10.1177/0142723713499848

Bordag, D. (2004). Interaction of L1 and L2 systems at the level of grammatical encoding: Evidence from picture naming. *EUROSLA Yearbook*, *4*, 203–230. doi:10.1075/eurosla.4.10int

* Bordag, D., & Pechmann, T. (2007). Factors influencing L2 gender processing. *Bilingualism: Language and Cognition*, *10*, 299-314. doi:10.1017/s1366728907003082

* Bordag, D., & Pechmann, T. (2008). Grammatical gender in translation. *Second Language Research*, *24*, 139–166. doi:10.1177/0267658307086299

Caramazza, A. (1997). How Many Levels of Processing Are There in Lexical Access? *Cognitive Neuropsychology*, *14*, 177-208. doi:10.1080/026432997381664

Caramazza, A., & Miozzo, M. (1997). The relation between syntactic and phonological knowledge in lexical access: evidence from the 'tip-of-the-tongue' phenomenon. *Cognition*, *64*, 309-343. doi:10.1016/s0010-0277(97)00031-0

Caramazza, A., Miozzo, M., Costa, A., Schiller, N., Alario, F-X. (2001). Lexical selection: A cross-language investigation of determiner production. In E. Dupoux (ed.), *Language, brain, and cognitive development: Essays in honor of Jacques Mehler*, pp. 209–226. Cambridge, MA: MIT Press.

Carroll, S. E. (1999). Input and SLA: Adults' Sensitivity to Different Sorts of Cues to French Gender. *Language Learning*, *49*, 37–92. doi:10.1111/1467-9922.00070

Comesaña, M., Bertin, P., Oliveira, H., Soares, A. P., Hernández-Cabrera, J. A., & Casalis, S. (2018). The impact of cognateness of word bases and suffixes on morpho-orthographic processing: A masked priming study with intermediate and high-proficiency Portuguese-English bilinguals. *PLOS One*, *13*, e0193480. doi:10.1371/journal.pone.0193480

Comesaña, M., Ferré, P., Romero, J., Guasch, M., Soares, A. P., & García-Chico, T. (2015). Facilitative effect of cognate words vanishes when reducing the orthographic overlap: The role of stimuli list composition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *41*, 614–635. doi:10.1037/xlm0000065

Corbett, G. G. (1991). *Gender*. Cambridge University Press. doi:10.1017/cbo9781139166119

Corrêa, L. M. S., Augusto, M. R. A., & Castro, A. (2011). Agreement and markedness in the ascription of gender to novel animate nouns by children acquiring Portuguese. *Journal of Portuguese Linguistics*, *10*, 121–142. doi:10.5334/jpl.103

Costa, A., Kovacic, D., Fedorenko, E., & Caramazza, A. (2003). The Gender Congruency Effect and the Selection of Freestanding and Bound Morphemes: Evidence From Croatian. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *29*, 1270–1282. doi:10.1037/0278-7393.29.6.1270

* Costa, A., Kovacic, D., Franck, J., & Caramazza, A. (2003). On the autonomy of the grammatical gender systems of the two languages of a bilingual. *Bilingualism: Language and Cognition*, *6*, 181–200. doi:10.1017/s1366728903001123

Cubelli, R., Lotto, L., Paolieri, D., Girelli, M., & Job, R. (2005). Grammatical gender is selected in bare noun production: Evidence from the picture–word interference paradigm. *Journal of Memory and Language*, *53*, 42-59. doi:10.1016/j.jml.2005.02.007

De Martino, M., Bracco, G., Postiglione, F., & Laudanna, A. (2017). The influence of grammatical gender and suffix transparency in processing Italian written nouns. *The Mental Lexicon*, *12*, 107-128. doi:10.1075/ml.12.1.05dem

Dell, G. S., Chang, F., & Griffin, Z. M. (1999). Connectionist models of language production: Lexical access and grammatical encoding. *Cognitive Science*, *23*, 517–542. doi:10.1207/s15516709cog2304_6

Dijkstra, T., Miwa, K., Brummelhuis, B., Sappelli, M., & Baayen, H. (2010). How cross-language similarity and task demands affect cognate recognition. *Journal of Memory and Language*, *62*, 284–301. doi:10.1016/j.jml.2009.12.003

Dijkstra, T., Wahl, A., Buytenhuijs, F., van Halem, N., Al-Jibouri, Z., de Korte, M., & Rekké, S. (2018). Multilink: a computational model for bilingual word recognition and word translation. *Bilingualism: Language and Cognition*. doi:10.1017/S1366728918000287

Doughty, C., & Williams, J. (1998*). Focus on form in classroom second language acquisition*. Cambridge University Press.

Duràn, C., & Pillon, A. (2011). The role of grammatical category information in spoken word retrieval. *Frontiers in Psychology*, *2*. doi:10.3389/fpsyg.2011.00338

Fung, K. T. D., & Murphy, V. A. (2016). Cross linguistic influence in adult L2/L3 learners: The case of French on English morphosyntax. *Journal of Education*, *3*, 6-15. doi:10.7603/s40742-015-0011-4

Grainger, J., & Dijkstra, T. (1992). On the representation and use of language information in bilinguals. *Advances in Psychology, 83*, 207–220. doi:10.1016/s0166-4115(08)61496-x

Grainger, J., Midgley, K., & Holcomb, P. J. (2010). Chapter 14. Re-thinking the bilingual interactive-activation model from a developmental perspective (BIA-d). *Language Acquisition Across Linguistic and Cognitive Systems*, 267–283. doi:10.1075/lald.52.18gra

Harris, J. W. (1991). The Exponence of Gender in Spanish. *Linguistic Inquiry*, *22*, 27-62.

Hedges, L. V., Tipton, E., & Johnson, M. C. (2010). Robust variance estimation in meta-regression with dependent effect size estimates. *Research Synthesis Methods*, *1*, 39-65. doi:10.1002/jrsm.5

Jackson, C. N. (2008). Proficiency Level and the Interaction of Lexical and Morphosyntactic Information During L2 Sentence Processing. *Language and Learning*, *58*, 875 – 909. doi:10.1111/j.1467-9922.2008.00481.x

Jansen, V. (2009). *Determiners in first language acquisition of monolingual English and German children – A corpus-based investigation* [unpublished master's thesis]. University of Wuppertal, Germany.

Karmiloff-Smith, A. (1979). *A Functional Approach to Child Language: A Study of Determiners and Reference*. Cambridge: Cambridge University Press.

* Klassen, R. (2016). The representation of asymmetric grammatical gender systems in the bilingual mental lexicon. *Probus*, *28*, 9-28. doi:10.1515/probus-2016-0002

Köpcke, K.-M. (1982). Untersuchungen zum Genussystem der deutschen Gegenwartssprache, *Linguistische Arbeiten*, *122*. doi:10.1515/9783111676562

Köpcke, K.-M., & Zubin, D. A. (1983). Die kognitive Organisation der Genuszuweisung zu den einsilbigen Nomen der deutschen Gegenwartssprache. *Zeitschrift für germanistische Linguistik*, *11*. doi:10.1515/zfgl.1983.11.2.166

Kroll, J. F., & Stewart, E. (1994). Category Interference in Translation and Picture Naming: Evidence for Asymmetric Connections Between Bilingual Memory Representations. *Journal of Memory and Language*, *33*, 149–174. doi:10.1006/jmla.1994.1008

Kroll, J. F., & Tokowicz, N. (2005). Models of bilingual representation and processing: Looking back and to the future. In J. F. Kroll & A. M. B. de Groot (Eds.), *Handbook of bilingualism: Psycholinguistic approaches* (pp. 531-553). New York, NY, US: Oxford University Press.

Kupisch, T., Müller, N., & Cantone, K. F. (2002). Gender in monolingual and bilingual first-language acquisition: Comparing Italian and French. *Lingue e Linguaggio, 1*, 107-150. doi:10.1418/7559

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

* Lemhöfer, K., Spalek, K., & Schriefers, H. (2008). Cross-language effects of grammatical gender in bilingual word recognition and production. *Journal of Memory and Language*, *59*, 312–330. doi:10.1016/j.jml.2008.06.005

Levelt, W. J. M. (2001). Spoken word production: A theory of lexical access. *Proceedings of the National Academy of Sciences of the United States of America*, *98*, 13464-13471. doi:10.1073/pnas.231459498

Levelt, W. J. M., Roelofs, A., & Meyer, A. S. (1999). A theory of lexical access in speech production. *Behavioral and Brain Sciences*, *22*, 1-75. doi:10.1017/S0140525X99001776

Lim, J. H. and Christianson, K. (2013). Integrating meaning and structure in L1–L2 and L2–L1 translations. *Second Language Research*, *2*, 233-256. doi:10.1177/0267658312462019

MacWhinney, B. (1997). Implicit and explicit processes. *Studies in Second Language Acquisition*, *19*, 277-281. doi:10.1017/s0272263197002076

* Manolescu, A., & Jarema, G. (2015). Grammatical gender in Romanian-French bilinguals. *The Mental Lexicon*, *10*, 390–412. doi:10.1075/ml.10.3.04man

Meyer, A. S. (2016) Same, different, or closely related: What is the relationship between language production and comprehension? *Journal of Memory and Language*, *89*, 1-7. doi:10.1016/j.jml.2016.03.002

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

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.

doi:10.1037/0278-7393.28.2.388

* Morales, L., Paolieri, D., & Bajo, T. (2011). Grammatical gender inhibition in bilinguals.

Frontiers in Psychology, *2*. doi:10.3389/fpsyg.2011.00284

Morales, L., Paolieri, D., Dussias, P. E., Valdés Kroff, J. R., Gerfen, C., & Bajo, M. T.

(2016). The gender congruency effect during bilingual spoken-word recognition.

Bilingualism: Language and Cognition, *19*, 294–310.

doi:10.1017/s1366728915000176

Naughton, J. (2005). *Czech: An Essential Grammar.* New York: Routledge: Taylor & Francis

Group.

* Paolieri, D., Cubelli, R., Macizo, P., Bajo, T., Lotto, L., & Job, R. (2010). Grammatical

gender processing in Italian and Spanish bilinguals, *The Quarterly Journal of*

Experimental Psychology, *63*, 1631-1645. doi:10.1080/17470210903511210

Paloeri, D., Lotto, L., Leoncini, D., Cubelli, R., & Job, R. (2011). Differential effects of

grammatical gender and gender inflection in bare noun production. *The British*

Psychological Society, *102*, 19-36.doi: 10.1342/000712610X496536

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*, 335-

347.doi:10.1080/09541440902916803

* Paolieri, D., Padilla, F., Koreneva, O., Morales, L., & Macizo, P. (2019). Gender

congruency effects in Russian–Spanish and Italian–Spanish bilinguals: The role of

language proximity and concreteness of words. *Bilingualism: Language and*

Cognition, *22*, 112–129. doi:10.1017/s1366728917000591

Pérez-Pereira, M. (1991). The acquisition of gender: What Spanish children tell us. *Journal of Child Language*, 18, 571–590. doi:10.1017/s0305000900011259

R Core Team (2018). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. URL: https://www.R-project.org/

Sá-Leite, A. R., Fraga, I., & Comesaña, M. (2019). Grammatical Gender Processing in Bilinguals: An Analytic Review. *Psychonomic Bulletin & Review*. doi:10.3758/s13423-019-01596-8

Sá-Leite, A. R., Oliveira, H., Soares, A. P., Carreiras, M., & Comesaña, M. (2017). *The processing of gender grammar during bilingual word recognition: Tackling the role of cross-language similarities*. Poster presented at the 5th Barcelona Summer School on Bilingualism and Multilingualism. Barcelona, Spain.

Sá-Leite, A. R., Tomaz, Â., Hernández-Cabrera, J. A., Fraga, I., & Comesaña, M. (2019). *The GAP Hypothesis: An Integrative View of Gender Acquisition and Processing*. Poster presented at the 60th Annual Meeting of the Psychonomic Society: Canada: Montreal.

* Salamoura, A., & Williams, J. N. (2007). The representation of grammatical gender in the bilingual lexicon: Evidence from Greek and German. *Bilingualism: Language and Cognition*, *10*, 257-275. doi:10.1017/s1366728907003069

Schiller, N. O., & Caramazza, A. (2003). Grammatical feature selection in noun phrase production: Evidence from German and Dutch. *Journal of Memory and Language*, *48*, 169–194. doi: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*, 841–850. doi:10.1037/0278-7393.19.4.841

Schriefers, H., & Jescheniak, J. D. (1999). Representation and Processing of Grammatical Gender in Language Production: A Review. *Journal of Psycholinguistic Research*, *28*, 575-600. doi:10.1023/A:1023264810403

Schriefers, H., Jescheniak, J. D., & Hantsch, A. (2002). Determiner selection in noun phrase production. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *28*, 941–950. doi:10.1037/0278-7393.28.5.941

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*, 1368–1377. doi:10.1037/0278-7393.26.6.1368

Soares, A. P., Iriarte, A., Almeida, J. J., Simões, A., Costa, A., França, P., Machado, J., & Comesaña, M. (2014). Procura-PALavras (P-PAL): Uma nova medida de frequência lexical do Português Europeu contemporâneo [Procura-PALavras (P-PAL): A new measure of word frequency for contemporary European Portuguese]. *Psicologia: Reflexão e Crítica, 27*, 1-14.

Soares, A. P., Oliveira, H., Comesaña, M., & Costa, A. (2018). Lexico-syntactic interactions in the resolution of second language relative clause ambiguities: The role of cognate status and second language proficiency. *Psicológica*, *39*, 164-197. doi:10.2478/psicolj-2018-0008

Soares, A. P., Oliveira, H., Ferreira, M., Comesaña, M., Macedo, F., Ferré, P., Acuña-Fariña, J. C., Hernández-Cabrera, J., & Fraga, I. (2019). The role of cognate status and second language proficiency on relative clause attachment preferences in a second language: An eye-tracker study. *PLOS One*.

Starreveld, P., & La Heij, W. (2004). Phonological facilitation of grammatical gender retrieval. *Language and Cognitive Processes*, *19*, 677–711. doi:10.1080/01690960444000061

Unsworth, S. (2013). Assessing the role of current and cumulative exposure in simultaneous bilingual acquisition: The case of Dutch gender. *Bilingualism: Language and Cognition*, *16*, 86-110. doi:10.1017/s1366728912000284

Van Berkum, J. J. (1997). Syntactic processes in speech production: the retrieval of grammatical gender. *Cognition*, *64*, 115–152. doi:10.1016/s0010-0277(97)00026-7

Van der Velde, M. (2003). *Déterminants et pronoms en néerlandais et en français: syntaxe et acquisition* [Unpublished doctoral dissertation], University of París 8, France. Available online at http://www.theses.fr/2003PA082332 (May, 2017)

Weber, A., & Paris, G. (2004). The origin of the linguistic gender effect in spoken-word recognition: Evidence from non-native listening. *Proceedings of the Twenty-Sixth Annual Meeting of the Cognitive Science Society*, 1446–1451.

Williams, S., & Hammarberg, B. (2005). Language switches in L3 production: Implications for a polyglot speaking model. *Processes in Third Language Acquisition*, *19*, 28–70. doi:10.3366/edinburgh/9780748635115.003.0003

Zubin, D. A., & Köpcke, K.-M. (1984). Affect classification in the German gender system. *Lingua*, *63*, 41–96. doi:10.1016/0024-3841(84)90031-7

Acknowledgements