

Psychonomic Bulletin and Review

Grammatical Gender Processing in Bilinguals: An Analytic Review

Ana Rita Sá-Leite¹, Isabel Fraga¹, & Montserrat Comesaña^{2,3}

¹Cognitive Processes & Behaviour Research Group, Department of Social Psychology, Basic Psychology & Methodology, University of Santiago de Compostela, Spain

²Centro de Enseñanza Superior Cardenal Cisneros, Universidad Complutense de Madrid, Spain

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

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@rai.usc.es

Abstract

In this review paper, we analyse how grammatical gender is represented and processed in the bilingual mind. To that end, we review the data from 13 existing behavioural studies with mainly late second language (L2) learners on the so-called gender congruency (GC) effect (a facilitated processing for translation equivalents with the same gender in comparison to those with different gender) in L2 production and comprehension. The majority of results shows a GC effect, regardless of the type of language involved. However, the state of cognateness of the target nouns, as well as the similarity between the gender systems of bilingual speakers and their L2 proficiency, modulates the results. Interestingly, a gender agreement context is not required in order to observe gender effects, in that they are also observed with bare nouns. Overall, the findings support an integrative view of bilingual gender representation, with competitive and inhibitory processes at different levels of language processing underlying cross-language GC effects.

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

Introduction

The representation and processing of grammatical gender in bilinguals has been examined in recent decades as a means of better understanding the cross-linguistic interactions that may take place at the lemma level during language production and comprehension. Lemma and lexeme retrieval compound the two major stages often described for lexical access (Kempen & Hoenkamp, 1987). Whereas the lexeme level is where the recovery of orthographic and phonological information takes place, lemma retrieval consists of the selection of semantic information, as well as of other grammatical features that define words, such as case, gender and number. The interplay of these grammatical features during language processing is an important focus of research (e.g., Domínguez, Cuetos, & Seguí, 1999). Yet, each of these features is conceptually different and has its own idiosyncrasies (see for instance Barber & Carreiras, 2005; Harris, 1991; for experimental evidence on differences between number and gender selection during language processing, see also De Vicenzi, 1999; Igoa, García-Albea, & Sánchez-Casas, 1999). Hence, it is necessary to try and understand their representational characteristics separately. In the case of grammatical gender, this is particularly important because, as Corbett (1991:1) noted, it is “the most puzzling of the grammatical categories”.

The complexity of grammatical gender is visible in its short but rich definition: it is described as an arbitrary, abstract property of nouns, being both lexical and syntactic, that is present in every language that has a gender system (Schriefers & Jescheniak, 1999). A gender system classifies words according to different categories or classes called gender values (Corbett, 1991). The number and name of these gender values differs depending on the language itself, ranging from the two classes usually featured in Romance languages (i.e., masculine and feminine) to seventeen or more in languages such as Wolof or Nigerian Fula (e.g., Babou & Loporcaro, 2016; Breedveld, 1995; for more details on other languages see

Corbett, 1991). As a formal feature of the lexicon, grammatical gender is inherent to nouns in the sense that it categorizes them according to an irreplaceable gender value. For instance, the European Portuguese noun *mesa* ('table' in English) is feminine and cannot be transformed into the masculine. Of note, grammatical gender is described as an arbitrary abstract characteristic because it is independent of its meaning (Lyons, 1968). In some cases, however, an exception must be made. In the majority of gendered languages there are some semantic rules, these often based on general conceptual splits such as male/female, human/non-human, or animate/inanimate, to classify nouns according to gender (Audring, 2016). Specifically, it is very common for languages to link gender to the semantic information of words through the dichotomy male/female. Hence, gender here is no longer irreplaceable or abstract, in that it matches the sex of the referent (male or female) and should thus be called “natural gender” or “semantic gender” rather than “grammatical gender” (Tight, 2006). In fact, only nouns of natural gender change morphologically according to the sex of the referent (e.g., in European Portuguese, *menino* –masculine- and *menina* –feminine- [boy and girl, respectively], see Corrêa, Augusto, & Castro, 2011 and Harris, 1999, for details). Thus, in general, with the exception of those animate nouns with natural gender, nouns have no conceptual basis for the distinction of grammatical gender (Barber & Carreiras, 2005). In addition, grammatical gender is not related to the morphological form of nouns but can be partially related to their phonological form. More specifically, in some languages labelled as transparent (e.g., Romance ones), we find statistical-phonological regularities in the distribution of the nominal endings in function of their gender value (e.g., in European Portuguese, Italian, or Spanish, nouns ending in “-a” tend to be feminine; Bates, Devescovi, Pizzamiglio, D’Amico, & Hernández, 1995). These regularities are cues that facilitate gender acquisition and processing, but they must not be interpreted as rules since there are many exceptions (e.g., ambiguous or opaque endings exist for both gender values,

such as “-e” in European Portuguese, being *torre* [tower] feminine and *corte* [cut] masculine, or irregular endings such as the “-o” in a feminine word like *tribo* [tribe] and the “-a” in a masculine word like *programa* [program]; see Pérez-Pereira, 1991). Finally, the main function of grammatical gender is to establish agreement between the different elements of speech (e.g., in European Portuguese, *Um copo* [masculine] *pequeno* [‘A small cup’] but *Uma casa* [feminine] *pequena* [‘A small house’]). Consequently, the status of gender-irreplaceability does not apply for words such as articles and adjectives, among others. Words of this kind have to agree with the gender value of a main noun, for which they have to adopt every gender value in a particular language. To achieve this, articles and adjectives change morphologically through gender inflections (i.e., in the previous example, “*um*” [masculine indefinite article, ‘a’], but “*um-a*” [feminine indefinite article, ‘a’]). As a result, grammatical gender is primarily a lexical property, but since it determines the form of other words during syntactic assembling, it is also defined as a syntactic property. Because its syntactic implications are essential in language comprehension and production, agreement relationships between the different elements of speech have received special attention in literature (see for example, Wicha, Moreno, & Kutas, 2004), particularly in comparison to the study of grammatical gender in its most lexical dimension.

In the present review, we are interested in the studies that have been carried out thus far on grammatical gender as an abstract lexical property to determine how the gender systems of a bilingual interact (or do not interact) during the lexical access of second language (L2) nouns. From a theoretical point of view, as can be seen in Figure 1 and Figure 2, two perspectives are commonly contrasted in bilingual gender representation: the integrative vs. the autonomous view. According to the integrative proposal (Klassen, 2016), the representation of grammatical gender values is shared across languages in a unique integrated gender system; therefore, the gender activation of words in a given language

influences the gender activation of words in the other language. Conversely, the autonomous proposal (Costa et al., 2003) claims that each language has its own specific gender system, and thus the fact that two translations either have or do not have the same gender value is irrelevant for the organization of L2 grammatical knowledge. That is, since gender values are not shared across languages, the gender of the target word would not receive any activation from the gender of its translation equivalent.

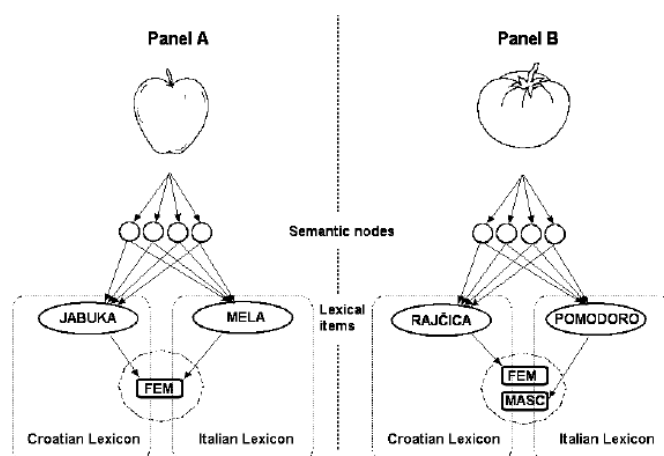


Figure 1. Representation of the gender integrated hypothesis for words of the same gender across languages (panel A) and those of different gender (panel B). *Jabuka* and *mela* mean “apple” in Croatian and Italian, respectively. *Rajčica* and *pomodoro* mean “tomato”. Gender features (feminine [fem] and masculine [masc]) are shared across languages. Figure taken from Costa, Kovacic, Frank, and Caramazza (2003).

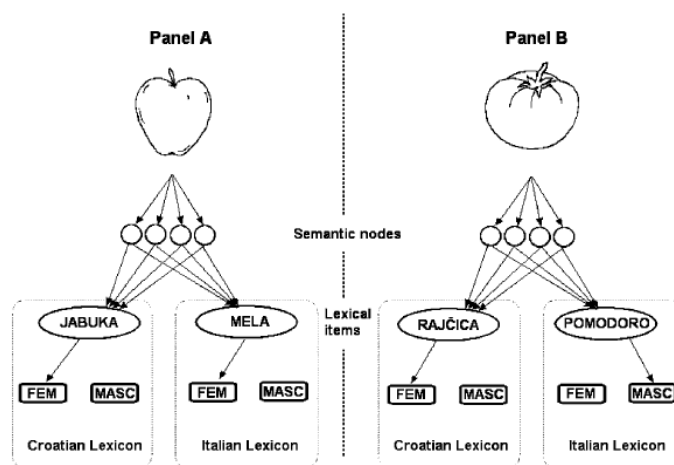


Figure 2. Representation of the gender autonomous hypothesis for words of the same gender across languages (panel A) and those of different gender (panel B). *Jabuka* and *mela* mean “apple” in Croatian and Italian, respectively. *Rajcica* and *pomodoro* mean “tomato”. Gender features (feminine [fem] and masculine [masc]) are independent for both languages. Figure taken from Costa, Kovacic, Frank, and Caramazza (2003).

Theoretical basis for bilingual grammatical gender representation

As cross-linguistic interactions have been consistently observed at the lemma level during semantic processing, as well as at the lexeme level during phonological and orthographical encoding (e.g., Comesaña et al., 2015, 2018; Costa and Caramazza, 1999), it would be reasonable to expect an interaction to occur during bilingual L2 gender processing. Indeed, in light of findings here, the most influential models of bilingual language processing predict cross-linguistic interactions at every level of lexical access. For example, the Bilingual Interactive Activation models are connectionist yet localist in nature, presenting four layers of nodes for written word recognition (features, letters, words and language nodes; BIA model, Dijkstra, van Heuven, & Grainger, 1998; BIA+ model, Dijkstra & van Heuven, 2002; see Figure 3 - A for a representation of the BIA+ model) and production (Multilink, Dijkstra et al., 2018). See the Figure 3.

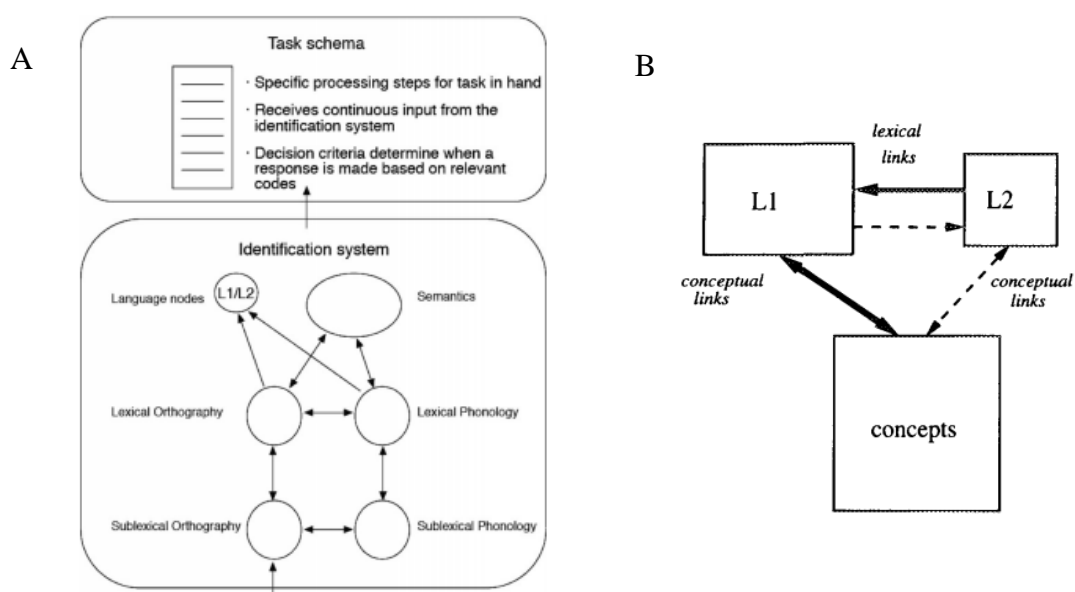


Figure 3. The BIA+ model (A). Arrows indicate activation flowing from different representational levels. Inhibitory connections within those levels are graphically omitted.

Figure taken from Dijkstra and van Heuven (2002). The RHM model (B). Dashed lines represent weaker links between levels. Figure taken from Kroll and Stewart (1994).

Regarding the bilingual interactive activation models, at the level of the languages' nodes, there is one different node for each of the languages of the bilingual (L1 and L2). Importantly, when we recognize letters, those letters send activation up to the word level. According to these models, words from both languages are simultaneously activated as they are fully connected to each other. Hence, the authors of these models argue for the existence of an integrated lexicon with a non-selective lexical access. Thus, for instance, according to the network architecture of the Multilink model of word production, the level of semantic would activate the levels of orthography and phonology (i.e., words phonemes and letters) through bidirectional links, regardless of the language that is being spoken. In the BIA and BIA+ models, the language nodes, influenced by many factors (specially the linguistic context), would have to inhibit the words from the other language so production is made in the language that is intended. Another classical model of bilingual language production, the Revised Hierarchical Model (RHM, Kroll & Stewart 1994; see Figure 3 - B for its representation), does not support an integrated lexicon for both languages, but it does predict direct influence between both languages at the level of lexical representation. Specifically, it proposes a unidirectional link between the concept (lemma) and the word form levels. At the level of word form, two separate lexicons (L1 and L2) would be influencing each other (these influences are assumed to be stronger in the forward direction [from L1 to L2] than in the inverse direction [from L2 to L1]).

Although interesting, neither of the aforementioned models have explicitly addressed the issue of how grammatical gender is represented and processed in the bilingual mind. In

fact, in neither of them is the location of grammatical features clearly identified. Since grammatical features are part of the lemma level, along with semantics, they might be included in the semantic/conceptual level for bilingual interactive activation models (indeed Mulder, Dijkstra, and Baayen, 2015 consider that grammatical morpheme effects such as those of morphological family size in bilinguals are mainly driven by semantic factors) as well as for the RHM. Thus, in sum, we can only speculate as to what might occur during abstract grammatical gender selection in bilinguals. Indeed, although there is strong evidence in favour of an integrated lexicon and non-selective lexical access (see Van Heuven, Dijkstra & Grainger, 1998; see also Comesaña et al., 2018), interference at the point of gender selection during the processing of nouns might not occur, this mainly because there are many differences between what gender itself is, depending on the language. This highlights the importance of being cautious when considering grammatical gender to be the same grammatical feature for the two languages of a bilingual person (Costa et al., 2003).

Particularly:

- It is possible that the structure of the gender systems of the two languages differ. Thus, the number of gender values may vary across languages.
- It is also possible that the correlation between the phonology of a word's nominal ending and its gender value might vary across languages. Thus, some languages classify nouns depending on their phonological gender transparency (e.g., Romance languages with “-a” being the most common nominal ending for feminine) whereas others do not follow such a classification (e.g., Dutch). The absence of these phonological cues hampers and slows down the acquisition and processing of grammatical gender (see Gathercole & Thomas, 2005; see also Unsworth, 2013).
- It is possible that grammatical and semantic gender coincide for one language but not for the other. For instance, in languages in which gender values are ‘masculine’ and

‘feminine’, such as Romance languages, nouns that refer to living entities have natural gender related to biological sex. This does not occur in languages where grammatical gender values are ‘common’ and ‘neuter’, such as in Dutch (“man” and “woman” in Dutch, *man* and *vrouw*, are both of common gender, where as in a Romance language such as European Portuguese, *homem* is masculine and *mulher* is feminine) and as a result, semantic mediation does not exist in the same way during gender acquisition and processing.

- It is possible that the morphological gender inflections that occur for agreement extend to different classes of words for each language (e.g., contrary to Romance Languages, Slavic languages tend to mark verb tenses by gender).
- It is possible that in one language (e.g., Spanish) the gender value of a given word has consequences for the selection of definite/indefinite determiners (which are essential for the acquisition of grammatical gender, e.g., Arnon & Ramscar, 2012), while in other languages (e.g., Croatian) such consequences are not present at all.

Therefore, considering grammatical gender as the same parameter in two languages is an assumption that might vary depending on the similarity of their respective gender systems. This similarity may determine the degree of integration of these gender systems and, hence, the influence that the non-target language has on gender processing in the target language. Such an idea is supported by many developmental studies, which have observed that the higher the similarity between L1 and L2 gender systems, the easier the process of gender acquisition in the L2 (Cornips & Hulk, 2006; Schwartz et al., 2015).

It seems clear that the characteristics of the gender systems of a bilingual might play an important role in the observation of cross-linguistic interferences during grammatical gender processing. Nevertheless, other variables must be borne in mind in order to have an accurate picture here. In fact, it has been shown that cross-linguistic interactions at the level

of semantic processing and phonological/orthographical encoding are subject to other variables, such as those related to the characterization of the bilingual population itself (e.g., age of L2 acquisition and L2 proficiency) and those related to word form (e.g., the cognateness status and the phonological transparency of nouns). Regarding the age of L2 acquisition (i.e., the age at which bilinguals started to learn their L2) and their L2 proficiency, it has been observed that: a) the later that acquisition takes place, the higher the influence of the L1 on the L2 and the higher the number and type of errors during L2 language processing, especially in terms of grammatical encoding (e.g., Bley-Vroman, 1990; Keating, 2016); and that b) L1 influence on L2 processing decreases as L2 proficiency increases (e.g., Elston-Güttler, Paulmann, & Kotz, 2005).

Regarding word form, two main variables must be considered. First, it has been observed that the cognate status of translation equivalents affects L2 processing, with cognate words (those that, besides meaning, share form; e.g., *mensagem* – message, in European Portuguese [EP] and English, respectively) facilitating lexical selection in comparison to non-cognates (words that share only meaning; *caneta* – pen in EP and English, respectively; see for instance Degani, Prior, & Hajajra, 2018; see also Comesaña et al., 2015 for inhibitory effects of cognate words as a function of cognate word type and task requirements).

Second, it is critical to consider the degree of phonological gender transparency of nouns. In transparent languages, we can distinguish three different lexical categories as a function of their degree of phonological gender transparency: a) Transparent nouns, which end in the letters that are mostly associated with their gender, such as the EP masculine noun *casaco* [jacket] that ends in *-o*, or the feminine noun *camisa* (shirt) that ends in *-a* [shirt]; b) opaque nouns, which end in other vowels or consonants that are not related mostly to any gender, such as the EP masculine noun *chocolate* (chocolate), or the EP feminine noun *rede* (net); and c) irregular nouns, which show inverse cues to those for transparent nouns, such as

the EP masculine noun *drama* (drama). Recently, studies with behavioural, electrophysiological, and/or neuroimaging techniques have shown that these categories influence the way that nouns are processed (e.g., Caffarra, Janssen, & Barber, 2014; Fuchs, Polinsky, & Scontras, 2015; Hernandez et al., 2004; Padovani et al., 2005; Urrutia, Domínguez, & Álvarez, 2009). Specifically, these studies showed that because phonological gender transparency affects the selection of the appropriate gender node at the grammatical level, there is an interaction between phonological and grammatical encoding. For instance, transparent nouns such as *casaco* seem to be processed faster and/or more accurately than opaque/irregular nouns like *chocolate* or *drama* (Bates et al., 1995; Bates, Devescovi, Hernandez, & Pizzamiglio, 1996; Caffarra et al., 2014; Gollan & Frost, 2001). This means that the selection of grammatical features (i.e., gender) does not have to be completed before the beginning of the phonological encoding, and a flow of information from the lower level of phonological encoding to the higher level of grammatical encoding is taking place. Such an idea supports interactive models of language production, this is, those that argue for the existence of bidirectional connections between different processing levels, such as connectionist ones (e.g., BIA, BIA+, and Multilink models for bilingual processing). On these lines, some authors have questioned whether bilinguals are sensitive to phonological gender cues in their L2s, and have wondered if gender processing in L2s is more serial in nature than in L1s.

In sum, cross-linguistic interactions at multiple levels of language processing have been consistently found in the literature. The majority of the models on both language production and comprehension in bilinguals account for these interactions (e.g., the BIA model and its successive updates). Yet grammatical gender is a very peculiar feature, and might function according to its own principles. Bearing in mind that the characteristics of gender can vary widely from one language to another, it is not clear whether or not the

concept of grammatical gender is the same for the two languages within the bilingual linguistic system. This has direct implications for the possible integration of the two gender systems in the bilingual mind. In this sense, even if the bilingual lexicon is an integrated one with non-selective access, the processing of gender for one language might not influence the processing of gender for the other language. Thus, modulations on the cross-linguistic interactions during the grammatical gender selection of nouns should be expected depending on the similarity of the gender systems. These modulations must be examined taking into account other factors related to the bilingual population itself and the characteristics of word form, with special attention to the degree of gender-phonological transparency of the target nouns.

Experimental assessment of grammatical gender processing in bilinguals

The most frequently used task to examine grammatical gender selection during noun processing in bilinguals is the naming task. Participants are asked to describe pictures using L2 nouns the gender of which can be congruent or incongruent with their L1 translation equivalents. Therefore, heterogeneric nouns (i.e., nouns that have one gender value in one language and the opposite one in the other language) are used along with homogeneric nouns (i.e., nouns that have the same gender in both languages) to create the conditions of gender congruence and incongruence between languages. The observation of a Gender-Congruency effect –GC- (response times are shorter when the L1 translations are gender-congruent with the L2 target nouns than when they are gender-incongruent, $GC < GI$) would point to an interaction between languages during grammatical gender selection in the L2. It should be noted that the GC effect is measured as a difference of response times, not of error rates, since errors usually indicate misnaming, or, if a gender value is incorrectly retrieved during the task, point to an instability of the gender representation due to an ongoing (or

unsuccessful) process of gender acquisition for that word (Lemhöfer, Spalek, & Schriefers, 2008). Thus, the GC effect here will be understood mainly as a timing effect.

When experimentally assessing grammatical gender processing in bilinguals, one major question must be addressed due to its theoretical implications in the bilingual as well as in the monolingual domain of word production: are agreement contexts necessary to retrieve grammatical gender when processing a noun? Indeed, the most influential model of monolingual language production (WEAVER++ model, Levelt, Roelofs, & Meyer, 1999) understands grammatical gender as a syntactic rather than as a lexical parameter. More specifically, this model proposes that there are three main levels (from the top to the bottom: [1] semantics, [2] lexical and syntactic representation –lemma- and [3] phonological form -lexeme-). The lemma stratum mediates between conceptual and phonological lexical information and includes the gender nodes, which represent the gender values of a language and are linked to the correspondent nouns of each gender. Gender selection takes place when these nodes reach a certain level of activation. Thus, it is a competitive process, but critically, the activation of gender nodes only occurs if grammatical gender is fulfilling its main function of agreement. For this to happen, the noun has to be embedded in a sentence or noun phrase (NP), where other words have to agree with it in gender. From these claims, it follows that looking at gender with bare nouns (BN) is not possible. This tenet has been classically examined through the so-called Picture-Word Interference Paradigm. In this paradigm, monolingual participants have to name aloud a picture while ignoring a superimposed noun that may or may not coincide in gender with the noun that denotes the picture. Following the tenet of the WEAVER++ model, variations on the RTs that arise from the congruence/incongruence between the gender of the target and the distractor are expected with NPs but not with BNs. Many studies with Dutch and German native speakers support the WEAVER++ claims, showing GC effects with NPs, but not with BNs (e.g., La Heij,

Mark, Sander, & Willeboordsde, 1998). However, more recent studies featuring multiple languages such as Italian, Spanish, and EP, have indeed found gender effects with BNs, which is interesting since the usage of BNs allows research to avoid potential interference from agreement processes and determiner selection (e.g., Cubelli et al., 2005; Paolieri, Lotto, Leoncini, Cubelli, & Job, 2011; Sá-Leite, Oliveira, Soares, Carreiras, & Comesaña, 2017; see however Finocchiaro et al., 2011, for null results). The results with Romance languages seem to be more in line with connectionist models of monolingual language processing (e.g., the Connectionist Model of Speech Production of Dell, Chang, & Griffin, 1999). The CMSP proposes an interactive view of language processing formed by three levels (semantics, words, and phonemes). Grammatical gender would be a pure lexical feature related to words that is activated by means of a competitive process when lexical access occurs, regardless of the presence of an agreement context.

Recent theories have noted the possibility that differences in gender acquisition found between these linguistic families (Germanic vs. Romance) might explain the mixed results observed with the Picture-Word Interference Paradigm. Briefly, as Romance languages are gender-transparent languages, gender acquisition is done not only through definite articles but also and very especially through nouns themselves, thanks to their form (e.g., using the cue “feminine nouns usually end in -a”, for more detail see Pérez-Pereira, 1991). On the other hand, because Germanic and Slavic languages are very gender-opaque, children acquire gender mainly through articles (the most common element agreeing with the gender of the noun, see Arnon & Ramscar, 2012). Hence, grammatical gender is more of a lexical property for Romance languages but more of a syntactic one for Germanic languages (Sá-Leite, Tomaz, Hernández-Cabrera, Fraga, & Comesaña, in preparation). Consequently, gender effects are obtained with NPs but not with BNs for the latter. Thus, the principles proposed by the WEAVER++ model might in fact be correct for gender-opaque languages, whereas

other models, such as Dell, would be more appropriate for transparent languages. If so, the principles of the WEAVER++ and the CMSP models might not be universal. Although developed for monolingual data, these models assess grammatical gender processing more directly than the ones described for bilingual language processing and hence we can consider them when examining the bilingual population, with special care taken in terms of the degree of phonological transparency of the language pair under study.

Purpose of the study

In the present review, we aim to arrive at a better picture of how grammatical gender is represented and selected in the bilingual mind, not only during L2 noun production but also during comprehension. The examination of production and comprehension studies should allow us to establish the extent to which gender effects are task-sensitive.

Specifically, following connectionist models of bilingual language processing (e.g., BIA, BIA+, Multilink), we acknowledge the rationale that expects a gender-integrated system where cross-linguistic interactions occur during L2 grammatical gender selection. However, as grammatical gender is quite a peculiar feature, with its characteristics varying across languages, these cross-linguistic differences, along with the aforementioned variables related to individual linguistic characteristics and word form, might modulate or even determinate such interactions. It is worth noting here that the study of grammatical gender processing in bilinguals contributes to a long-term discussion about the requirement of an agreement context to observe gender effects, something that might also be affected by the linguistic family of the bilinguals' languages pair. This directly tests the universality of the tenets of certain models (WEAVER++, CMSP), highlighting the peculiarities of the grammatical features of each evaluated language.

Method

A systematic search was conducted of the databases ERIC (Educational Resources Information Center), LLBA (Linguistics and Language Behaviour Abstracts), Psychology Database, PsycINFO, and Google Scholar. Dissertations and patents were not considered. The search for the first four databases included the keywords ‘gender congruency’, ‘gender processing’ and ‘bilingualism’. To avoid an excessive number of duplicates, the search using Google Scholar included only the more precise terms ‘gender congruency’ and ‘bilingualism’. In total, 305 results were obtained. We individually screened all the articles and applied the following criteria for inclusion: (1) the paper focuses mainly on grammatical gender processing as an abstract lexical characteristic of nouns rather than on agreement relationships; (2) the paper explores the influence of the L1 on the L2 and not the other way around; that is, the study of how L2 acquisition modifies the processing of the L1 (L2-L1) analyses different mechanisms and is based in different principles to those used to study the influence in the inverse direction (L1-L2, see for instance, Lim & Christianson, 2012); (3) the paper tests participants that are bilingual speakers of two gendered languages, without informed influence from a third language, this because many studies have found cross-linguistic influence from the L3 in L2 processing (e.g., Fung & Murphy, 2016); and (4) the study uses behavioural techniques.

Following the application of the inclusion criteria, 12 studies were identified as relevant (Bordag, 2004; Bordag & Pechmann, 2007, 2008; Costa, Kovacic, Frank, & Caramazza, 2003; Klassen, 2016; Lemhöfer et al., 2008; Manolescu & Jarema, 2015; Morales, Paolieri, & Bajo, 2011; Morales et al., 2016; Paolieri et al., 2010; Paolieri, Padilla, Koreneva, Morales, & Macizo, 2018; Salamoura & Williams, 2007). Citations from these studies were also inspected, and as a result, one further study was selected (Weber & Paris, 2004). In total, 10 of these assess language production (mainly using picture-naming tasks,

but also forward-translation tasks), two assess language comprehension (lexical decision task and visual world paradigm), and one assesses both. All studies created the congruent and incongruent conditions through the use of homogeneric/heterogeneric target nouns. A detailed description of all the studies is given at the end of the manuscript in Table 1, with special attention to every variable that was controlled in each one. In the following pages, we will describe the literature, in two sections: studies on language production (with two subsections: studies that have used picture-naming tasks and those that have used forward translation-tasks), and studies on language comprehension. A brief summary of the results is given at the end of each section. Finally, a discussion with the main theoretical rationale and conclusions is presented.

Grammatical gender processing in bilingual production

Studies with picture-naming tasks

The first study to explore grammatical gender processing in bilinguals as an inherent characteristic of nouns was that of Costa et al. (2003). Until today, it is also the only one to find null results of gender using solely L2 picture-naming tasks. The authors ran five experiments with highly proficient and native-like bilinguals, including multiple language pairs. In their first three experiments, participants were Croatian-Italian bilinguals. Of note, Croatian and Italian have quite different and asymmetric gender systems. Croatian has three gender values: masculine, feminine and neuter (with the latter not being used in the first three experiments), and Italian has two gender values (masculine and feminine); moreover, they have different proportions of gender values, feminine being the less common gender in Italian but the more common in Croatian. As stated in the introduction about picture naming tasks in this line of research, pictures were named in the L2 (i.e., Italian). The participants used NPs (article + noun) whose translation could have either the same gender in the L1 ('same gender picture set' or homogeneric nouns, e.g., '*mela*' [fem.] in Italian is '*jabuka*' [fem.] in Croatian, 'apple' in English) or a different gender ('different gender picture set' or heterogeneric nouns, e.g., '*pomodoro*' [masc.] in Italian is '*rajčica*' [fem.] in Croatian, 'tomato' in English). The authors considered three possibilities regarding the results they could obtain. If the gender value of the nouns in the non-response language interacted with the gender value of the words in the response language, a difference on the response times between the two types of picture sets would be obtained. Conversely, if gender systems were autonomous, differences would not be found. A third possibility based in the Independent Network [IN] model (Caramazza & Miozzo, 1997) was also considered. This stated that the gender value of a target noun becomes automatically selected for processing when the noun's lexical node is selected. From this perspective, gender access is a direct consequence of lexical selection and

competition through activation is not considered. As competitive effects cannot be observed, gender selection cannot be experimentally assessed. Hence, the absence of differences between conditions could be positive evidence for either a gender-autonomous perspective or an automatic gender selection one.

Results revealed no differences in RTs or error rates between picture sets for any of the groups in any of the three first experiments. The null results in the third experiment are particularly interesting, as it consisted on a mixed-language picture-naming task. Within the mixed-language paradigm, the “language mode” of the participant is manipulated. “Language mode” is the name given to the state of activation of the bilingual’s languages a given point in time (Grosjean, 1997a, b) and it varies in function of the context depending on whether the bilingual is speaking to a monolingual or a bilingual person. When stimulation is given by two languages, instead of being in a ‘monolingual mode’ and hence ignoring the lexical activation of the non-response language, participants from the same population are forced to be in the so-called ‘bilingual mode’. Thus, in this case, by mixing up naming trials in L1 with trials in L2, the possible influence of the L1 on the L2 increases (for more details on ‘language modes’, see Grosjean, 1997a, b). Also, in this experiment, the authors included adjectives in the NPs, this because Croatian has no determiners and therefore no agreement relations were necessary in the L1 translations of the L2 Italian marked NPs in the first two experiments. The authors considered that, according to the classical models of language production (e.g., WEAVER++), this could be preventing gender from being activated in its first language. However, despite the direct co-activation of the L1 and the explicit requirement of an agreement context in both languages, no differences in naming times between picture sets were found. The authors interpreted this absence of cross-language effects in their first three experiments as a possible consequence of the dissimilarity between gender systems, as it has been described above (e.g., 3 Croatian vs. 2 Italian gender values).

Thus, in a fourth and fifth experiment, they replicated their first one with native-like Spanish-Catalan and Italian-French bilinguals, language pairs with very similar gender systems (structurally, morphologically and proportionally). In both experiments, control groups were included with native monolingual Catalan and French speakers, respectively. Although the results revealed differences between same vs. different gender sets in both groups of participants (i.e., faster responses for items from the same gender set), such differences were also observed in the control groups. The finding of GC effects in both types of population (mono vs. bilingual) was no further explored and the authors interpreted all their findings as evidence against the integrated-gender representation view or as favourable evidence to automatic-selection models (which hold that the observation of any gender effects is impossible, even if the grammatical gender systems of both languages are integrated).

Many criticisms have been made of this study, since multiple factors could have resulted in the failure to obtain gender effects. First, bilingual participants were native-like in both languages in all the experiments (Croatian-Italian; Spanish-Catalan; Italian-French), as shown in their extremely low error rates in comparison to the other reviewed studies (up to 10%). This might have led to more stable gender representations and to a reduction of gender assignment and agreement errors, as well as to a higher control of L1 activation. Therefore, instead of understanding the results as support for a gender-representation autonomous view, they should be interpreted in terms of the role of proficiency, age of acquisition, or way of L2 acquisition on L1 activation control in general. Similarly, results should not be interpreted as evidence for automatic gender-selection models (IN Model, Miozzo & Caramazza, 1998), this because the absence of evidence (absence of gender congruency effect) cannot be taken as the evidence for the absence (i.e., lack of the effect). Hence, the null effect observed in Costa et al. first three experiments do not necessarily support the automatic selection models. Besides, the fact that Croatian and Italian have quite different and asymmetric gender systems

could have been another factor promoting the decrease of interaction between languages. Croatian is a Slavic language and Italian is a Romance one, both have a different number of gender values and inverse proportions for each of them. In addition, the nominal ending ‘-o’ in Croatian is related to a neuter value, but in Italian it is the typical termination of masculine nouns. Importantly, morphological variations due to gender agreement occur between different classes of words in both languages. For example, in Croatian but not in Italian, verbs carry morphological inflections in agreement with the gender of the main noun. Articles do not generally vary by gender in Croatian but do so in Italian, and thus they do not play a principal role in gender acquisition and processing in the former but are fundamental in the latter. This amplifies the differences between both gender systems to a greater degree than in any other pair of languages that will be reviewed here after. Another potential caveat with this study is the fact that only 10 bilingual participants were recruited per experiment in their first three experiments, in which no effects were found for either bilinguals or monolinguals. This could have affected the statistical power of the effects. Furthermore, Costa et al. (2003) did not manipulate the cognate status of the nouns, collapsing both types of translations (cognates and non-cognates) into one analysis. In fact, their stimuli were mostly non-cognates. The inclusion of more cognate translations might have led to the appearance of a gender congruency effect, as has been observed in other studies (e.g., Lemhöfer et al., 2008). Another important factor is that, in comparison with other Germanic and Slavic languages, Romance languages such as Italian (tested in the first three experiments) have special properties that have produced controversial results when exploring GC effects with NPs, even in the monolingual literature. Thus, although gender effects have been consistently obtained with BNs in Italian and other Romance languages, the scenario is not the same with NPs (e.g., null effects in Alario & Caramazza, 2002). The use of BNs in Costa et al.’s (2003) study would have been desirable. Finally, since an effect of GC was obtained in their fourth

and fifth experiment for both bilingual and monolingual groups, it is possible that the experimental materials were not correctly selected (e.g., the number of phonological and orthographic neighbours of the nouns used to depict the pictures was not controlled and may have affected naming times; for more details, see Andrews, 1997). However, the authors did not further analyse any variable related to the stimuli and provided no further explanation of the shared effect between bilinguals and the monolingual control groups. Hence, it would be useful to replicate Costa et al.'s experiments with another set of materials in order to assess whether or not it is possible to talk about an autonomous-gender representation view with these language pairs.

Apart from Costa's et al. study, all successive works with picture-naming tasks have found gender effects (Bordag, 2004; Bordag & Pechmann, 2007; Klassen, 2016; Lemhöfer et al., 2008; Manolescu & Jarema, 2015; Morales et al., 2011; Paolieri et al., 2010) and thus support an integrative representation of grammatical gender. The effects are however, modulated by different variables. One of the main variables we have pointed out as fundamental is the similarity between gender systems. Studies that have used bilingual populations with quite similar gender systems have found clear GC effects. For instance, Paolieri et al. (2010) conducted two highly controlled picture-naming experiments in L2 with intermediate to highly proficient Italian-Spanish bilinguals along with a control group of Spanish monolingual speakers. Italian and Spanish have similar gender systems (structurally, proportionally, phonologically and morphologically). Results from both experiments revealed a main effect of GC for both BNs and NPs. Morales et al. (2011) replicated Paolieri's et al. results in a picture-naming study with Italian-Spanish bilinguals, this time of intermediate proficiency. The authors were also interested in investigating whether inhibitory mechanisms are responsible for solving between-language competition at the grammatical representational level. In a first picture-naming experiment, to create more or less L1 inhibition, they

manipulated the number of presentations of each picture (one or five times). In a second experiment, participants were presented with the same pictures and instead of naming them in their L2, they had to produce L1 definite determiners that agreed with the gender of the L1 nouns that denoted the pictures. They expected that the gender-incongruent pictures presented and named five times in the L2 in the first experiment would create more inhibition in the L1 than those pictures presented only once. A GC effect was obtained in the first experiment, replicating the data of Paolieri et al. (2010). Also, in the second experiment, the retrieval of L1 grammatical information (i.e., retrieving the appropriate article) of incongruent picture target words took longer relative to the retrieval of the appropriate article for gender-congruent pictures. Interestingly, the greater the number of L2 repetitions in the previous experiment the bigger the GC effect in this second experiment when retrieving the appropriate articles. Together, these two studies suggest the existence of a gender-integrated system where cross-linguistic interactions take place during lexical access at the level of grammatical encoding. Particularly, the study of Morales et al. (2011) defends connectionist models of bilingual language processing as it confirms the hypothesis of co-activation and competition for selection of both gender values, supporting the idea of an inhibitory control mechanism that works in a competitive situation between different gender values of target nouns across languages.

As the differences between the languages grow, the GC effect still appears but some small modulations due to other variables such as L2 proficiency or the phonological transparency of the target words come up as important. Specifically, Bordag (2004) conducted two mixed-language picture-naming tasks (participants, thus, were in the ‘bilingual mode’, Grosjean, 1997a, b) with intermediate to advanced proficiency German-Czech bilinguals and with Czech-German bilinguals of the same proficiency. Structurally, the gender systems of both languages are identical (three genders: feminine, masculine and

neuter). However, Czech does not have articles and so grammatical gender acquisition is slightly different from that in German. The results revealed a gender congruency effect with both BNs and NPs in RTs (shorter response times for gender-congruent pictures than for incongruent ones –GC < GI) and errors (measured as agreement error rates and only obtainable for the NP condition). Interestingly, Bordag and Pechmann (2007) replicated those effects with Czech-German bilinguals and, in a second experiment, moved the participants to the ‘monolingual mode’ by asking them to name all the items in German (L2). The GC effect was once again obtained for BNs and NPs. This means that the L1’s grammatical gender was also activated despite the fact that L1 was not required for the task. In a third experiment, they replicated previous experiments with the main aim of testing the role of the phonological gender transparency, dividing the target nouns into transparent, opaque and irregular.

Additionally, participants were more fluent than the ones from the previous experiments.

They analysed the variables of gender congruency and phonological gender transparency in two different ANOVAs, so no interaction effects were reported. Again, the results showed a GC effect for both BNs and NPs, although a numerically smaller one, a fact that the authors attributed to the higher proficiency level of the participants. Also, subjects were significantly slower when they named irregular nouns (no differences between transparent and opaque nouns) and they made more errors when naming opaque or irregular nouns in comparison to transparent nouns. The results not only support a gender-integrated system even when the gender systems of the bilingual are slightly different and the participant is in a “monolingual mode”, but also support interactive models of language production (i.e., connectionist models). Having controlled the frequency of the targets, irregular nouns took more time to be processed than transparent and opaque nouns and irregular and opaque nouns produced more agreement errors. This points toward a bigger (and not necessarily more precise) processing effort being made to access grammatical gender when a misleading or ambiguous cue is

present. In other words, the advantage obtained for transparent nouns seems to suggest that the nominal ending of the noun is being processed to select a gender value and, thus, to access a noun. The interaction between lexeme (phonological form of the noun) and lemma (grammatical characteristics of the noun) levels to select a proper gender value points to the existence of bottom-up and top-down processes. In any case, outright affirmations cannot be made as the interaction between the factors gender congruency and phonological gender transparency was unfortunately not analysed.

Another variable that we pointed out as fundamental is the state of cognateness of the target translations. Lemhöfer et al. (2008) addressed its impact on the GC effect through two picture-naming experiments that replicated Bordag and Pechmann's work but with intermediate to advanced German-Dutch bilinguals. German and Dutch have similar gender systems, but they are structurally different. German has three gender values (masculine, feminine and neuter) and Dutch has two (common and neuter), although in the latter, the masculine and feminine gender existed some decades ago and recently collapsed into the common gender. Consequently, when the German gender value was 'feminine' or 'masculine' and the Dutch gender value was 'common', picture target nouns were gender-congruent, as well as when they were both neuter. Otherwise (fem.-neuter, masc.-neuter, neuter-common) they were gender incongruent. In addition, half of the picture target nouns were cognates, defined as highly phonologically and orthographically similar (e.g., *hund* - *hond* [dog]), while the other half were non-cognates (e.g., *kleid* - *jurk* [dress]). The participants were also asked to name the pictures using BNs and NPs (definite determiner + noun) in their L2. The results for the first naming task experiment revealed an effect of GC based in error rates which was restricted to NPs. Specifically, they found less gender agreement errors for gender-congruent Dutch nouns than for gender incongruent ones (GC < GI). The effect was significantly larger for cognates but existed in both types of translations.

However, the results on RTs were not significant. The authors noted that the error rates were extremely high (40.5% in the NP condition). This loss of statistical power could be responsible for the failure to find any effects of GC in the RT analysis. Hence, in their second experiment, they tried to reduce error rates and to explore the origin of the effects of GC. Thus, participants were trained in terms of the gender of items beforehand, as a means of reducing the high number of errors. Also, since this training consisted of participants repeatedly producing the experimental stimuli in gender-marked phrases (and receiving feedback), a measure became available that indicated the stability of the gender representation. This made it possible to study whether the difficulties they were experiencing with gender incongruent target nouns arose from problems during L2 gender acquisition (that led to unstable gender representations) rather than from online lexical competition processes between conflicting gender information. The results revealed: 1) that the training was effective as a way of reducing error rates; 2) a GC effect (RTs) restricted to NPs; 3) that cognates primarily carried the gender effects, with higher degrees of facilitation on the gender-congruent condition and of interference in the gender incongruent condition in comparison to the effects found with non-cognates; 4) the stability analysis indicated that the interference observed with gender incongruent pairs were carried especially by nouns with unstable gender representations. The GC effect disappeared for stable nouns. Thus, the primary mechanism causing the gender effects might be an increased difficulty in acquiring correct and stable gender representations. Of course, this does not discount the possibility of an ‘online’ competition between the two conflicting gender representations. In fact, when looking at cognates only, there was an effect in the error rates and a trend on the RTs even for the ‘stable’ group. In any case, it is crucial to keep in mind that in this experiment, the stability of the gender representations could have played a special role in comparison to previous studies, since Dutch is known to have a very difficult gender system to master for

both monolinguals and bilinguals. Unstable gender representations are very common and remain a problem even for bilingual children of Dutch, whose age of acquisition is low and exposure to the L2 extremely high (e.g., Blom, Poliřenská, & Weerman, 2008; Van der Velde, 2003). Regarding the absence of a GC effect for BNs, as we have stated in the introduction, even in the monolingual literature GC effects are not usually obtained with BNs in Germanic languages. Even more, if the absence of results with BNs is due to the opaqueness of languages (Sá-Leite et al., in preparation), the results here are not surprising as Dutch is the most opaque Germanic language and the one carrying the majority of null results with BNs in the literature (La Heij, Mak, Sander, & Willeboordse, 1998). Thus, whereas the connectionist models of language production clearly support the results seen with Romance languages (Morales et al., 2011; Paolieri et al., 2010) and even those with slightly more transparent Germanic languages (Bordag, 2004; Bordag & Pechmann, 2007), the WEAVER++ grammatical gender processing statements seem to adjust better for Lemhöfer et al. (2008) results with L2 Dutch.

An integrative view and non-selective access to the bilingual lexicon still makes it to the picture here when explaining the interaction between the factors ‘cognateness’ and ‘gender congruency’. Specifically, it seems that when a L2 noun has to be processed, its L1 translation becomes active as well (non-selective access), along with its grammatical features (i.e., gender). If the gender values of both translations are incongruent, then there will be competition for selection, hampering L2 gender processing. The co-activation of the translation equivalent and its gender is larger for cognates than for non-cognates due to cross-language similarities (see Comesaña et al., 2015, for more details; also Costa, Caramazza, & Sebastián-Gallés, 2000). According to the authors, this is why both the facilitation and the interference of gender-congruent and incongruent translations is higher for cognate translations than for non-cognate translations.

Finally, there are two papers in which the difference between languages is quite high but the GC effect is still successfully obtained (Klassen, 2016; Manolescu & Jarema, 2015). More specifically, Manolescu and Jarema (2015) replicated Paolieri et al.'s (2010) study with Romanian and French highly proficient bilinguals that had started to learn French in childhood after immigrating to Montreal. They also recruited native French speakers as a control group. Structurally, Romanian has one more gender (masculine, feminine, and neuter) than French (masculine and feminine) does. The authors state that pairs of 'neuter-masculine' or 'neuter-feminine' translations can be treated as potentially gender incongruent pairs. Moreover, the distribution of definite determiners is different in the two languages, and thus gender acquisition and processing differs slightly, not only morphologically but also syntactically. The participants had to name the pictures using a BN or a NP (indefinite det. + noun). The results for both experiments replicate those in Paolieri et al. (2010), that is, a GC effect for both BNs and NPs which was restricted to bilinguals. However, we should point out that although incongruent conditions (masculine-feminine and feminine-masculine) and potentially incongruent conditions (masculine-neuter, feminine-neuter) yielded higher RTs than congruent conditions, no differences were observed between incongruent and potentially incongruent conditions (there was only a tendency for marginally shorter RTs for the potentially incongruent condition than for the incongruent one).

In the other hand, Klassen's (2016) study is especially interesting because, similarly to Costa et al. (2003), it explores a group of bilinguals of *supposedly opposed* linguistic families (i.e., bilinguals of a Germanic and a Romance language). She conducted a picture-naming task with Spanish-German intermediate bilinguals, along with a control group of native German speakers. Spanish and German have quite different gender systems, especially in terms of structure, since German has a tripartite gender system (masculine, feminine and neuter) whereas Spanish only has two gender values (masculine and feminine). In addition,

Spanish is a highly phonological gender transparent language, contrary to German, which is quite opaque. As Manolescu and Jarema (2015), the author considered one further condition, the ‘potentially incongruent condition’. That is, instead of a mismatch in the gender systems between the two languages, there is an absence of neuter gender in Spanish. Thus, pairs like masculine-neuter or feminine-neuter would be potentially incongruent. In addition, they featured two conditions depending on the naming instruction: BN or NP. Results revealed a main effect of GC for both BNs and NPs in the experimental bilingual group, significant by participants but not by items. The effect was due to gender incongruent pairs since whereas RTs were significantly lower for potentially incongruent than for gender incongruent, differences between potentially incongruent and gender-congruent ones showed only a tendency. Naming latencies did not differ significantly between naming conditions (BNs and NPs). Error rates replicated the results obtained in RTs. Apart from the fact that gender effects support the gender-integrated representation hypothesis, an interesting result was observed with neuter nouns. As the author explains, although both the gender-incongruent and neuter conditions are, by definition, L1–L2 gender incongruent, neuter gender value is only present in the L2. This might cause the linguistic system to encode the neuter gender value (without an equivalent in the L1) separately from the masculine and feminine shared gender nodes. This separate representation is not subject to interference from the activation of the masculine and feminine shared gender nodes. Interestingly, in their study Manolescu and Jarema (2015) found that the behaviour of neuter (the absent gender in the L1) was somehow quantitatively different from that of incongruent nouns (i.e., marginally shorter RTs), but congruent nouns still had the shortest RTs by far. This means that the ‘gender value in discord’ (in comparison to the L1 gender system) was still causing trouble, although it might not have been entirely incongruent.

In sum, with the exception of Costa et al. (2003), research using picture-naming tasks supports an integrated view of the bilingual gender representation. Importantly, the similarity between the gender systems of the pair of languages at issue has not determined the GC effect here. Thus, the GC effect has been obtained in highly similar pairs such as Spanish and Italian (Morales et al., 2011; Paolieri et al., 2010), similar pairs such as Czech and German (Bordag, 2004; Bordag & Pechmann, 2007) and dissimilar pairs such as Romanian and French or German and Spanish (Klassen, 2016; Manolescu & Jarema, 2015). The fact that the GC effect has been observed with participants in both a monolingual and a bilingual mode, as defined by Grosjean (1997a, b), supports the idea of a non-selective access in a highly integrated lexicon. Importantly, this degree of integration extends to the gender systems of the bilingual. Regarding the participants, there are other characteristics that have shown to affect the results. When Bordag & Pechmann (2007) used bilinguals that were more proficient and whose age of acquisition was lower (10.7 years old on average, for more detail, see Table 1) in their third experiment, the GC effect turned out smaller in relation to their other experiments. Other interesting findings are in line with interactive models of language production that support the existence of bidirectional links between the lemma and lexeme levels. Specifically, the degree of phonological transparency of the L2 target nouns affected the naming times and error rates, with irregular nouns taking more time and producing more naming errors (still, the interaction with the GC effect was not explored). This suggests that there is a direct influence of the phonological encoding of the word on the selection of the proper gender value at the lemma level. Top-down mechanisms as those supported by connectionist models of language production would be behind these results. On the other hand, these models also support the effect of cognateness here obtained: cognates seem to boost the GC effect, in the way that the interference caused by incongruent gender translations is higher for cognates, but the facilitation caused by those that are congruent is

also higher. Only a system with non-selective access and competitive/inhibitory mechanisms such as those proposed by Morales et al. (2011) would explain these results.

Besides, and most importantly, the effect seems to be obtained with both BNs and NPs, something that goes against the postulates of the WEAVER++ model that understands grammatical gender more of as a pure syntactic property only activated when agreement is necessary. Interestingly, the only study with Dutch as target language (L2) found no GC effects with BNs, something coherent with what has been observed in the monolingual literature (i.e., absence of GC effects for every study conducted with Dutch participants using BNs).

Studies with forward-translation tasks

Salamoura and Williams (2007) were the first to use a forward-translation task to explore the GC effect in bilinguals. In this task, participants are presented with an L1 word on a screen and they have to translate it into the L2 as fast and accurately as possible. The L2 translation can be heterogeneric or homogeneric in relation to the L1 target word. The authors tested the gender-integrated hypothesis with BNs and NPs and they looked at whether the degree of cognateness between L1 and L2 nouns was a prerequisite for a shared representation of gender features and, thus, of gender processing interaction between both languages. Greek-German advanced proficiency bilinguals were recruited. Greek and German have structurally similar gender systems (masculine, feminine and neuter) although these are proportionally and phonologically different. Participants were presented on a screen with either BNs or NPs (adjective BIG or SMALL + noun) and had to translate them into the L2. Two blocks were created, one of cognate nouns (defined by the authors as translations that have a highly similar phonological form, with orthographic overlap varying since Greek and German have different alphabets) and the other of non-cognate nouns. The results

revealed an effect of GC across languages only with NPs for RTs and agreement error rates, which were numerically greater for cognates than for non-cognates. These results give support to a shared representation of gender features between L1 and L2 and replicate the results obtained with cognates in picture-naming tasks (Lemhöfer et al., 2008). They also give support to the WEAVER++ model, as the effects were observed only with NPs, this is, when agreement must be established.

Bordag and Pechmann (2008) conducted three forward translation experiments with Czech-German upper-intermediate to advanced proficiency. We might bear in mind that these two languages share the same number of gender values (three) and their major difference consists in the absence of articles for Czech, which makes gender acquisition different from German. The stimuli set was identical to the one reported in Bordag (2004), that is, nouns were transparent, half belonging to the congruent gender translation condition and half to the incongruent gender translation condition. Two conditions featuring short (BNs) and long (NP: adjective SMALL or BIG + noun) responses were included. No significant differences were observed between the two critical conditions (congruent vs. incongruent), either for BNs or for NPs. In their second experiment, they replicated the previous one but using a new set of materials, those from the picture-naming task used in Bordag and Pechmann's (2007) study. Hence, the phonological gender transparency was also manipulated (16 transparent, 16 opaque and 16 irregular nouns distributed evenly). A new group of participants drawn from the same population as before was selected. Again, no effects of GC were obtained. However, transparent nouns were faster and more accurately translated than opaque and irregular nouns, although only in the long condition (NPs). To increase the probability of L1 gender retrieval (and thus, of cross-linguistic gender congruency effects), the authors decided to conduct what they called a 'pure' translation task in their third experiment. This was because in the previous experiments participants did not have to perform a pure translation of NPs

(i.e., in both the long and short conditions, they only translated the noun; the size of the adjective they had to produce in the long condition was determined by the size of a dot that appeared in front of the word). This new experiment was a replication of Experiment 2, but with BNs and complex NPs (gender marked adjective + noun). The participants were drawn from the same population as in the other experiments. No GC effects were obtained. Again, the same effect of phonological gender transparency for NPs in RTs was obtained, but the differences were marginal in error rates. As in the picture-naming study of Bordag and Pechmann (2007), despite the fact that no effects of GC were obtained, gender was being processed at least in the NP condition, since the faster retrieval of nouns with a gender transparent termination indicates a computation of the gender value in which gender transparency is playing a role.

Even though these results seem to not fully support a gender-integrated system, other studies with translation tasks did find GC effects, even when non-cognates were included in the stimuli (Manolescu & Jarema, 2015; Paolieri et al., 2010). For instance, Paolieri et al. (2010) conducted a forward translation task (from L1 to L2, Experiment 3) with Italian-Spanish highly proficient bilinguals drawn from the same population as their experiments with picture-naming tasks. Likewise, the same materials as those in their previous experiments were used. Participants had to translate either a BN or a NP (Definite determiner + noun). An effect of GC was obtained with both BN and NPs, again revealing faster responses to gender-congruent nouns than to gender-incongruent ones. Manolescu and Jarema (2015) replicated this experiment with the materials used in their picture-naming task and obtained the same effects (a GC effect for both BNs and NPs) with highly proficient Romanian-French bilinguals.

Recently, Paolieri et al. (2018) conducted two experiments based on oral translation tasks with a threefold aim. First, comparing once again the retrieval of grammatical gender in

a BN vs. a NP condition; second, exploring more deeply the role of similarity between the gender systems of a bilingual's two languages on gender congruency effects. For this, in their first experiment they recruited Russian-Spanish bilinguals, two languages whose gender systems are structurally, morphologically and phonologically different. They then compared the results to those obtained in a second experiment with Italian-Spanish bilinguals, two languages with very similar gender systems. Finally, their third objective was to explore the influence of semantic mediation on the occurrence of grammatical gender effects in bilinguals. Specifically, they based their hypothesis on the fact that abstract words seem to have fewer semantic features than concrete words (e.g., de Groot, 1989; see also Ferré, Sánchez-Casas, Comesaña, & Demestre, 2017). Consequently, the number of semantic elements that abstract words can share with their translations is very much reduced in comparison to concrete ones. This allows them to test what Konishi (1993) and Boroditsky, Schmidt, and Philips (2003) claimed about arbitrary syntactic features such as the grammatical gender of nouns being part of the conceptual representation of the objects they refer to. If grammatical gender and semantic information interact, concreteness should interact with the gender congruency effect. Thus, in the first experiment, Russian-Spanish bilinguals were asked to translate nouns from their L1 to the L2 by producing either a BN or a NP. The L1 nouns could be gender-congruent, gender incongruent or, since Russian has a third gender value that Spanish does not (i.e., neuter), neuter-incongruent with their L2 translations. Concreteness was also manipulated (concrete vs. abstract nouns). Results revealed modulation in the GC effect (faster responses for congruent gender pairs than for incongruent or neutral-incongruent gender pairs) as a function of concreteness and response type (BN vs. NP). For abstract nouns, the gender congruency effect was only significant with NPs, whereas for concrete nouns the effect was present for both conditions. In the second experiment, Italian and Spanish advanced bilinguals were recruited and the materials were

controlled for the same variables, with the same number of transparent nouns present in both languages. An effect of GC was obtained both for BNs and for NPs, as well as for concrete and abstract nouns, although the magnitude of the effect was greater for the concrete ones. Thus, the evidence obtained was: (1) in favour of the interaction between both gender systems even when they are quite different, although this interaction seems to be greater when the similarity between them is higher; (2) against an agreement context (i.e., NPs) as a requirement for processing gender, although it indeed boosted gender effects; and (3) in favour of an interaction between encoding at the semantic and grammatical levels, perhaps because gender is partially related to semantics contrary to what is commonly presented in its definition (i.e., that grammatical gender is totally independent from semantics).

Summary

To summarise, we have seen thus far that the majority of production studies on gender processing with bilinguals who speak languages with more or less similar gender systems (see Table 1 for more detail) found GC effects (i.e., 9 out of 11 studies). Overall, these findings give support for a gender-integrated representation perspective. This is true both for Romance and Germanic languages (as L2) with intermediate (e.g., Bordag, 2004) and highly proficient bilinguals (e.g., Manolescu & Jarema, 2015; Paolieri et al., 2010). The only study that seems to present strong evidence against this view is that of Costa et al. (2003) with three different bilingual populations (Croatian-Italian, Spanish-Catalan, and Italian-French bilinguals). Leaving aside its methodological limitations, it is worth noting here that some inconsistencies persist even with bilinguals who speak the same languages (e.g., Czech-German bilinguals showed gender effects in Bordag and Pechmann [2007], but null effects were observed in Bordag and Pechmann [2008]). Task requirements might be responsible for the seemingly contradictory results in these latter studies, probably because the time course of the activation of the gender features in L1 and L2 in translation and picture-naming tasks

might have been different (Bordag & Pechmann, 2008). For the latter, the activation spreads from the concept in common to both L1 and L2 lemmas in parallel, to the level of grammatical encoding. Thus, the L1 and L2 gender nodes (or the same gender node) are activated at the same time and they can compete for selection. In forward translation, the L1 word form and its lemmas are activated firstly, then the activation spreads to the lemma of the L2 translation equivalent (or to the concept and then to the L2 lemma). This means that the L2 gender node becomes activated after the L1 gender node. However, the fact that other studies with languages as different as Czech and German have found gender effects regardless of task requirements (e.g., Manolescu & Jarema, 2015 with Romanian and French), weakens this hypothesis. Therefore, the reason of the null GC effects observed in Bordag and Pechmann's (2008) remains unclear. It is likely that gender effects are not as easily obtained in forward translation-tasks as in picture-naming tasks, especially when languages other than Romance ones are involved. Note that the two studies (out of five) with translation tasks that did not obtain gender effects when participants were asked to use BNs (Bordag & Pechmann, 2008; Salamoura & Williams, 2007) were the ones featuring non-Romance languages. In comparison to other languages (such as German, the one used as L2 in the two experiments that failed to obtain gender effects with BNs), Romance languages might activate gender to a greater extent, this probably because they have a highly transparent gender system and the nominal ending of the noun is always relevant to the system. In fact, in Spanish, Italian, and French the phonological cues for gender have been shown to affect BN processing (e.g., Sá-Leite et al., 2017; Urrutia et al., 2009), something that has not been observed in languages like German where phonological cues do not play a central role.

Regarding other characteristics of the languages at issue, it is important to point out that the GC effects were more robust for cognate translations. This may be explained due to the resemblance of form and meaning between cognates (Salamoura and Williams, 2007).

Taking into account that the studies that used cognates selected highly similar phonological translations (although no information about identical cognates is provided, a careful inspection of the materials from all the selected studies showed no identical cognates in the lists), this overlap might lead not only to language-shared semantic, phonological and/or orthographic representations, but also to a correspondence between L1 and L2 cognates in terms of lemma (including its grammatical features such as gender). The cognateness may prompt learners to link the new L2 word to the gender value of its L1 translation until evidence to the contrary is available. When such evidence is provided, the strong link to the L1 lemma must be progressively reduced. The formal dissimilarity of non-cognates, on the contrary, may make learners more cautious about equating L2 with L1 gender, forcing them to develop stronger L2 lemma independently of links to the L1. On the other hand, evidence is obtained in favour of bidirectional mechanisms between the lemma and lexeme levels similar to those proposed by connectionist models of language processing. Specifically, it seems that transparency facilitates L2 noun processing, something that has been shown in other studies featuring picture-naming tasks aimed exclusively at testing this effect (e.g., Bordag & Pechmann, 2007; Bordag, Opitz, & Pechmann, 2006). Thus, support is obtained for the idea that bilinguals are sensitive to phonological gender transparency.

Finally, regarding the need for an agreement context to process gender, and thus to capture gender effects, the reviewed studies suggest that an agreement context is not mandatory. Seven out of ten production studies that tested this obtained gender effects using BNs (see Table 1). It seems, though, that an agreement context might boost these gender effects. The studies that did not obtain effects with BNs were those featuring forward-translation tasks, as discussed above (Bordag & Pechmann, 2008; Salamoura & Williams, 2007) and one featuring a picture-naming task (Lemhöfer et al., 2008). Following the reasoning given in the discussion, the target language in this picture-naming task (Lemhöfer

et al., 2008) was Dutch (L2), one of the most gender-opaque Germanic languages.

Furthermore, the L1 was another gender-opaque language, German. As we have said, studies in the monolingual domain using these languages also failed to observe gender effects with BNs (e.g., La Heij et al., 1998; also Sá-Leite et al., 2017).

Gender processing in bilingual comprehension

Regarding language comprehension, three studies on grammatical gender processing in bilinguals have used either Lexical Decision Tasks (LDT) or the visual world paradigm. These studies are particularly interesting, as the monolingual models of language processing that have assessed grammatical gender selection (WEAVER++, CMSP) have been developed to explain language production rather than language comprehension. Nevertheless, connectionist models of bilingual language processing characterized by a highly integrated lexicon with no-selective access, sustain the existence of cross-language interactions at all levels during word recognition (BIA, BIA+) and production (Multilink) and thus gender effects would be expected with BNs and NPs. Again, the similarity between the gender systems of the bilinguals might determine these interactions.

The first authors to conduct an LDT were Lemhöfer et al. (2008) and tested bilinguals of two highly opaque and similar languages: German-Dutch bilinguals that had been living in the Netherlands for 1.5 to 11 years (probably intermediate to high proficient bilinguals). They were presented with a string of letters and they had to decide as fast and accurately as possible whether or not it was a real word in the L2 Dutch. The pseudo-words (matched in length to experimental words) were constructed by changing one or more letters in existing Dutch words. Participants saw a determiner (prime) and shortly after that (250 ms) a letter string. Half of the trials were presented with the indefinite determiner (gender unmarked) and the other half with the definite determiner (gender marked). Along with the gender-

congruency condition, they also explored the role of cognateness (cognates vs. non-cognates, with cognates being defined as highly phonologically and orthographically similar, as in their previously described naming tasks) and the prime type (definite vs. gender-neutral indefinite determiner). Results revealed an effect of GC regardless of prime type: gender-congruent translations were faster recognised than gender incongruent translations when preceded for both definite and indefinite determiner ($GC < GI$). Although not significant, the data pattern for cognates and non-cognates looked qualitatively different: there seemed to be a crossover interaction of gender congruency and prime type in the cognates, but this was absent for non-cognates. Indeed, the analysis of cognates revealed a significant interaction between gender congruency and prime type. For incongruent gender cognate pairs, RTs were 31 ms longer for definite (gender marked) than for indefinite (gender unmarked) determiner primes. For congruent gender cognate pairs, there was a marginal significant advantage of 16 ms for definite determiners primes over indefinite determiner ones. Summarising, a GC effect in comprehension was found which was somewhat modulated by cognateness similarly to what happened in production. These findings give support to: i) a gender-integrated representation view; and ii) a cross-language interaction during grammatical gender encoding which seems to be intensified by the status of cognateness.

Weber and Paris (2004) and Morales et al. (2016) tested oral rather than visual L2 word recognition. This is particularly interesting since it extends, if present, the cross-linguistic influences at the level of grammatical encoding to the oral and more common means of language comprehension. Following the connectionist interactive activation models (BIA, BIA+ and Multilink), through the activation of the phonemes, concepts from both languages should be activated and, thus, if the results are similar to those in the previous study, the gender systems of the bilingual person should influence each other. Weber and Paris (2004) conducted a study on gender processing using a spoken-word recognition task

within the visual world paradigm along with the eye-tracking technique to assess this question. The authors recruited French-German highly proficient bilinguals as participants, with French and German having quite different grammatical gender systems (structurally, morphologically, and phonologically) and belonging to different language families (Romance and Germanic, respectively). Following spoken instructions in German (L2), they were required to click on a target picture on a screen from a set of four pictures while their eye movements were recorded. Two of those pictures were fillers phonologically unrelated to the target item, which could coincide or not in gender with them (i.e., the gender value of the fillers was randomised). One of the other two was the target, which always had a translation of the same gender in the L1 (e.g., *kassette* [tape] is a feminine noun in German and it translates to another feminine noun in French, *cassette*). The noun corresponding to the target picture they had to select was said in the instructions while they were seeing the four pictures. That target's noun was preceded by a definite article that agreed in gender with it (e.g., '*Wo befindet sich die kassette?*' [Where is the tape?]). The remaining picture was a competitor, whose corresponding noun overlapped in onset (initial phonemes) with the target picture noun in both languages as a means of creating an ambiguous situation (e.g., *kanone* [cannon]). In the L2, target and competitor nouns always shared gender (e.g., *kassette* [tape] and *kanone* [cannon] are both feminine), so the gender marking on the article could not exclude the competitor as a lexical candidate. Critically, target and competitor translations could differ in gender in the non-target language (e.g., in French *cassette* [tape] is feminine and *canon* [cannon] is masculine). In trials with the target and competitor of different genders in the L1 (e.g., *cassette* - *canon*), and where participants did not on average look more at the competitor (*kanone*) than at the filler images, then participants were seen to be using L1 French gender information (*canon* is masculine) to disambiguate between target and competitor in their L2 in a situation where the L1 is irrelevant. Results revealed that in trials

with same-gender pairs (i.e., same gender in French L1), participants on average fixated on competitor objects significantly more than filler objects, as also happened with the target. However, the pattern of results changed for trials with different-gender pairs (i.e., different gender in French L1). Participants no longer fixated on competitor objects significantly more than filler objects on average, the target picture being the one on which they fixated significantly more. This means that gender information provided by the French translation of the competitor could constrain competitor activation in the L1, as the author notes: ‘Despite its phonological similarity with the target noun, the competitor was not activated when the article of the target noun did not match in gender with the competitor in French’ (Weber & Paris, 2004, p. 1450). In a second experiment with native monolingual German speakers, results showed that, conversely, German listeners looked more often at the competitor than at the unrelated fillers both in same-gender and in different-gender pairs. This study adds evidence to the activation of the L1 translations grammatical gender in a pure L2 oral context with highly proficient bilinguals, at least when NPs are involved.

Morales et al. (2016) replicated Weber and Paris’ study (2004) with upper intermediate bilinguals of highly similar and transparent Romance languages: Italian and Spanish. Spanish monolinguals were also included as a control group. Here, participants viewed only a pair of pictures (objects) on a computer screen, so the ambiguity was guaranteed by the same gender of the two pictures. As in the previous study, instructions were presented in Spanish L2 and a NP always preceded the target noun (*‘Encuentra la bufanda’* [find the scarf]). Interestingly, in this experiment materials were controlled in both Spanish (L2) and Italian (L1) (for more information on the controlled variables, see Table 1). The grammatical gender of the objects’ name was manipulated so that pairs of objects had the same (congruent) or different (incongruent) gender in Italian, although the gender in Spanish was always the same. The results revealed that participants looked at the target

pictures significantly less (less fixations) when their translation equivalents in Italian were incongruent in gender. Hence, target items that did not share gender across languages yielded a reduced proportion of fixations on target pictures relative to target items in the congruent gender condition. Also, no interaction was found between gender value (masculine, feminine) and condition (congruent, incongruent), thus the effect did not vary across gender values (note that this made it possible to discount the potential impact of the definite feminine article ‘*la*’ [the] being the same in both languages). Importantly, Morales et al. (2016) explored the time course of the gender effect and observed that the influence of L1 grammatical gender affected spoken-word recognition from 498 ms onwards (until the end of the time window, 900 ms).

In a second experiment, using participants from the same population as those in Experiment 1, Morales and colleagues increased the experimental control of new materials (care was taken to include only non-cognate translations) in order to guarantee that they did not contain any phonologically related words between languages and that target and distractor nouns never shared orthographic onset in either language. In addition, they included a third condition, where the gender of the two pictures was different in both L1 and L2. As well as replicating the previous results, this experiment aimed to compare the time course of the two gender effects through this third condition. Thus, included were ‘same-different gender’ trials (in the L2, nouns such as ‘*pentola*’ [FEM] and ‘*farfalla*’ [FEM], ‘pot’ and ‘butterfly’ have the same gender and nouns such as ‘*cuscinio*’ [MASC] and ‘*pecora*’ [FEM], ‘pillow’ and ‘sheep’ have different gender) and congruent-incongruent gender trials (in the previous example, both pairs’ L1 translations were gender-congruent, since ‘pot’ and ‘butterfly’ as well as ‘pillow’ and ‘sheep’ are also feminine in L1 Spanish, ‘*olla*’ and ‘*mariposa*’, ‘*almohada*’ and ‘*oveja*’; if any of the pairs had an L1 gender incongruent translation, then they would be of the same gender in L2 but gender incongruent in relation to L1). The factor ‘same-different gender in

L2' would indicate the moment when L2 determiner gender processing starts, as the determiner disambiguates the choice, while the second factor 'congruent-incongruent gender L1-L2' would signal the moment at which L1 co-activation occurs. Results revealed, first, that L2 determiner gender processing initiated before L1/L2 gender co-activation. Second, there was a significantly higher number of target fixations in the congruent condition than in the incongruent condition, with no interaction with the factor 'target gender value'. Third, the difference between target fixations in the congruent and incongruent conditions became significant at 360 ms from article onset. Morales et al. (2016) conclude that the grammatical gender of the L1 modulates spoken-word L2 recognition shortly after L2 gender information becomes available to bilinguals. Results with a monolingual control group of native speakers of Spanish revealed null effects of gender congruency/incongruency.

Summary

To summarise, all the comprehension studies reviewed saw gender effects across-languages, both for visual and oral recognition, with an early gender-effect of the L1 on the spoken L2 (Morales et al., 2016). The effect was obtained even when the pair of languages tested had highly different gender systems probably because in all of them one of the languages was Romance and thus transparent (it would be interesting that future studies assess if this is a *sine qua non* condition to observe GC effects). In visual recognition tasks, participants showed lower RTs when the L1 translations had the same gender as the target nouns (GC effects). In oral recognition tasks, the competitor pictures prompted more interference when their gender was the same of its L1 translation. In addition, as in production studies, gender effects seem to be boosted by the cognateness of the translations (Lemhöfer et al., 2008).

Discussion

In this review paper, we aimed to explore grammatical gender processing in bilinguals during the production and comprehension of L2 nouns. Two main views of bilingual gender representation were contrasted: the integrative vs. the autonomous one. Whereas the first one supports an integrated gender system for both languages of the bilingual and thus, it predicts cross-linguistic effects during grammatical encoding, the latter expects the opposite, this is, two separate gender systems that do not influence each other. The majority of the models on both language production and comprehension in bilinguals would expect an interaction to occur between both languages during grammatical gender selection, especially connectionist models such as the BIA (Dijkstra et al., 1998), BIA+ (Dijkstra & van Heuven, 2002), and the Multilink model (Dijkstra et al., 2018). These models firmly defend a non-selective access to an integrative lexicon at (supposedly) all levels. However, grammatical gender is a very peculiar feature whose characteristics can vary widely from one language to another, and hence, it cannot be assumed that our linguistic system considers grammatical gender from different languages to be the same parameter. Thus, the integration of both gender systems and their possible interactions were analysed with special focus on the similarity between the gender systems of the participants and the linguistic families (Romance vs. Slavic/Germanic). Likewise, other factors affecting the L1 influence on the L2 and related to the characteristics of the bilingual population at issue (AoA, L2 proficiency) and the languages themselves (state of cognateness, phonological gender transparency) were considered. The degree of phonological gender transparency of the languages and the stimuli was particularly interesting as it tested the bidirectional links proposed by certain connectionist models between the lemma and lexeme levels (e.g., BIA, BIA+, Multilink). Finally and more elementally, it was explored whether or not an agreement context is a mandatory condition for processing the gender of a single noun, putting the universality of some models developed

in the monolingual domain into question (CMSP, Dell, 1999; WEAVER++, Levelt et al., 1998).

Considering all the findings, we conclude that the interaction between the L1 and the L2 lexicon occurs at the level of grammatical encoding. A cross-language gender congruency effect (GC effect, a facilitation of the processing of L2 nouns when their translation into L1 is congruent in gender in comparison to when it is incongruent) has been systematically observed during both L2 production (9 out of 11 studies) and comprehension (3 out of 3 studies). The effect was observed even when the gender systems of both languages were quite different, at least whenever one of the two spoken languages was transparent (e.g., Manolescu & Jarema, 2015; Paolieri et al., 2018; Salamoura & Williams, 2007). This means that, even in these cases, grammatical gender is considered as the same linguistic parameter in the bilingual mind. Although, Costa et al.'s (2003) and Paolieri et al.'s (2018) studies suggest that the greater the distance between the languages, the lesser the effect. This is in line with some studies that use eye-tracking techniques and electrophysiological measures, which show that the higher the similarity between grammatical rules in both languages the higher the sensitivity of the participants to agreement errors (e.g., Foucart & Frenck-Mestre, 2012; Sabourin & Stowe, 2008). Regarding the characteristics of the participants at issue, this effect seems to be restricted to adult bilinguals, mostly unbalanced, with a late AoA (+10 years old). The only study that features bilingual participants whose AoA was listed as less than 10 is Manolescu and Jarema (2015), since L2 was acquired during childhood, but no further information is specified. Importantly, even in this case, gender effects were observed. In addition, the effect was obtained with bilinguals both in the monolingual and in the bilingual mode (Grosjean, 1997a, b). Regarding the proficiency on the L2, the influence between languages during gender processing occurred both for intermediate and high proficiency bilinguals, although L2 higher proficiency and native-likeness might contribute to

a reduction of the effect (Bordag & Pechmann, 2007; Costa et al., 2003). As to the analysed languages, the effect has been obtained for both transparent and opaque L2s. Yet, for opaque ones the effect seems to be slightly harder to obtain, perhaps due to the absence of phonological gender cues that facilitate and may prompt nouns' gender activation baseline to a higher level than opaque languages (e.g., Bordag & Pechmann, 2008; Lemhöfer et al., 2007).

In terms of language production, gender effects seem to be more easily obtained with picture-naming tasks than with oral-translation tasks, probably because the time course of L1 and L2 gender activation differs as a function of task requirements (see subsection Summary on the production studies section for more detail). Regarding language recognition, the effect was observed in all the reviewed studies, both in visual and oral recognition tasks. These gender effects on comprehension are in line with the tenet of connectionist models (BIA, BIA+ and Multilink) that multiple word candidates are simultaneously activated during recognition regardless of the language, and that they compete against each other.

Thus, evidence theoretically supports a gender-integrated representation view of the bilingual mind where both languages share the same gender nodes. In fact, after reviewing every study paying attention to modulations in the GC effect as a function of L2 proficiency, a further explanation is provided in the form of the existence of a strong dependency of the L2 on the L1 during its acquisition. Late-learners tend to acquire L2 lemmas through inheriting the grammatical features of their L1 translation equivalents, including grammatical gender values. In fact, the higher the similarities between translations, the higher is the degree of the inheriting here. This is consistent with the fact that cognates boost the gender congruency effect as well as the interference caused by gender incongruent translations during the processing of target nouns. As MacWhinney pointed out when defining the Competition Model (1997), this '*parasiting*' state of the L2 lexicon over that of the L1 only

begins to decrease as L2 proficiency increases, as two of our reviewed studies suggest (i.e., Bordag & Pechmann, 2007; Costa et al., 2003). Since the original links to L1 lemmas and their grammatical features may persist (even when new links to the appropriate features for the L2 lemmas start to form), this causes a great deal of competition between the two languages at the level of grammatical encoding. If this explanation is right, such competition is what we have seen in the reviewed studies, although it diminishes as proficiency increases. Future studies might seek to analyse, through a longitudinal experiment, how increasing L2 proficiency affects the GC effect.

Regarding the role of the phonological gender transparency of nouns, since participants have been shown to be sensitive to L2 phonological gender cues (e.g., Bordag & Pechmann, 2007; Bordag & Pechmann, 2008), evidence from the studies reviewed here seems to support a model that allows interaction between the levels of phonological encoding and grammatical encoding in an L2. Again, this is in line with the postulates of the main connectionist models of monolingual and bilingual language processing (e.g., CMSP, BIA, BIA+, Multilink). Recent electrophysiological evidence matches these results, as bilinguals seem to be highly sensitive to gender cues since they are more helpful for them to recover gender than for monolinguals, at least when agreement is involved (Cafarra, Barber, Molinaro, & Carreiras, 2017).

Concerning the question of whether or not an agreement context is a mandatory condition for the lexical selection and processing of grammatical gender, the answer seems to be negative, at least for late-learners bilinguals. Although the GC effect has been more easily observed with NPs (gender-marked utterances apparently prompt gender processing to a greater extent, e.g., Salamoura & Williams, 2007), it has been obtained extensively with BNs (production) and unmarked utterances (comprehension). In production, this goes against the postulates of the classical activation-dependant model, the WEAVER++ (Levelt et al., 1999).

In any case, when the L2 is Dutch (and especially, when the L1 is also an opaque Germanic language such as German), the postulates of the WEAVER++ model do explain the results (null for BNs, GC for NPs). This can be interpreted as supporting evidence to the theoretical idea that grammatical gender is more of a syntactic property for opaque languages, with Dutch being the most opaque one (Sá-Leite et al., in preparation). In any case, competitive mechanisms between gender values do seem to be behind the results here. According to Morales et al. (2011), this competition is resolved through inhibition mechanisms, so that the L1 has to be inhibited to process L2's gender, at least in production. This is in line with the Inhibitory Control model proposed by Green (1998), in which a word is selected from a certain language, and the lemmas from the other language(s) (potential competitors) have to be suppressed by virtue of their language tags. In addition, as also stated by Green (1998), it will take longer to process a language that has been more suppressed, which is the case for the results in Morales et al.'s (2011) study. In their last experiment, the gender-incongruent nouns produced most in L2 during the picture-naming task were the ones whose L1 gender retrieval was most likely to fail during the post-experimental picture-naming task in the L1. Furthermore, L2 gender-congruent nouns did not inhibit gender processing on their L1 translations.

When comparing the results on bilingual gender processing with the results in the monolingual literature, a question arises. As we have said, studies with Germanic languages have failed to observe gender effects with BNs in the monolingual field, and the results are sometimes controversial either with BNs or with NPs when it comes to Romance languages (as stated in the introduction, see Finocchiaro et al., 2011, for null results). Hence, why are gender effects more easily obtained with bilinguals rather than with monolinguals in both NPs and BNs and with both kinds of languages? Bordag (2004) proposes that, since in picture-naming tasks with bilinguals the distractor is a translation of the target word, both

lemmas are closely related semantically. If the arbitrary syntactic features such as grammatical gender of nouns are part of the conceptual representation of the objects they refer to (Boroditsky et al., 2003; Konishi, 1993), when both L1 and L2 nouns have the same grammatical gender, the corresponding concepts share more features than concepts corresponding to translations with different gender. Thus, the interaction between the two languages during a picture-naming task might be not only lexical but also semantically mediated. This maximises the competition for gender selection. In the picture-naming tasks within the picture-word interference paradigm with monolinguals, the distractor and the picture have nothing in common. The hypothesis proposed by Bordag is supported by Paolieri et al.'s study (2018), in which the size of the GC effect was greater for concrete words in comparison to abstract words. Having controlled many variables in the experiment (see Table 1), the main difference between them was the amount of semantic features they had, and consequently that they could share. Note that, in a study on gender processing with Italian monolinguals and using a picture-word interference paradigm, Cubelli et al. (2005) found no interaction between the semantic relatedness and the GC between target and distractor. However, other authors did find it (Schriefers & Teruel, 2000). Therefore, Bordag's hypothesis might indeed be valid and deserves to be further examined.

Taken together, the data on gender processing (as an abstract lexical property) in bilinguals collected thus far appears to be in favour of activation-dependant yet interactive models that also consider the existence of inhibitory connections across languages. As we have seen, connectionist models generally define the selection of grammatical features such as gender as a competitive process (selection by activation). Furthermore, the CMSP, a main connectionist model of monolingual language processing, does not consider agreement context as a requirement for gender selection and processing. Importantly, connectionist models hold that different levels of language processing influence each other bi-directionally

(bottom-up and top-down) through excitatory and inhibitory connections. Thus, they state that, first, superior levels of language processing, such as syntactic processing, can affect the inferior levels of grammatical selection, something that would explain why agreement contexts may boost grammatical gender activation. Second, they predict that the processing of the lower level of the lexeme (namely the phonological properties of a noun) affects the selection at the higher level of the lemma, which would explain why phonological gender transparency affects grammatical gender selection.

Moreover, results support a connectionist model that focus on language acquisition during adulthood (or, at least, when an L1 is already established) and not only on general bilingual language processing. Specifically, predictions should be made based on: (i) a high dependency of the L2 on the L1 grammatical features during the first steps of L2 late acquisition that will later progressively fade, and (ii) cross-language effects even when bilinguals are highly proficient and are in the monolingual mode (Grosjean, 1997a, b). These are all main characteristics of a developmental adaptation of the BIA: The BIA-d model (Grainger, Midgley, & Holcomb, 2010), a connectionist yet localist model that aims to explain L2 lexical acquisition. According to this model, at least in late bilinguals, both languages are always active to some degree and co-activated representations from the irrelevant language will affect target language processing. However, the influence of the L1 on the L2 will be higher the lower the proficiency of the bilingual is, because when L2 is being acquired direct links between the L1 and their L2 translations are created, especially when these translations are cognates, sharing form and semantic information (note that the BIA-d incorporates both the postulates of the BIA (Dijkstra et al., 1998) and the RHM [Kroll & Stewart, 1994]). The lemma features of the L1 will be used in the L2 lexical entries at least until the bilingual person is proficient enough to create L2 independent linguistic nodes. The links between the L2 lexical representations and their own independent features will compete

with the direct links between the L2 lexical representations and the L1 features of their translation equivalents. As the direct links between L2 lexical representations and their own grammatical characteristics are further strengthened, the connections between translation equivalents (L1-L2) are modified as L2 lexical representations are integrated into a single lateral inhibitory network, apart from the network for words from the L1. In this way, in the presence of a word, all activated representations from L1 and L2 henceforth enter the competition for word identification, but the probability that a given word is in fact the perceived stimulus is regulated by the probability that the stimulus is a word in one or the two languages spoken by the bilingual person. Competition takes place between representations until one of them, along with its language node, is selected and the other representations are inhibited, especially the ones belonging to the other language node, which will also remain inhibited. In sum, this model includes the main characteristics of the connectionist models, which allow us to: (i) explain the competitive and inhibitory processes underlying cross-language gender congruency effects, (ii) predict the boosting of these effects when cognates and/or NPs are being processed and (iii) predict the interaction between lexeme and lemma processing during grammatical gender selection. Moreover, it also allows us to explain the effects from a developmental perspective, explaining why gender effects are found in adults even when they are in a L2 monolingual mode, and why proficiency seems to have a role in the interaction between languages at the level of grammatical encoding.

In future research, efforts to preserve a high level of experimental control are crucial since several variables have been shown to modulate noun processing (see Table 1). Indeed, some of the reviewed studies neglected critical variables such as phonological gender transparency and word frequency. For example, Manolescu and Jarema (2015) only controlled two variables: the number of syllables and the number of letters of the L2 target nouns, neglecting word frequency, transparency and familiarity, which have shown to affect

naming times (Bartram, 1973; Crepaldi, Che, Su, & Luzzatti, 2012; Gao, Li, Chou, & Wu, 2016; Sá-Leite et al., 2017). Furthermore, concreteness and/or imageability were not controlled in the picture-naming studies of Klassen (2016), Manolescu and Jarema (2015), or Paolieri et al. (2010), despite the fact that they are strong predictors of naming latencies (Crepaldi et al., 2012). Likewise, given that Weber and Paris (2004) and Morales et al. (2016) are studies with visual world tasks also involving pictures, they should have considered these factors. Another variable that has been widely ignored is the number of phonological and orthographic neighbours in the L2, although it has been shown that naming latencies for words with large neighbourhoods are shorter than latencies for words with smaller neighbourhoods (see Andrews, 1989, 1992, 1997; Sears, Hino, & Lupker, 1995). In addition, most of the studies only controlled the reviewed variables for the L2 target nouns. Although we recognise the difficulty of achieving this, some control of L1 translations, such as that in Morales et al. (2016), would have been desirable. Importantly, in picture-naming tasks, the L1 translations that participants consider for the L2 target pictures are extremely relevant, in that some L1 synonyms might vary the gender congruency between the translations. Despite this, some studies have failed to fully control for this variable (Bordag, 2004; Bordag & Pechmann, 2007; Costa et al., 2003). For instance, Bordag (2004) and Bordag and Pechmann (2007) conducted a pre-experimental familiarisation phase in which participants were shown the pictures along with the L1-L2 translation pairs they were supposed to activate while performing the naming task. However, in this way authors cannot control for which L1 translations are actually being activated during picture naming in L2. If participants use a noun they were not supposed to while naming in L2, researchers should remove such responses. However, if they are activating the incorrect L1 translation (for example, a synonym of the opposite gender to the translation noun they are supposed to activate) during the naming of a certain picture, researchers would be unaware of this and should have to

analyse the answer as gender-congruent or incongruent despite the fact that, for this participant, it was the opposite. Thus, the ideal scenario would be one in which participants are asked to name the stimuli pictures in the L1 at the end of the experimental session (after the L2 picture-naming task), so that incorrect translations indicate which trials of the L2 picture-naming task must be removed (e.g., Paolieri et al., 2010). In this regard, it is essential that word databases to be developed control translations between specific languages, with information on not only the gender congruency or incongruency between them, but also on their degree of translation ambiguity. This would include, for example, the existence of synonyms or homonyms in both translations, the degree of formal ambiguity between the translations and other nouns (e.g., *interlingual* homographs) and the degree of cognateness, among other factors.

Besides, as future areas of research, it might be interesting to replicate Manolescu and Jarema's (2015) and Klassen's (2016) studies using the highest stimuli control possible, in order to achieve further evidence on how neuter gender behaves as a mismatch gender value between structurally gender-asymmetrical languages. In this vein, it would be pertinent to look at the stability of these asymmetrical gender representations, as in Lemhöfer's et al. (2008) study. Since in quantitative terms neuter behaved different from gender incongruent pairs in Manolescu and Jarema's (2015) and Klassen's (2016) studies, perhaps a stability analysis can shed some light on this issue. This would tell us whether differences between neuter and congruent/incongruent conditions are a consequence of either 'incongruence' (i.e., its representational status as a competitive gender value influencing on-line processing) or the 'novelty' of that gender value (i.e., little experience processing that specific gender value). Also of interest would be an examination of the interaction between Romance and Germanic gender systems, which are very different. For instance, the interaction between Dutch and Spanish languages would be useful to explore, since the gender system of Dutch is

structurally, proportionally, morphologically and phonologically different from the Spanish gender system, mainly in that Dutch gender values are ‘common’ and ‘neuter’, meaning that there is no natural gender mediation, contrary to Spanish. Thus, questions such as the following arise: does a Dutch-Spanish bilingual treat grammatical gender as the same parameter in both languages? Would a ‘gender congruency’ effect emerge even when there is no gender value shared by both languages? Finally, as bilinguals have shown to be sensitive to L2 phonological gender transparency, it would be interesting to test whether, during L2 processing, the phonological gender cues of the L1 interfere with the phonological gender cues of the L2. For example, researchers might select transparent nouns in the L1 whose L2 translations are opaque or irregular and compare them to translations that are transparent in both languages. In this sense, studies could be made examining the differences on the GC effect depending of the degree of phonological gender transparency of the L1 as a whole vs. the L2 (e.g., even in transparent linguistic families such as Romance languages, differences on the GC effect might be found between a pair of highly transparent languages, as it is Italian and Spanish, vs. a pair in which a more opaque Romance language is present, such as French).

Table 1

List of reviewed studies and their description

Study	Tasks	Phrase Type	Population	Proficiency on the L2	Bilinguals AoA	Phonological Gender Transp.	Cogn.	Targets Freq.	Targets N° Syllables / Letters	Targets semantic category	Concr.	GoD	FAM	COMPL	Effect of GC
Bordag (2004)	Mixed-language picture-naming task	BNs NPs	German-Czech	Intermediate to low advanced	Late learners (AoA = 20yo)	Contr (Transparent, Opaque and Irregular)	All non-Cognates	Contr	Contr	Unc	Contr	Contr	Contr	Unc	✓
	Mixed-language picture-naming task	BNs NPs	Czech-German	Intermediate to advanced	Late learners (AoA = 14yo)	Contr (Transparent, Opaque and Irregular)	All non-Cognates	Contr	Contr	Unc	Contr	Contr	Contr	Unc	✓
Bordag and Pechmann (2007)	Mixed-language picture-naming task	BNs NPs	Czech-German	Upper intermediate to advanced	Late learners (AoA = 13.2 yo)	Contr (Transparent, Opaque and Irregular)	All non-Cognates	Contr	Contr	Unc	Contr	Contr	Contr	Contr	✓
	Picture-naming task in L2	BNs NPs	Czech-German	Intermediate to advanced	Late learners (AoA = 12.4 yo)	Contr (Transparent, Opaque and Irregular)	All non-Cognates	Contr	Contr	Unc	Contr	Contr	Contr	Contr	✓
	Mixed-language picture-naming task	BNs NPs	Czech-German	Upper intermediate to advanced	Late learners (AoA = 10.7 yo)	Contr (Transparent, Opaque and Irregular)	All non-Cognates	Contr	Contr	Unc	Unc	Unc	Contr	Unc	✓ (smaller)
Bordag and Pechmann (2008)	Forward-Translation task	BNs NPs	Czech-German	Upper-intermediate to advanced	Late learners (AoA > 9 yo)	Contr (transparent)	All non-Cognates	Contr	Contr	Unc	Contr	-	-	-	Null
	Forward-Translation task	BNs NPs	Czech-German	Upper-intermediate to advanced	Late learners (AoA > 9 yo)	Contr (Transparent, Opaque and Irregular)	All non-Cognates	Contr	Contr	Unc	Contr	-	-	-	Null
	Forward-Translation task	BNs Complex NPs	Czech-German	Upper-intermediate to advanced	Late learners (AoA > 9 yo)	Contr (Transparent, Opaque and Irregular)	All non-Cognates	Contr	Contr	Unc	Contr	-	-	-	Null
Costa et al. (2003)	Picture-naming task in L2	NP	Croatian-Italian	Nativelike	Early learners	Unc	Unc	Contr	Contr	Contr	Unc	Unc	Unc	Unc	Null

Study	Tasks	Phrase Type	Population	Proficiency on the L2	Bilinguals AoA	Phonological Gender Transp.	Cogn.	Targets Freq.	Targets N° Syllables / Letters	Targets semantic category	Concr.	GoD	FAM	COMPL	Effect of GC
	Speeded-naming task in L2	NP	Croatian-Italian	Nativelike	Early learners	Unc	Unc	Contr	Contr	Contr	Unc	Unc	Unc	Unc	Null
	Mixed-language picture-naming task	NP	Croatian-Italian	Nativelike	Early learners	Unc	Unc	Contr	Contr	Contr	Unc	Unc	Unc	Unc	Null
	Picture-naming task in L2	NP	Spanish-Catalan	Nativelike	Early learners	Unc	Unc	Contr	Contr	Contr	Unc	Unc	Unc	Unc	Null – Replicated in Control group
	Picture-naming task in L2	NP	Italian-French	Nativelike	Early learners	Unc	Unc	Contr	Contr	Contr	Unc	Unc	Unc	Unc	Null – Replicated in Control group
Klassen (2016)	Picture-naming task in L2	BNs NPs	Spanish-German	Intermediate	Late learners	Contr for L1 (transparent)	All non-Cognates	Contr	Unc	Unc	Unc	Unc	Unc	Unc	✓ (participants)
	Picture-naming task in L2	BNs NPs	German-Dutch	Intermediate to advanced	Late learners	Contr (opaque)	Contr (Cognates and non-cognates)	Contr	Contr	Unc	Contr	Unc	Unc	Unc	NPs Higher for cognates
Lemhöfer et al. (2008)	Picture-naming task in L2	BNs NPs	German-Dutch	Intermediate	Late learners	Contr (opaque)	Contr (Cognates and non-cognates)	Contr	Contr	Unc	Contr	Unc	Unc	Unc	NPs Greater for cognates
	Lexical decision task in L2	BNs (unmarked utterances) NPs	German-Dutch	Intermediate to high-proficient	Late learners	Contr (opaque)	Contr (Cognates and non-cognates)	Contr	Contr	Unc	Contr	-	-	-	✓
	Picture-naming task in L2	BNs NPs	Romanian-French	Highly proficient	Early learners (not specified)	Unc	All non-Cognates	Contr	Contr	Unc	Unc	Unc	Unc	Unc	✓
Manolescu and Jarema (2015)	Forward-Translation task	BNs	Romanian-French	Highly proficient	Early learners (not specified)	Unc	All non-Cognates	Contr	Contr	Unc	Unc	Unc	Unc	Unc	✓
Morales et al. (2011)	Picture-naming task in L2	BNs	Italian-Spanish	Upper intermediate	Late learners	Unc	All non-Cognates	Contr	Contr	Unc	Unc	Unc	Unc	Unc	✓
Morales et al. (2016)	Visual world paradigm in L2	NPs	Italian-Spanish	Upper intermediate	Late learners (AoA > 20 yo)	Unc	Contr (different degrees of)	Contr (For L1 too)	Contr (For L1 too)	Unc	Unc	Unc	Unc	Unc	✓

Study	Tasks	Phrase Type	Population	Proficiency on the L2	Bilinguals AoA	Phonological Gender Transp.	Cogn.	Targets Freq.	Targets N° Syllables / Letters	Targets semantic category	Concr.	GoD	FAM	COMPL	Effect of GC
Paolieri et al. (2010)	Picture-naming task in L2	BNs NPs	Italian-Spanish	Intermediate to highly proficient	Late learners (AoA = 21.8 yo)	Contr (transparent and opaque)	All non-Cognates phonological overlap)	Contr	Contr	Unc	Unc	Unc	Unc	Unc	✓
	Picture-naming task in L2	BNs	Italian-Spanish	Intermediate to highly proficient	Late learners (AoA = 22.2 yo)	Contr (transparent and opaque)	All non-Cognates	Contr	Contr	Unc	Unc	Unc	Unc	Unc	✓
	Forward-Translation task	BNs NPs	Italian-Spanish	Intermediate to highly proficient	Late learners (AoA = 24.7 yo)	Contr (transparent and opaque)	All non-Cognates	Contr	Contr	Unc	Unc	Unc	Unc	Unc	✓
Paolieri et al. (2018)	Forward-Translation task	BNs NPs	Russian-Spanish	Advanced	Late learners (AoA = 22 yo)	Contr (transparent L1; same n° transparent L2)	All non-Cognates	Contr	Contr	Unc	Contr (concrete vs. abstract)	-	-	-	✓ (NPs for abstracts)
	Forward-Translation task	BNs NPs	Italian-Spanish	Advanced	Late learners (AoA = 28.7 yo)	Contr (transparent L1 = transparent L2)	All non-Cognates	Contr	Contr	Unc	Contr (concrete vs. abstract)	-	-	-	✓ (Higher for concrete)
Salamoura and Williams (2007)	Forward-Translation task	BNs NPs	Greek-German	Advanced	Late learners (AoA = 10 yo)	Unc	Contr (Cognates and non-cognates)	Contr	Contr	Unc	Unc	-	-	-	✓ (NPs Greater for cognates)
Weber and Paris (2004)	Visual world paradigm in L2	NPs	French-German	Highly proficient	Late learners (AoA = 12 yo)	Unc	Contr (different degrees)	Contr	Unc	Unc	Unc	Contr	Unc	Unc	✓

Note. AoA = Age of Acquisition; Phonological Gender Transp. = Phonological Gender Transparency ; Cogn. = State of cognateness of target nouns; Targets Freq = Frequency per million of target nouns; Concr. = Concreteness; GoD = Goodness of Depiction; FAM = Familiarity of target pictures; COMPL = Complexity of target pictures; Unc = Uncontrolled; Contr = Controlled; - = The variable does not apply for this task. The AoA presented for every study is an average.

References

- Alario, F.-X., & Caramazza, A. (2002). The production of determiners: evidence from French. *Cognition*, 82(3), 179–223. doi:10.1016/s0010-0277(01)00158-5
- Andrews, S. (1989). Frequency and neighborhood effects on lexical access: Activation or search? *Journal of Experimental Psychology: Learning, Memory, & Cognition*, 15, 802-814. doi:10.1037//0278-7393.15.5.802
- Andrews, S. (1992). Frequency and neighborhood effects on lexical access: Lexical similarity or orthographic redundancy? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 18(2), 234. doi:10.3758/bf03214334
- Andrews, S. (1997). The effects of orthographic similarity on lexical retrieval: Resolving neighborhood conflicts. *Psychological Bulletin & Review*, 4, 439-461. doi:10.3758/bf03214334
- Arnon, I., & Ramscar, M. (2012). Granularity and the acquisition of grammatical gender: How order-of-acquisition affects what gets learned. *Cognition*, 122(3), 292–305. doi:10.1016/j.cognition.2011.10.009
- Audring, J. (2016). Gender. *Oxford Research Encyclopedia of Linguistics*. doi:10.1093/acrefore/9780199384655.013.43
- Babou, C. A., & Loporcaro, M. (2016). Noun classes and grammatical gender in Wolof. *Journal of African Languages and Linguistics*, 37(1), 1-57. doi:10.1515/jall-2016-0001
- Barber, H., & Carreiras, M. (2005). Grammatical Gender and Number Agreement in Spanish: An ERP Comparison. *Journal of Cognitive Neuroscience*, 17(1), 137–153. doi:10.1162/0898929052880101
- Bartram, D. J. (1973). The effects of familiarity and practice on naming pictures of objects. *Memory & Cognition*, 1(2), 101–105. doi:10.3758/bf03198077

- Bates, E., Devescovi, A., Hernandez, A., & Pizzamiglio, L. (1996). Gender priming in Italian. *Perception & Psychophysics*, *58*(7), 992-1004. doi:10.3758/bf03206827
- 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
- Bley-Vroman, R. (1990). The logical problem of foreign language learning. *Linguistic Analysis*, *20*, 3-49.
- Blom, E., Polišenská, D., & Weerman, F. (2008). Articles, adjectives and age of onset: The acquisition of Dutch grammatical gender. *Second Language Research*, *24*, 297-331.
doi:10.1177/0267658308090183
- 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., Opitz, A., & Pechmann, T. (2006). Gender processing in first and second languages: The role of noun termination. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *32*(5), 1090-1101. doi:10.1037/0278-7393.32.5.1090
- Bordag, D., & Pechmann, T. (2007). Factors influencing L2 gender processing. *Bilingualism: Language and Cognition*, *10*(03). doi:10.1017/s1366728907003082
- Bordag, D., & Pechmann, T. (2008). Grammatical gender in translation. *Second Language Research*, *24*(2), 139-166. doi:10.1177/0267658307086299
- Boroditsky, L., Schmidt, L., & Phillips, W. (2003). Sex, syntax, and semantics. In D. Gentner & S. Goldin-Meadow (Ed.), *Language in mind: Advances in the study of language and cognition* (pp. 61-80). Cambridge, MA: MIT Press.

- Breedveld, J. O. (1995). *Form and Meaning in Fulfulde: A Morphophonological Study of Maasinankoore*. Leiden: Research School CNWS.
- Caffarra, S., Barber, H., Molinaro, N., & Carreiras, M. (2017). When the end matters: influence of gender cues during agreement computation in bilinguals. *Language, Cognition and Neuroscience*, *32*(9), 1069–1085.
doi:10.1080/23273798.2017.1283426
- Caffarra, S., Janssen, N., & Barber, H. A. (2014). Two sides of gender: ERP evidence for the presence of two routes during gender agreement processing. *Neuropsychologia*, *63*, 124-134. doi:10.1016/j.neuropsychologia.2014.08.016
- 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*(3), 309-343. doi:10.1016/S0010-0277(97)00031-0
- 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*(3), 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*(3), 614–635. doi:10.1037/xlm0000065
- Corbett, G. G. (1991). *Gender*. Cambridge University Press. doi:10.1017/cbo9781139166119
- Cornips, L., & Hulk, A. (2006). External and internal factors in bilingual and bidialectal language development: Grammatical gender of the Dutch definite determiner. En C.

- Lefebvre, L. White, & C. Jourdan (Eds.), *L2 acquisition and creole genesis: Dialogues*, pp. 355-378. Amsterdam: John Benjamins. doi:10.1075/lald.42.21cor
- 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(1), 121–142. doi:10.5334/jpl.103
- Costa, A., & Caramazza, A. (1999). Is lexical selection in bilingual speech production language-specific? Further evidence from Spanish–English and English–Spanish bilinguals. *Bilingualism: Language and Cognition*, 2(3), 231-244. doi:10.1017/s1366728999000334
- Costa, A., Caramazza, A., & Sebastian-Galles, N. (2000). The cognate facilitation effect: Implications for models of lexical access. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 26(5), 1283–1296. doi:10.1037/0278-7393.26.5.1283
- 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(3), 181–200. doi:10.1017/s1366728903001123
- Crepaldi, D., Che, W.-C., Su, I.-F., & Luzzatti, C. (2012). Lexical-Semantic Variables Affecting Picture and Word Naming in Chinese: A Mixed Logit Model Study in Aphasia. *Behavioural Neurology*, 25(3), 165–184. doi:10.1155/2012/178401
- 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. doi:10.1016/j.jml.2005.02.007

- de Groot, A. M. (1989). Representational aspects of word imageability and word frequency as assessed through word association. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *15*, 824–845. doi:10.1037/0278-7393.15.5.824
- De Vincenzi, M. (1999). Differences between the morphology of gender and number: Evidence from establishing coreferences. *Journal of Psycholinguistic Research*, *28*(5), 537–553. doi:10.1023/A:1023272511427
- Degani, T., Prior, A., & Hajajra, W. (2017). Cross-language semantic influences in different script bilinguals. *Bilingualism: Language and Cognition*, *21*(4), 782-804. doi:10.1017/S1366728917000311
- Dell, G. S., Chang, F., & Griffin, Z. M. (1999). Connectionist Models of Language Production: Lexical Access and Grammatical Encoding. *Cognitive Science*, *23*(4), 517–542. doi:10.1207/s15516709cog2304_6
- Dijkstra, T. & van Heuven, W. J. B. (2002). The architecture of the bilingual word recognition system: From identification to decision. *Bilingualism: Language and Cognition*, *5*, 175–197. doi:10.1017/s1366728902003012
- Dijkstra, T., van Heuven, W. J. B. & Grainger, J. (1998). Simulating cross-language competition with the bilingual interactive activation model. *Psychologica Belgica*, *38*, 177–196.
- 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*. Published online. doi:10.1017/s1366728918000287
- Domínguez, A., Cuetos, F., & Seguí, J. (1999). The Processing of Grammatical Gender and Number in Spanish September. *Journal of Psycholinguistic Research*, *28*(5), 485-498. doi:10.1023/A:1023216326448

- Elston-Güttler, K., Paulmann, S., & Kotz, S. A. (2005). Who's in control? Proficiency and L1 influence on L2 processing. *Journal of Cognitive Neuroscience*, *17*(10), 1593-1610. doi:10.1162/089892905774597245
- Ferré, P., Sánchez-Casas, R., Comesaña, M., & Demestre, J. (2017). Masked translation priming with cognates and noncognates: Is there an effect of words' concreteness? *Bilingualism: Language and Cognition*, *20*(04), 770–782. doi:10.1017/s1366728916000262
- Finocchiaro, C., 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. *Rivista di Linguistica*, *23*(2), 161-198.
- Foucart, A., & Frenck-Mestre, C. (2012). Can late L2 learners acquire new grammatical features? Evidence from ERPs and eye-tracking. *Journal of Memory and Language*, *66*(1), 226-248. doi:10.1016/j.jml.2011.07.007
- Fuchs, Z., Polinsky, M., & Scontras, G. (2015). The differential representation of number and gender in Spanish. *The linguistic review*, *32*(4), 703-737. doi:10.1515/tlr-2015-0008
- Fung, K. T. D., & Murphy, V. A. (2016). Cross linguistic influence in adult L2/L3 learners: The case of French on English Morphosyntax. *GSTF Journal on Education*, *3*(2), 6-15. doi:10.7603/s40742-015-0011-4
- Gao, X.-Y., Li, M.-F., Chou, T.-L., & Wu, J.-T. (2016). Comparing the Frequency Effect Between the Lexical Decision and Naming Tasks in Chinese. *Journal of Visualized Experiments*, *110*. doi:10.3791/53815
- Gathercole, V. C. M., & Thomas, E. M. (2005). Minority language survival: Input factors influencing the acquisition of Welsh. En J. Cohen, K. McAlister, K. Rolstad, & J.

- MacSwan (Eds.), *Proceedings of the 4th International Symposium on Bilingualism*, pp. 852-874. Somerville, MA: Cascadilla Press.
- Gollan, T. H., & Frost, R. (2001). Two Routes to Grammatical Gender: Evidence from Hebrew. *Journal of Psycholinguistic Research*, 30(6), 627-651.
doi:10.1023/a:1014235223566
- 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
- Green, D. W. (1998). Mental control of the bilingual lexicosemantic system. *Bilingualism: Language and Cognition*, 1, 67–81. doi:10.1017/s1366728998000133
- Grosjean, F. (1998a). Transfer and language mode. *Bilingualism: Language and Cognition*, 1, 175–176. doi:10.1017/s1366728998000285
- Grosjean, F. (1998b). Studying bilinguals: Methodological and conceptual issues. *Bilingualism: Language and Cognition*, 1, 131–149.
doi:10.1017/s136672899800025x
- Harris, J.W. (1991). The exponence of gender in Spanish. *Linguistic Inquiry*, 22, 27–62.
- Hernandez, A. E., Kortz, A. S., Hofmann, J., Valentin, V. V., Dapretto, M., & Bookheimer, S. (2004). The neural correlates of grammatical gender decisions in Spanish. *NeuroReport*, 15(5), 863–866. doi:10.1097/00001756-200404090-00026
- Igoa, J. M., García-Albea, J. M., & Sánchez-Casas, R. (1999). Gender–number dissociation in sentence production in Spanish. *Rivista di Linguistica*, 11, 163–193.
- Keating, G. D. (2016). L2 Proficiency matters in comparative L1/L2 processing research. *Bilingualism: Language and Cognition*, 20(4), 700-701.
doi:10.1017/S1366728916000912

- Kempen, G., & Hoenkamp, E. (1987). An incremental procedural grammar for sentence formulation. *Cognitive Science*, *11*, 201–258. doi:10.1207/s15516709cog1102_5
- Klassen, R. (2016). The representation of asymmetric grammatical gender systems in the bilingual mental lexicon. *Probus*, *28*(1). doi:10.1515/probus-2016-0002
- Konishi, T. (1993). The semantics of grammatical gender: A cross-cultural study. *Journal of Psycholinguistic Research*, *22*, 519–534. doi:10.1007/bf01068252
- 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*(2), 149–174. doi:10.1006/jmla.1994.1008
- La Heij, W., Mak, P., Sander, J., & Willeboordsde, E. (1998). The gender-congruency effect in picture word task. *Psychological Research*, *61*(3), 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*(3), 312–330. doi:10.1016/j.jml.2008.06.005
- 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., & Christianson K. (2012). Second language sentence processing in reading for comprehension and translation. *Bilingualism: Language and Cognition*, *16*(03), 518-537. doi:10.1017/s1366728912000351
- Lyons, J. (1968). *Introduction to Theoretical Linguistics*. London: Cambridge University Press.
- MacWhinney, B. (1997). Implicit and explicit processes. *Studies in Second Language Acquisition*, *19*(02). doi:10.1017/s0272263197002076

- Manolescu, A., & Jarema, G. (2015). Grammatical gender in Romanian-French bilinguals. *The Mental Lexicon*, 10(3), 390–412. doi:10.1075/ml.10.3.04man
- 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(02), 294–310. doi:10.1017/s1366728915000176
- Mulder, K., Dijkstra, T., & Baayen, R. H. (2015). Cross-language activation of morphological relatives in cognates: the role of orthographic overlap and task-related processing. *Frontiers in Human Neuroscience*, 9(16), 1-18. doi:10.3389/fnhum.2015.00016
- Padovani, R., Calandra-Buonaura, G., Cacciari, C., Benuzzi, F., & Nichelli, P. (2005). Grammatical gender in the brain: Evidence from an fMRI study on Italian. *Brain Research Bulletin*, 65, 301–308. doi:10.1016/j.brainresbull.2004.11.025
- 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(8), 1631-1645. doi:10.1080/17470210903511210
- Paolieri, 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., Padilla, F., Koreneva, O., Morales, L., & Macizo, P. (2018). Gender congruency effects in Russian–Spanish and Italian–Spanish bilinguals: The role of language proximity and concreteness of words. *Bilingualism: Language and Cognition*, 1–18. doi:10.1017/s1366728917000591

- Pérez-Pereira, M. (1991). The acquisition of gender: what Spanish children tell us. *Journal of Child Language*, 18(03), 571-590. doi:10.1017/s0305000900011259
- Sabourin, L., & Stowe, L. A. (2008). Second language processing: When are L1 and L2 processed similarly. *Second Language Research*, 24, 397–430.
doi:/10.1177/0267658308090186
- 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(03). doi:10.1017/s1366728907003069
- Sá-Leite, A. R., Oliveira, H., Soares, A. P., Carreiras, M., & Comesaña, M. (2017, April). *Is the (in)congruency-gender effect modulated by the animacy of bare nouns? A study with native speakers of European Portuguese*. Oral presentation at the XIII International Symposium of Psycholinguistics, Braga, Portugal.
- Sá-Leite, A. R., Tomaz, A., Hernández-Cabrera, J. A., Fraga, I., & Comesaña, M. (in preparation). Gender Acquisition and Processing (GAP): An Integrative Model of Gender Processing.
- Schwartz, M., Minkov, M., Dieser, E., Protassova, E., Moin, V., & Polinsky, M. (2015). Acquisition of Russian gender agreement by monolingual and bilingual children. *International Journal of Bilingualism*, 19(6), 726-752.
doi:10.1177/1367006914544989
- Schriefers, H., & Jescheniak, J. D. (1999). Representation and Processing of Grammatical Gender in Language Production: A Review. *Journal of Psycholinguistic Research*, 28(6), 575-600. doi:10.1023/A:1023264810403
- 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. doi:10.1037/0278-7393.26.6.1368

- Sears, C. R., Hino, Y., & Lupker, S. J. (1995). Neighborhood size and neighborhood frequency effects in word recognition. *Journal of Experimental Psychology: Human Perception and Performance*, *21*(4), 876. doi:10.1037/e665412011-271
- 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*(1), 86-110. doi:10.1017/s1366728912000284
- Urrutia, M., Domínguez, A., & Álvarez, C. J. (2009). Gender activation in transparent and opaque words. *Psicothema*, *21*(1), 1-8.
- Van der Velde, M. (2003). *Déterminants et pronoms en néerlandais et en français: syntaxe et acquisition* (Doctoral dissertation), University of Paris 8, France. Retrieved from <http://www.theses.fr/2003PA082332> (May, 2017)
- van Heuven, W. J. B., Dijkstra, T., & Grainger, J. (1998). Orthographic Neighborhood Effects in Bilingual Word Recognition. *Journal of Memory and Language*, *39*(3), 458-383. doi:10.1006/jmla.1998.2584
- 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.
- Wicha, N. Y., Moreno, E. M., & Kutas, M. (2004). Anticipating Words and Their Gender: An Event-related Brain Potential Study of Semantic Integration, Gender Expectancy, and Gender Agreement in Spanish Sentence Reading. *Journal of Cognitive Neuroscience*, *16*(7), 1272-1288. doi:10.1162/0898929041920487

Acknowledgements

We would like to thank Dr. Sedy Cafarra for her helpful comments on earlier drafts of the manuscript.

This paper was funded by the Government of Spain-Ministry of Education, Culture and Sports– through the Training program for Academic Staff (Ayudas para la Formación del Profesorado Universitario, FPU grant BOE-B-2017-2646), the research project with reference PSI2015-65116-P granted by the Spanish Ministry of Economy and Competitiveness and the grant for research groups with reference GRC 2015/006 given by the Galician Government, as well as by the FCT (Foundation for Science and Technology, Portugal) through the state budget with reference IF / 00784/2013 / CP1158 / CT0013. Finally, the study has also been partially supported by the FCT and the Portuguese Ministry of Science, Technology and Higher Education through national funds and co-financed by FEDER through COMPETE2020 under the PT2020 Partnership Agreement (POCI-01-0145-FEDER-007653).



Cofinanciado por:

