Bilingualism: Language and Cognition

cambridge.org/bil

Language balance rather than age of acquisition: A study on the cross-linguistic gender congruency effect in Portuguese–German bilinguals

Ana Rita Sá-Leite[1,2], Cristina Flores[3], Carina Eira[4], Juan Haro[5] and Montserrat Comesaña[6,7]

[1]Cognitive Processes & Behaviour Research Group, Department of Social Psychology, Basic Psychology, and Methodology, University of Santiago de Compostela, Santiago de Compostela, Spain; [2]Institut für Romanische Sprachen und Literaturen, Goethe-Universität Frankfurt, Frankfurt am Main, Germany; [3]Center for Humanities, School of Arts and Humanities, University of Minho, Braga, Portugal; [4]School of Arts and Humanities, University of Minho, Braga, Portugal; [5]Department of Psychology, Research Center for Behavior Assessment (CRAMC), Universitat Rovira i Virgili, Tarragona, Spain; [6]Psycholinguistics Research LineCIPsi, School of Psychology, University of Minho, Braga, Portugal and [7]Centro de Investigación Nebrija en Cognición (CINC), Universidad Nebrija, Madrid, Spain

Research Article

Cite this article: Sá-Leite, A. R., Flores, C., Eira, C., Haro, J., Comesaña, M. (2023). Language balance rather than age of acquisition: A study on the cross-linguistic gender congruency effect in Portuguese–German bilinguals. *Bilingualism: Language and Cognition*, **26**, 1079–1092. https://doi.org/10.1017/S1366728923000378

Received: 8 March 2022
Revised: 22 March 2023
Accepted: 22 April 2023
First published online: 26 May 2023

Keywords:
age of acquisition; bilingualism; language balance; gender congruency effect; grammatical gender; proficiency

Corresponding author:
Ana Rita Sá-Leite; Email: anar.saleite@gmail.com

Abstract

The cross-linguistic gender congruency effect (GCE; a facilitation on gender retrieval for translations of the same gender) is a robust phenomenon analysed almost exclusively with late bilinguals. However, it is important to ascertain whether it is modulated by age of acquisition (AoA) and language proficiency. We asked 64 early and late bilinguals of European Portuguese and German to do a forward and backward translation task. A measure of language balance was calculated through the DIALANG test. Analyses included this factor along with the gender congruency between translations, the target language, and the AoA of both languages, among others. Results showed a GCE for European Portuguese that was independent of the AoA and greater the higher the language imbalance. We propose that changes in proficiency in any of the languages create situations of dependency between them which allow cross-linguistic gender interaction to occur and effects to emerge depending on gender transparency.

Introduction

When communicating, bilinguals have to control or even inhibit to a certain extent the language that they do not intend to produce. Interference of the non-target over the target language must be kept minimal for accurate output to be achieved. Yet, this task can be especially tricky for certain aspects of grammar. There is one feature notorious for being problematic and creating situations of long-term interference during language processing that are not easily overcome, and that is grammatical gender (Carroll, 1989; Franceschina, 2005; Hawkins, 2009).

Grammatical gender is an inherent abstract characteristic of nouns that partially determines the form of other words in speech[1]. It is present in gendered languages, which have gender systems that classify nouns according to different values. The number (and type) of gender values depends on the language itself (Corbett, 2013). For instance, in European Portuguese (EP) nouns can be either masculine (M, clock: "*relógio*"), or feminine (F, table: "*mesa*"), whereas in German nouns can also be neuter (N, room: "*Zimmer*") besides feminine and masculine. Due to the abstract nature of grammatical gender, there is no particular reason for assigning one value or another to nouns, which makes gender assignment arbitrary in terms of semantics. Because of this arbitrariness, the gender value assigned to a certain noun might differ across languages. Thus, it is easy to imagine how tricky this feature can be for bilinguals, both in terms of acquisition and of processing. Indeed, even though in EP "clock" is masculine and "table" is feminine, in German this classification is reversed and hence "clock" is feminine ("*Uhr*"), and "table" is masculine ("*Tisch*"). Other nouns, however, keep the same gender (e.g., "door", which is feminine in both EP and German, "*porta*" and "*Tür*", respectively). This cross-linguistic (in)congruency between gender values is reflected in two specific terms: heterogeneric and homogeneric translations. Heterogeneric pairs have

This article has earned badges for transparent research practices: Open Data and Open Materials. For details see the Data Availability Statement.

[1]This study does not assess natural gender, a semantic-based feature that does not comply with the abstract and arbitrary nature of grammatical gender.

different gender values and thus are gender incongruent, whereas homogeneric pairs have the same gender value and thus are gender congruent. This situation of (in)congruency evidently increases the difficulty of learning and correctly using the gender of nouns across multiple languages (Franceschina, 2005). Yet, the target-like assignment of gender is still crucial to assure accurate output in terms of agreement, as the form of other words such as articles or adjectives changes in order to agree with the gender of the head noun. For instance, in EP we would say "*O relógio caro*" (the expensive clock) but "*A mesa cara*" (the expensive table).

The repercussions that the mismatches between gender systems can have on the acquisition, representation and retrieval of gender in bilinguals have been a focus of interest of many researchers (e.g., Egger et al., 2018; Kupisch et al., 2022; Lemhöfer et al., 2008; Unsworth, 2008, among many others). In the present study, we focused on the representation and processing of grammatical gender in both early and late bilinguals. More specifically, we asked EP and German bilinguals to translate bare nouns in backward (from L2 to L1) and forward (from L1 to L2) direction in order to understand how gender values are selected during the lexical access to nouns depending on the gender congruency between equivalent translations. We did so by focusing on the so-called gender congruency effect (Sá-Leite et al., 2019, 2020). Importantly, we took into consideration two relevant individual background variables that have remained poorly assessed in the literature on gender processing by adult speakers: the age of acquisition (AoA) of the two languages and the language balance, assessed through a proficiency-based measurement of each language.

Tackling gender retrieval during bilingual noun production

The study of the representation and processing of grammatical gender during bilingual language production has attracted the attention of many scholars during the last decades. Particularly, the debate has focused on whether the gender systems of each language are represented separately in the bilingual's mind or rather, languages share a unique system (the autonomous vs. the integrative view, see Sá-Leite et al., 2020). To understand this debate, we must first comprehend how gender is represented in the linguistic system as a feature. It is widely consensual among theories of lexical selection that gender is located at the lexico-syntactic level of word representation in the form of one node per gender value (e.g., the M, F, and N gender nodes; Caramazza, 1997; Levelt et al., 1999). These gender nodes are connected to all the nouns in the lexicon as a function of their gender values, so that "*Tisch*" ("table", M) is connected to the masculine node (see Figure 1). Yet, bilinguals could either have an autonomous gender system per language, each one with its own nodes (e.g., a masculine gender node for EP and a masculine gender node for German), or rather an integrated system in which all lexical entries belonging to the same value are connected to the same gender node, regardless of the language they belong to (e.g., "*porta*" and "*Tür*" ["door"] would be connected to the same feminine gender node).

To explore this issue, researchers have relied on a specific effect, the cross-linguistic gender congruency effect (GCE) obtained mainly with two different experimental settings: picture naming and forward translation tasks. In the former, participants are asked to name pictures orally in the second language (L2, e.g., bilinguals of EP [first language, L1] and German [L2] would say "*Tisch*" when presented with the picture of a table); in the latter, participants have to orally translate to the L2 nouns written in their L1 (e.g., saying "*Tisch*" [German] when presented with the word "*mesa*" [EP]). In both tasks, conditions of gender congruency and incongruency are created through the selection of homogeneric and heterogeneric translation pairs. According to the integrative view of bilingual gender representation, variations on the response times (RTs) of the participants are expected depending on the type of translation pair. These variations on the RTs are not expected if each language has its own autonomous gender system with its own nodes. Imagine that a bilingual of EP and German wanted to produce the German noun "*Tisch*". The processing of this noun would require first the activation and selection of its semantic features, that later would activate the grammatical and syntactic information of the word associated to that concept (*tisch*) as well as its morpho-phonological realization (/tisch/)[2]. Hence, the masculine gender node would receive activation coming from the semantic representation of "TISCH" and would thus be selected.2 In the experimental context of a naming or translation task, when bilingual participants are presented with the image of a table or the word "table" in their L1 (e.g., in EP, "*mesa*"), they activate not one but two appropriate lexical entries ("*Tisch*" and "*mesa*"; e.g., Hatzidaki et al., 2011; Klaus et al., 2018). The lexico-syntactic representation of *tisch* would activate the masculine gender node, conversely to *mesa*, which would activate the feminine node. Since both gender nodes are active, they compete for selection (see Figure 2). This process of competition is mostly addressed in the WEAVER++ model of language processing by Levelt et al. (1999). According to the authors and as it has been shown in several computational simulations, for a node to be selected, activation has to reach a threshold defined by the difference of activation across nodes (the Luce ratio, see Roelofs, 1992). Hence, the greater the activation strength of a non-target node is, the harder it is for the target node to reach the difference of activation marked by the threshold for selection. Conversely, if different sources of activation converge in the same node, facilitation is observed. Yet, if gender systems were autonomous and thus separated depending on the language, convergence on the same node would not occur and facilitation would not be observed (and the same would apply for mechanisms based on competition rather than on facilitation as competition would not occur between autonomous systems).

Results from the naming and translation tasks support the integrative view of the bilingual gender representation as they reveal a consistent cross-linguistic GCE by which RTs are significantly lower for the gender congruent condition (homogeneric translations) than for the gender incongruent one (heterogeneric translations; for a review, see Sá-Leite et al., 2019). In other words, in the case of homogeneric translations, the threshold for selection is being reached more easily in comparison to when activation is spread to the opposite gender. This effect has been tested almost exclusively in research on late bilinguals (but see Fuchs, 2022, for an ongoing study on Spanish heritage speakers), focusing on their L2, and observed in picture naming tasks (Bordag, 2004; Bordag & Pechmann, 2007; Klassen, 2016; Lemhöfer et al., 2008; Manolescu & Jarema, 2015; Morales

[2]Note that how this selection occurs is a matter of discussion in the literature, especially when concerning monolingual and first language processing. Some authors argue that gender is selected automatically without the intervention of competitive mechanisms (Caramazza et al., 2001). Others argue that gender is selected competitively, but only in the presence of elements of agreement, for which gender selection is required to encode the form of the other words (Levelt et al., 1999). In this study, we will adopt a competitive view on gender selection that is however independent on the presence of elements of agreement, as this is thus far the only way we are capable of explaining the results on bilingualism (see Sá-Leite et al., 2019, 2020 for overviews).

https://doi.org/10.1017/S1366728923000378 Published online by Cambridge University Press

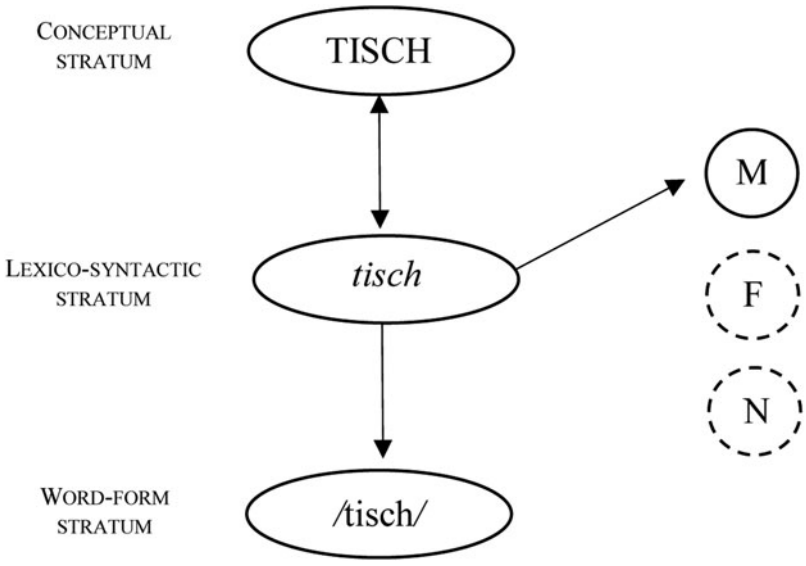

Figure 1. Representation of gender selection during lexical access in German

Note. Representation has been simplified as it intends to be illustrative of gender selection during noun production through the spread of activation. The conceptual stratum represents the abstract semantic features associated with each word, here illustrated through the English noun "table". Continuous bold lines indicate selection; discontinuous lines represent features (N) that have been neither activated nor selected. M = Masculine; F = Feminine; N = Neuter. Figure based mainly on the WEAVER++ model of lexical access (Levelt et al., 1999).

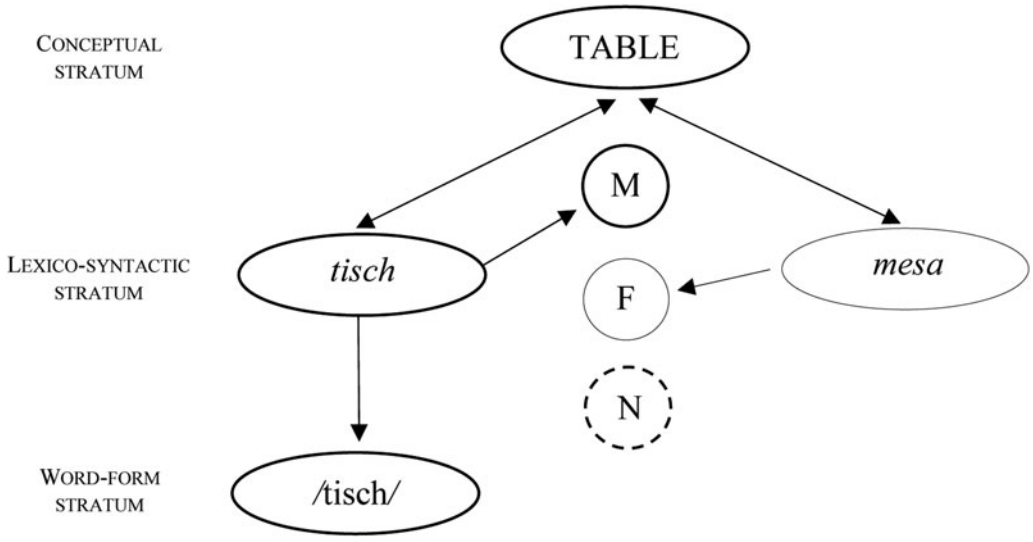

Figure 2. Production of "table" in German in a shared gender system with Portuguese

Note. Lexical access to the word "table" for a Portuguese and German bilingual (without mechanisms of inhibition or control considered). Representation has been simplified as it intends to be illustrative of gender selection and gender nodes within bilingualism during noun production. The conceptual stratum includes the abstract semantic features associated with each word, here represented by the English noun "table". Continuous bold lines indicate selection; continuous fine lines indicate activation but not selection; discontinuous lines represent features (N) that have been neither activated nor selected. Spread of activation starts on the conceptual stratum. The masculine gender node is selected, whereas the feminine gender node is activated by the lexical representation of the word "table" in Portuguese ("*mesa*"), and has hence been a competitor for selection. M = Masculine; F = Feminine; N = Neuter. Figure based mainly on the WEAVER++ model of lexical access (Levelt et al., 1999).

et al., 2011; Paolieri et al., 2010; but for null results see Costa et al., 2003), as well as in translation tasks (Bordag & Pechmann, 2008; Manolescu & Jarema, 2015; Paolieri et al., 2010, 2019; Salamoura & Williams, 2007) featuring multiple language pairs from the Germanic, Slavic, and Romance language families (i.e., Czech, Dutch, French, German, Greek, Italian, Romanian, Russian, and Spanish).

Current gaps in the literature

The cross-linguistic GCE is without a doubt highly informative when it comes to the organization of grammatical gender within the bilingual mind and appears to clearly support the integrative view. However, this scenario is likely to be more complex than what has been assumed so far. On the one hand, there are subtle differences in the outcome obtained by the above-mentioned tasks. Indeed, naming tasks seem to better detect the cross-linguistic GCE in comparison to translation tasks. This idea was first pointed out by Bordag and Pechmann (2008), who failed to obtain the effect in three translations tasks with Czech and German late bilinguals. Still, they had obtained the effect in naming tasks with participants of the same population (Bordag & Pechmann, 2007). The authors proposed that differences across tasks in the time course of gender activation were responsible

https://doi.org/10.1017/S1366728923000378 Published online by Cambridge University Press

for the absence of the effect. A recent proposal on the mechanisms underlying gender retrieval supports this idea by noting that gender effects are sensitive to the time course of lexical access because of the inherent low degree of activation of gender nodes (Sá-Leite, 2021). Indeed, in naming tasks, the activation spreads in parallel from the shared concept across languages to both L1 and L2 lemmas. So, gender nodes of one language and another are activated practically at the same time, creating an ideal situation for facilitation or competition to arise. Yet, in forward translation tasks, the word-form representation of the L1 noun that appears on the screen is activated along with its lemma earlier than that of the L2 translation equivalent. In fact, it is the activation of the L1 noun that spreads to the lemma of the L2 translation equivalent, for which the gender nodes are inevitably activated after the activation of the L1 gender node. If gender nodes accumulated high degrees of activation, this probably would not be a problem, but meta-analytic research by Sá-Leite et al. (2020, 2022) very much suggests that gender nodes accumulate low levels of activation, which in turn originates slippery experimental gender effects characterized by a high degree of heterogeneity and small sizes. In this sense, Sá-Leite (2021) explains that the level of activation would depend on the language itself. Languages like Italian, EP or Spanish have a high degree of transparency as more than 60% of their nouns end in the quite simple ortho-phonological gender cues "-a" (for feminine nouns) and "-o" (for masculine nouns). Many studies have shown that transparent nouns are more accurately processed and require more cognitive resources to be accessed (Caffarra et al., 2014). Following the Dual-Route model of gender retrieval by Gollan and Frost (2001), transparent nouns seem to rely on an extra route of gender selection besides the lexical memory-based route: the form-based route represented by the noun ending. This does not seem to be the case for languages like French, German, or Dutch, which rely on an extremely complex and sometimes contradictory body of gender-form regularities and are hence considered less transparent for gender (Kupisch et al., 2018). For instance, in German, Köpcke (1982) and Köpcke and Zubin (1983) found a quite high number of gender regularities in monomorphemic and extended monomorphemic nouns (44 of which are not coincident with any gender morphemes), which, in addition, depend on the case (e.g., nominative vs. dative), as gender intertwines with declension. Consequently, regularities in opaque languages are not as useful as the typical "-a" and "-o" of transparent languages and the retrieval of gender seems to rely mainly on one source: the lexical memory-based route. Sá-Leite (2021) proposes that the existence of an extra source of gender activation for most nouns in gender-transparent languages increases the resting level of activation of gender nodes. Ultimately, in these cases, gender selection involves higher levels of activation, which might make gender competition for selection (e.g., masculine vs. feminine) more easily observed. To be precise, bilinguals are the population in which genuine effects of gender congruency are most consistently obtained, even for languages with low degrees of transparency (Bordag & Pechmann, 2007; Lemhöfer et al., 2008; for a review see also Sá-Leite et al., 2019). This should not surprise us, as bilinguals have a systematic double source of gender activation due to the activation of equivalent translations. Yet, we may note that when the time course of activation of both competing nodes differs, effects may become slippery, especially when less transparent languages such as German are involved. To be precise, null effects within the area of the cross-linguistic GCE have been obtained for Germanic languages in translation tasks but not in naming tasks (Bordag & Pechmann, 2008; Salamoura & Williams, 2007). In sum, the slipperiness of the results obtained with translation tasks depending on the transparency of the language is still a matter to be explored.

On the other hand, all the studies on the cross-linguistic GCE have focused on the impact of the L1 on the L2, but not of the L2 on the L1, even though there is increasing evidence suggesting that the acquisition and presence of an L2 within the linguistic system may modulate the representation and processing of the L1 (e.g., Hamann et al., 2017; Ulbrich & Ordin, 2014). More specifically, evidence is quite robust for long-term residents of the L2 environment who achieve near-native L2 proficiency (Schmid, 2009). Thus, a comprehensive view of the bilingual gender selection requires a thoughtful examination of language interference when gender is retrieved for the L1 and not only for the L2. In this sense, two variables that have been shown to be critical for language processing in bilingualism remain mainly unexplored in this area: AoA and language proficiency.

Regarding the AoA, most of the studies on the cross-linguistic GCE have tested participants that acquired their L2 later in life (after the age of 10 – Bordag, 2004; Bordag & Pechmann, 2007, 2008; Klassen, 2016; Lemhöfer et al., 2008; Morales et al., 2011; Paolieri et al., 2010, 2019; Salamoura & Williams, 2007). We believe this bias for late learners should be broken in order to draw more precise conclusions on the way gender is organized and selected in bilinguals. Indeed, only two published studies have tested early bilinguals (Costa et al., 2003; Manolescu & Jarema, 2015). Costa et al. (2003) conducted picture naming tasks with speakers of Croatian and Italian, Spanish and Catalan, and Italian and French that had acquired their L2 after an average age of five. However, the study used small samples ranging from 10 to 22 participants and was appointed with multiple methodological flaws (for an overview, see Sá-Leite et al., 2019; see also Lemhöfer et al., 2008). The other study that previously tested early bilinguals was that of Manolescu and Jarema (2015), who unfortunately provided scarce information regarding the linguistic background of their participants. They only stated that they were children of Romanian parents that moved to Montreal and started acquiring French during "childhood", which is not particularly specific. In sum and due to the scarcity of research with early bilinguals, thus far conclusions on the organization of the bilingual gender system derived from research with late learners cannot be generalized to early learners. Hence there is a need to test early learners of an L2 or even simultaneous bilinguals, assuring a detailed assessment of their linguistic background, so that we can understand gender retrieval in bilingual language production in a broader sense.

Finally, language proficiency is perhaps the variable that concerns us the most. As pointed out by Sá-Leite et al. (2020) in their meta-analysis on the cross-linguistic GCE, proficiency in the L2 has never been included as a factor in the analyses of any study on this effect. It has been, however, controlled in many different ways: informal interviews that took place prior to the experiment (e.g., Bordag & Pechmann, 2007), different self-informed subjective questionnaires made by the authors (e.g., Costa et al., 2003; Lemhöfer et al., 2008; Morales et al., 2011; Paolieri et al., 2010, 2019) – in some cases the authors did not report any information about what the participants were asked, which skills were considered, or what scores were obtained in average (e.g., Manolescu & Jarema, 2015) –, self-informed standardized questionnaires such as Hermans et al.'s (1998) or Bachman and Palmer's (1989; e.g., Bordag & Pechmann, 2007; Salamoura & Williams 2007), ratings

https://doi.org/10.1017/S1366728923000378 Published online by Cambridge University Press

in measures such as the familiarity of the target nouns (e.g., Bordag & Pechmann, 2007), and official language tests that had been successfully completed by the participants (although it is not said when – e.g., Bordag, 2004; Salamoura & Williams, 2007). To our knowledge, only Klassen (2016) seems to have used an objective measure of proficiency (the proficiency test of the Goethe-Institut, 2010). In sum, most studies have exclusively used subjective and non-standardized ways of measuring L2 proficiency. Likewise, when checking the self-informed data on the L2 AoA, L2 age of first exposure, and L2 time of exposure across all studies, Sá-Leite et al. (2020) noticed great variations across studies for populations that were said to have the same level of proficiency. This might be another index of a subjective and non-precise way of understanding what constitutes proficiency.

It is certainly quite baffling that a variable that has been shown as extremely relevant for bilingual language processing in many other areas (see, for instance, Bultena et al., 2015; Lim & Christianson, 2015; Prior et al., 2007) has been mostly neglected in this specific area. In fact, we believe that the language proficiency of the participants should not only be properly controlled and reported in the subsequent works on bilingual gender processing but should also be included in the analyses. The literature on other areas of bilingualism shows enough evidence to suspect it may have a great role in determining the way languages interact and hence in how gender is selected in one language or another (Kupisch et al., 2013; Soares et al., 2019). More precisely, it is essential to test bilinguals' proficiency in both of their languages and to assess the balance between them, i.e., their relative language dominance. We therefore ought to (1) use more objective measurements to test proficiency and (2) include this variable in the analysis, along with gender congruency, to truly understand its role in the effect.

Theoretical background for the role of AoA and proficiency in gender retrieval

We will now present a proposal on how AoA and proficiency may impact grammatical gender representation and processing. We will base ourselves on the idea of relative language dominance (henceforth: language balance), which refers to the degree of balance between the languages of the bilingual speaker. To do so, we will rely on two popular and highly supported models of bilingual lexical processing: the developmental Bilingual Interactive Activation model (BIA-d, Grainger et al., 2010) and the Multilink (Dijkstra et al., 2019). The BIA-d model is an extension of the connectionist but also localist BIA model (Grainger & Dijkstra, 1992) that offers an interesting view on the adaptations that our linguistic system experiences during the acquisition of an L2. The Multilink, however, is the most recent connectionist model of bilingual language production and comprehension. By combining both models (and also considering that language learning involves an on-going fine-tuning of the learning system that continues across life span, see Ramscar et al., 2014; also Chuang & Baayen, 2021), we believe we can address the dynamics of a flexible system susceptible to changes in proficiency depending on variations in word frequency of exposure and use.

Following the BIA-d model, as well as extensive empirical evidence on the matter, when the L2 proficiency increases, the L2 dependency on the L1 decreases, and hence the ability of the L1 to create interference during L2 processing decreases as well (Abutalebi & Green, 2007; Pivneva et al., 2012). Indeed, it is known that the higher the L2 proficiency of the bilingual, the better their performance during L2 production at different levels (Costa & Caramazza, 1999; Kroll et al., 2002; Pivneva et al., 2012). The tenets of the Multilink are in line with this: as the frequency of encounters with words increases, the links between their lexico-syntactic representations (i.e., lemmas) and their grammatical features become stronger. We hypothesize that this would also affect grammatical gender and, hence, the stronger the link between the L2 lemma and the gender node, the easier the retrieval of gender and the less the dependency on the L1 gender node as well as its ability to interfere. Consequently, effects of cross-linguistic gender interaction such as the GCE should be smaller, the higher the proficiency in the target language.

Regarding the AoA, it is still an open question whether early bilinguals have a better language control (Berken et al., 2016, 2017; Bonfieni et al., 2019; Wattendorf et al., 2014), yet what seems clear is that other kinds of factors, such as daily exposure to each language or changes in the language environment, can modulate the interaction of both languages during language processing regardless of when the L2 was acquired (for a review, see Van Hell & Tanner, 2012; see also Bonfieni et al., 2019). This occurs largely because these factors directly influence the degree of proficiency of the speaker in each language (e.g., Dussias & Sagarra, 2007; Levy et al., 2007; but for more detail, see Van Hell & Tanner, 2012). Therefore, whether or not a cross-linguistic GCE is observable should depend mainly on the degree of proficiency in each language, and being an early bilingual should not be an impediment to language interaction happening (for a study in language comprehension that supports this prediction, see Paolieri et al., 2020). In fact, many studies assume that AoA effects may be leveled by increasing proficiency (see Gagarina & Klassert, 2018). We thus believe that modulations in proficiency of either of the two languages should produce changes in the strength of the links connecting grammatical gender to lemmas. The changes in strength create states of dependence or interference of one language over another, somehow regressing in the phases of acquisition defined by the BIA-d model (see our proposal in Figure 3). In this sense, note that both the L2 or the L1 could suffer changes in their representational state, depending on these fluctuations in proficiency, in line with previous evidence (e.g., Dussias & Sagarra, 2007; Guo et al., 2011; Linck et al., 2009; see also Morales et al., 2014). Thus, we do not expect AoA to be a better predictor of the GCE than proficiency.

The present study

In the present study, we aimed to: (1) explore the cross-linguistic GCE not only in an L2 but also in an L1 by assessing EP and German bilingual adult speakers to understand how cross-linguistic interference might modulate gender retrieval as a whole; (2) test bilinguals who speak languages with different degrees of transparency (a more transparent [EP] and a more opaque [German] language) to examine if the degree of gender transparency of the languages may have any impact on the cross-linguistic GCE; (3) for the first time test the role of AoA and language balance in the effect. To do so, we conducted translation tasks as is traditionally done in this field of research. However, instead of using only forward translation tasks as it has been previously done in the literature (Bordag & Pechmann, 2008; Manolescu & Jarema, 2015; Paolieri et al., 2010; Salamoura & Williams, 2007) we also used backward translation tasks, which allowed us to explore the effect of one language on another and vice-versa during language production. We recruited participants

https://doi.org/10.1017/S1366728923000378 Published online by Cambridge University Press

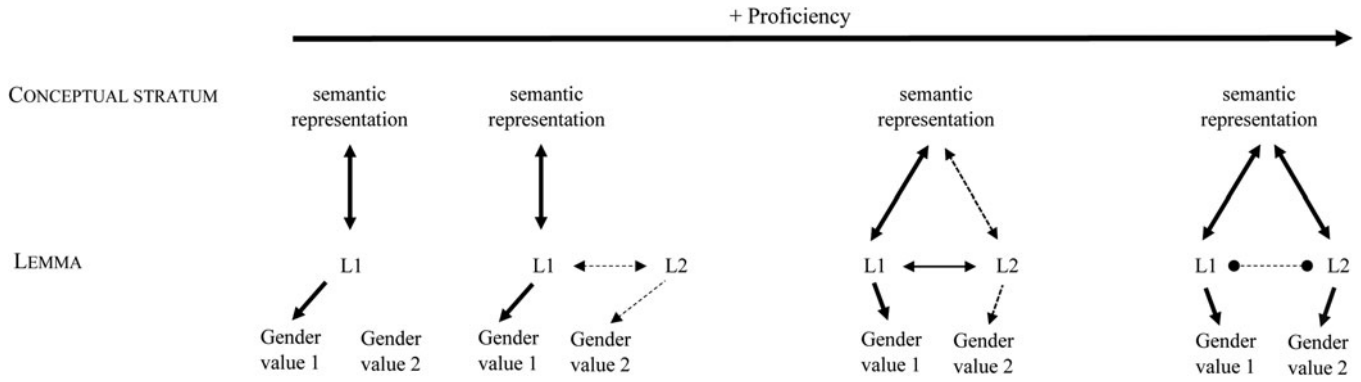

Figure 3. How L2 gender representation develops during acquisition following the BIA-d model
Note. L1 = First language; L2 = Second language. Discontinuous lines represent weak connections. The thinner the line, the weaker the connection. In our predictions, the representational state of the linguistic system may vary depending on the proficiency of one language or another. Figure based on Grainger et al. (2010).

from two different populations: (a) the so-called heritage speakers, i.e., early bilinguals of EP and German that had Portuguese parents and were born or lived during childhood in a German-speaking country (either Germany or Switzerland) with EP as home language, and (b) monolingual-raised native speakers of EP who started to learn German after the age of 10 in a classroom setting. The early bilinguals differed in the number of years they had been in German-speaking countries, with some living there even during adulthood and others leaving during childhood. At the moment of testing, 29 early bilinguals lived in Portugal, 12 in Germany and one participant in the German-speaking part of Switzerland. All early bilinguals have in common having acquired German in childhood through immersion before the age of 7 (following Hyltenstam & Abrahamsson's [2003] cut-off of early bilinguals). Consequently, they differed in their degree of proficiency in each language, suffering certain imbalances between languages due to varying proficiency in either EP or German (from now on, L1 and L2).[3] According to our proposal, for some participants, the strength of the links between the gender nodes and the L1 lemmas should have likely decreased, creating a situation of possible interference between languages, even though they were early bilinguals and highly proficient in the L2. We hence considered not only the proficiency of the L2 but also that of the L1. More specifically, we used a standardized and highly valid and reliable measurement of the degree of proficiency in both languages, the DIALANG Vocabulary Size Placement Test (see Alderson, 2005). We obtained a measure of language balance by subtracting the DIALANG score in one language from the score in the other, thus obtaining a differential-based dominance index, as suggested by Birdsong (2015).

We therefore tested the following hypotheses: (a) the cross-linguistic GCE is dependent on the language balance, so that the higher the imbalance, the greater the effect; (b) the effect can be obtained in both an L1 (EP) and an L2 (German), even though it might be modulated by the opaqueness of the Germanic language, being null for this language as obtained in previous studies with translation tasks assessing German as L2 (Bordag & Pechmann, 2008; Salamoura & Williams, 2007); (c) hence the effect does not depend on the AoA. The data and scripts used in this study are available online at the following link: http://doi.org/10.17605/OSF.IO/UE9XH

Method

Participants

Seventy-four voluntary adult bilinguals of EP and German (62 female; $M_{age} = 38.12$ years, $SD = 9.73$) were recruited online via email and social media and personal contact was made with each one of them. The requirements for participation for late bilinguals were (1) having started to learn German in a classroom context as teenagers or adults; and (2) having studied German for 5 years or more or having lived in a German-speaking country as adults. As for early bilinguals, the requirements were (1) being born or having immigrated to a German-speaking country before age 7 (Hyltenstam & Abrahamsson, 2003); and (2) having lived there for 6 or more years. Ten of them reported moderate to high proficiency in another gendered language apart from these languages in the Language History Questionnaire (LHQ, Li et al., 2020). Those were French, Spanish and Italian.[4] In addition to AoA, country of residence and years living in a German-speaking country, participants were also asked to self-rate their proficiency in all languages they knew and to estimate their degree of contact with each language in their daily life (by dividing 100% of contact among all relevant languages). The early bilinguals (n = 42) grew up as Portuguese-descendant second generation immigrants in Germany or in the German-speaking part of Switzerland. They started to acquire EP from birth as heritage language, i.e., as the main language spoken within the family. As is typical for heritage speakers, contact with the majority language (German) started either from birth or during pre-school age.[5] Thus, all speakers were either

[3]All early bilinguals acquired EP as their heritage language, which is an L1 (Flores, 2015). For the majority of these speakers German (the societal language) is an early L2, although some had at least some contact with German from birth since they were born in Germany (n = 21). For the sake of simplicity, we will refer to EP as the L1 and to German as the L2 throughout the text, especially since we treated the AoA as a continuous variable in the model of analysis (see Results section).

[4]We evaluated the effect of knowledge of a gendered language (moderate to high proficiency) on our models. However, we found that this variable had no impact on the outcome. The main effect of this variable was not significant, and the significance of the other effects remained unchanged. This means that the results that were statistically significant before remained significant (all *p*s < .05) and the ones that were not remained unchanged (all *p*s > .05). Based on these findings, we determined that this variable has no influence on the results and therefore we did not include it in the models presented in the manuscript.

[5]No participant from Switzerland spoke Swiss German within the family; all acquired standard German in pre-school age.

https://doi.org/10.1017/S1366728923000378 Published online by Cambridge University Press

simultaneous or early successive bilinguals who became German-dominant in childhood.[6] Due to various reasons (remigration, changes to the family constellations, professional reasons, among others), the degree of contact with either EP or German in daily life is diverse across speakers.

The late L2 learners started to acquire German in a classroom setting after age 10. All late learners are highly proficient in German, either because they studied German at the university or moved to Germany or Switzerland for professional or personal reasons. At the moment of testing, 7 participants were living in Germany, 2 in a German Swiss canton, 12 in Portugal and one participant moved recently from Germany to Northern Spain (Galicia).[7]

All participants signed informed consents for experimentation with human subjects previously approved by the Ethics Council of the University of Minho (CEICSH 120/2020) through Google Forms.

Materials

Measurement of language balance

We assessed proficiency in each language through the DIALANG Vocabulary Size Placement Test (VSPT, version 1) for EP and German. Note that the lexical competence has been shown to be a reliable predictor of language proficiency (Bonvin et al., 2021; Laufer & Nation, 1999; Treffers-Daller & Korybski, 2016), since the learners' lexical knowledge grows when proficiency increases, and adequate lexical knowledge is a prerequisite of effective language use. The DIALANG VSPT is a questionnaire that assesses lexical competence through a list of 75 words, of which 50 are real words and 25 are pseudo-words (for a more detailed explanation of the concept of language proficiency and balance as well as of supporting evidence of the DIALANG VSPT as a reliable proficiency indicator see Appendix S1). Participants were requested to indicate whether or not each word was an existing word in EP or German. Following Alderson (2005), the test score was computed based on the total of words correctly identified as either real words or pseudo-words. A measurement of language balance was obtained by calculating between-language subtractive differentials, i.e., the score obtained for German was subtracted from that of EP. Negative values indicate higher proficiency in EP and positive values in German (e.g., German: 62 - EP: 72; dif: -10). Values close to zero indicates high language balance.

Complementing this, we asked participants to self-assess their proficiency in Portuguese and German, on a scale from 1 to 7 in speaking and writing. For the quantification we added the ratings for both skills and obtained a total self-assessment score on a scale from 1 to 14 for both languages. We then computed the differential between both language scores to define a value for self-assessed relative proficiency. Again, negative values indicate higher self-estimated proficiency in Portuguese and positive values in German (e.g., Portuguese speaking: 7 + Portuguese writing: 6 = 13; German speaking: 5 + German writing: 5 = 10, dif: -3). A positive strong correlation between the language balance score obtained through the DIALANG VSPT and the self-assessment scores will further support the reliability of the DIALANG VSPT as a proficiency test (for more detail, check footnote 6 in the Results section; see also Flores et al., 2022).

Stimuli

We selected 180 EP inanimate nouns from the P-PAL database (for all the stimuli, see Appendix S2; Soares et al., 2018) and translated them into German. They were selected by taking into account the gender value in both languages, so that we had the same number of stimuli in each of these 6 translation types: heterogeneric feminine-masculine ("abóbora" [F] in EP, "Kürbis" in German [M], "pumpkin"), heterogeneric masculine-feminine ("journal" [M] in EP, "Zeitung" in German [F], "newspaper"), homogeneric feminine ("cenoura" in EP, "Karotte" in German, "carrot"), homogeneric masculine ("bosque" in Portuguese, "Wald" in German, "forest"), feminine-neuter ("perna" in EP, "Bein" in German, "leg"), and masculine-neuter ("carro" in EP, "Auto" in German, "car"). We avoided EP nouns with more than one German translation that had similar frequencies of use according to SUBTLEX-DE (Brysbaert et al., 2011). Besides, we avoided nouns in German that could also be verbs (e.g., "Leben" [life/to live]), nouns with high positive or negative affective valence related to death or sexuality (e.g., corpse, death, penis, etc.), German nouns that were cognates in English (e.g., "Butter"), and nouns that had more than one possible translation, when these diverged in gender within that language (e.g., "Miete" [rent, F] in German translates to "renda" [F] or "aluguer" [M] in EP). In terms of ortho-phonological gender transparency, we did not select any irregular nouns in EP, and included 144 transparent nouns and 36 opaque nouns evenly distributed across the six translation types (24 transparent nouns and 6 opaque nouns per type). We created two different blocks as a function of translation direction (EP to German and vice-versa), each composed of 90 stimuli, so that the presentation of both blocks was counterbalanced across participants. All EP nouns were controlled through a one-way ANOVAs across the 6 translation types for per million and logarithmic frequency, number of phonological and orthographic neighbours, number of letters, and mean logarithmic bigram frequency, taken from P-PAL (all $ps > .247$, Soares et al., 2018), logarithmic frequency, taken from SUBTLEX-PT ($p = .544$, Soares et al., 2015), and subjective frequency, concreteness, and imageability, taken from the Minho Word Pool (all $ps > .525$, Soares et al., 2017). See mean values in Table S1 of Supplementary Materials.

German nouns were controlled across the six translation types through a one-way ANOVA for absolute logarithmic frequency, number of letters, logarithmic number of neighbours based on the Levenshtein distance, initial logarithmic bigram frequency (normalized), and familiarity, as taken from dlexDB (all $ps > .091$; Heister et al., 2011), and logarithmic frequency as taken

[6]Research has shown that in early stages of language development simultaneous and early successive bilingual language acquisition may show developmental differences (e.g., Meisel, 2008); but these differences are overcome in older ages so that the language competence of simultaneous and early L2 speakers of a given language may become indistinguishable at least at adolescence (Montrul, 2016). Flores (2020), for instance, did not find any AoA effects in the competence of the Portuguese–German speakers analysed in her attrition study. Since the early bilinguals analysed in the present study are adults, who lived for an extended period of time in a German environment, there is no empirical support/evidence to further separate simultaneous from early acquirers of German. In fact, a clear separation between both acquisition types is typically not possible in heritage speakers who were born in the host country because it is hard (almost impossible) to determine the exact onset of exposure to the majority language of immigrant infants who are raised in a minority language environment (Montrul, 2016).

[7]The location where the participants were recruited (i.e., their place of residence) was included in the final model, but it had no significant effect ($p > .05$). The other effects in the model remained unchanged. We also compared the models that included this variable with those that did not and found that none of the comparisons were significant (all $ps > .05$). As a result, this variable was not included in the models presented in the manuscript.

https://doi.org/10.1017/S1366728923000378 Published online by Cambridge University Press

from SUBTLEX-DE (Brysbaert et al., 2011; $p = .29$). See Table S2 of Supplementary Materials for means and standard errors. Translation pairs across translation types were controlled for equivalent measures – namely, number of letters, and subjective frequency/familiarity ($ps > .563$). Although logarithmic frequency (SUBTLEX-PT and SUBTLEX-DE) showed significant differences, these differences were not between the conditions that subsequently showed significant results – namely, gender incongruency vs. congruency ($ps > .143$). The translations were also controlled for orthographic overlap using the NIM database (Guasch et al., 2013) and phonological overlap using the PHOR-in-One database (Costa et al., 2021; all $ps > .174$). See Table S3 of Supplementary Materials for means and standard errors of overlap measures.

Conditions regarding the stimuli were created taking into consideration the factors of Gender Congruency (gender congruent, gender incongruent), Target Gender (masculine, feminine), and Target Language (EP, German). Note that target gender and target language refer to the gender of the noun to be produced, and the language to be produced, respectively. Importantly, the factor of gender congruency included the four experimental conditions: heterogeneric masculine and feminine (incongruent) and homogeneric masculine and feminine (congruent) nouns. This allowed us to make a direct comparison between the gender systems of EP and German. Neuter gender in German would constitute a third category, in which rather than gender incongruency there is a situation in which one gender value does not exist in the other language, and this may change the representation of that value and the interaction between languages when it comes to its retrieval (for more information on the bilingual representation of differing gender nodes such as that of German and Spanish bilinguals in regard to the neuter node, see Klassen, 2016, and Klassen et al., 2022) . The scope of our study is to replicate the cross-linguistic GCE and test the activation of gender nodes and the processes of competition that may arise between them depending on language proficiency. These aims can be fulfilled with heterogeneric and homogeneric nouns, avoiding a greater degree of complexity in the experimental design, as we did. Nevertheless, neuter nouns were included among our stimuli in order to avoid artificial contexts that might in some way influence the performance of our participants.

Procedure

The experiment was conducted online due to the public international health emergency caused by the COVID-19 pandemic. We followed Burke and James' (2006) recommendations for online research and data collection.

Participants started by filling out the LHQ (Li et al., 2020) using Google Forms. This questionnaire allowed us to explore their linguistic background and check their knowledge of any other gendered language. Afterwards, links were sent for the EP and German versions of the DIALANG, a standardized lexical test to objectively assess the proficiency in each of these languages while guaranteeing a high degree of validity and reliability (Alderson, 2005). The task was timed through the Google Add-on *Quilgo,* which, in addition to the timer, set for 7 minutes in total, allows for screen tracking, thus informing us if the participants kept focused on the task. Only one participant was excluded due to unfocussed participation.

Participants then received a link that opened the experiment in a browser. The experiment was programmed using the JavaScript library jsPsych (de Leeuw, 2015). Two blocks of 90 nouns each (90 in EP and 90 in German) were created, but its order of presentation and the language to be produced (i.e., the target language) was counterbalanced, so that we had four different links depending on these two factors. After clicking on the link, the participant first read the instructions regarding the procedure for the whole experiment. These instructions appeared in the target language of the first sub-block. Then, a familiarization phase started in which the 90 translation pairs were sequentially presented, one by one. Participants controlled the presentation of the stimuli using the spacebar. The aim of the familiarization phase was to decrease mistranslations and non-responses. After that, participants tested their microphone, following instructions on the screen, and once checked, instructions appeared for the translation of the first block. They were asked to translate each noun into the target language as fast as possible, avoiding mistakes and trying to speak loudly and clearly into the microphone. Upon starting, participants went through a session of eight training items (different from the experimental items), then the experimental trials started. Each experimental trial had the following structure: a fixation point (+) at the centre of the computer screen, for 500 ms; the target noun for 3,000 ms or until response; a blank space for 500 ms as an inter-trial interval. Trials were presented randomly per participant. For the second sub-block, instructions appeared in the other language (which would be the new target language). The familiarization phase and translation task occurred again with the remaining 90 nouns. Responses were recorded and saved in a private directory on the University of Rovira i Virgili server. RTs were calculated offline from the presentation of the noun to be translated to the onset of the translation response using the PRAAT software (Boersma & Weenink, 2018).

The session for each block lasted approximately 20 min.

Results

We removed the RTs of incorrect responses (18.04% data points), RTs above 3,000 ms and RTs that exceeded 2.5 *SD* of each participant's mean (2.35% data points). We also removed the data from ten participants that made more than 40% of incorrect responses. Hence, the final sample was composed of 64 participants. Table 1 gives an overview over the predictor variables for these 64 participants – namely, AoA, self-rating in each language, self-assessed relative proficiency, proficiency in each language and language balance score.

RTs were analysed using linear mixed-effect models (e.g., Baayen, 2008; Baayen et al., 2008). To this end, we used the lme4 package of R (Bates et al., 2015). We created a fixed structure model to examine the hypotheses of the study, with the inverse of RTs (-1000/RT) as the dependent variable. As fixed effects, the model included the triple interaction and second-order interactions between Gender Congruency (GC, GI), Target Language (language that was produced: Portuguese or German) and absolute Language Balance Score (through the DIALANG)[8], the triple interaction and second-order interactions between Gender Congruency, absolute Language Balance Score and Age of Acquisition (AoA) of the German Language, the interaction

[8]It should be noted that the self-assessed relative proficiency differential and the DIALANG differential were highly correlated, $r = .79$, $p < .001$. We therefore decided to introduce the DIALANG differential in the models, instead of the self-assessed, as it is a more objective measure of the participants' language proficiency.

https://doi.org/10.1017/S1366728923000378 Published online by Cambridge University Press

Table 1. Sociolinguistic background of the 64 analyzed participants and DIALANG results

	AoA of seven or earlier (n = 42)				AoA higher than ten (n = 22)			
	M	SD	Range		M	SD	Range	
			min	max			min	max
Age	38.45	9.54	18	55	38.27	10.22	21	63
Average AoA	2.1	2.01	0	7	15.77	2.67	10	25
Self-rating Portuguese	12.55	1.82	6	14	13.55	1.14	9	14
Self-rating German	12.48	1.86	6	14	10.55	2.39	6	14
Self-assessed relative proficiency	−0.07	1.63	−5	3	−3	2.64	−8	3
Average European Portuguese proficiency (Dialang)	64.1	4.7	51	72	67.68	3.23	62	75
Average German proficiency (Dialang)	68	5.04	52	75	61.36	7.46	49	71
Language balance score (Dialang)	3.88	6.98	−17	20	−6.32	7.73	−26	4

Note. M = mean; SD = standard deviation; min = minimum; max = maximum. Negative values indicate higher proficiency in Portuguese in Self-assessed relative proficiency and Language Balance Score (DIALANG).

between Block Order (first and second) and Target Language, and, finally, Target Gender (gender of the noun that was produced: Feminine or Masculine). Continuous variables were centered and transformed into Z-scores. In addition, following the guidelines of Schad et al. (2020), all dichotomous variables were coded using sum contrast coding (-0.5 for the first level and +0.5 for the second level of each factor); Gender Congruency: GC (-0.5), GI (+0.5); Target Language: GER (-0.5), PT (+0.5); Target Gender: FEM (-0.5), MASC (+0.5), and Block Order: first (-0.5), second (+0.5). We also examined the multicollinearity of the fixed effects introduced in the model (R VIF function). All VIF values were less than 3, suggesting non multicollinearity (Zuur et al., 2010).

Participants and words were included as grouping factors for random effects. We followed a maximal random-effects structure (Barr et al., 2013) by adding as random slopes the most complex structure that allowed convergence. We incorporated Target Language and Gender Congruency into the random slope of participants, and Target Gender into the random slope of words. The structure of the models for evaluating the RTs in Portuguese and German naming was the same as above, but excluding Target Language for participants.

The significance of interactions was determined using log-likelihood ratio tests (R ANOVA function). We assessed the contribution of each interaction by comparing a model that included them with another model in which they were not included. We also report the results of the t-test analyses for the coefficient estimates of fixed effects and interactions. To this end, we used Satterthwaite's approximations to the degrees of

Table 2. Results of the linear mixed-effects model

Predictors	Estimate	SE	t	p
Intercept	−0.93	0.01	−64.86	<.001
Gender Congruency	0.01	0.02	0.40	0.691
Target Gender	0.01	0.01	0.90	0.371
Target Language	−0.02	0.01	−3.78	<.001
Language Balance Score	0.01	0.01	0.64	0.523
AoA German	−0.00	0.01	−0.23	0.815
AoA German * Gender Congruency	−0.01	0.00	−1.13	0.259
AoA German * Language Balance Score	−0.01	0.01	−0.73	0.464
AoA German * Language Balance Score * Gender Congruency	0.00	0.01	0.33	0.743
Gender Congruency * Target Language	−0.01	0.01	−0.57	0.567
Gender Congruency * Language Balance Score	0.01	0.01	1.24	0.217
Language Balance Score * Target Language	−0.01	0.01	−0.94	0.345
Gender Congruency * Language Balance Score * Target Language	0.02	0.01	2.12	0.034

Table 3. Mean RTs and standard errors

Target Language	Target Gender	Gender Congruency	Mean RTs	*SE*
GER	F	GC	1121	10.5
GER	F	GI	1130	11.1
GER	M	GC	1161	11.4
GER	M	GI	1152	10.6
PT	F	GC	1092	10.1
PT	F	GI	1124	10.5
PT	M	GC	1138	10.8
PT	M	GI	1114	11.3

Note. Results reported for the conditions of Gender Congruency taken into consideration Target Language and Target Gender.

https://doi.org/10.1017/S1366728923000378 Published online by Cambridge University Press

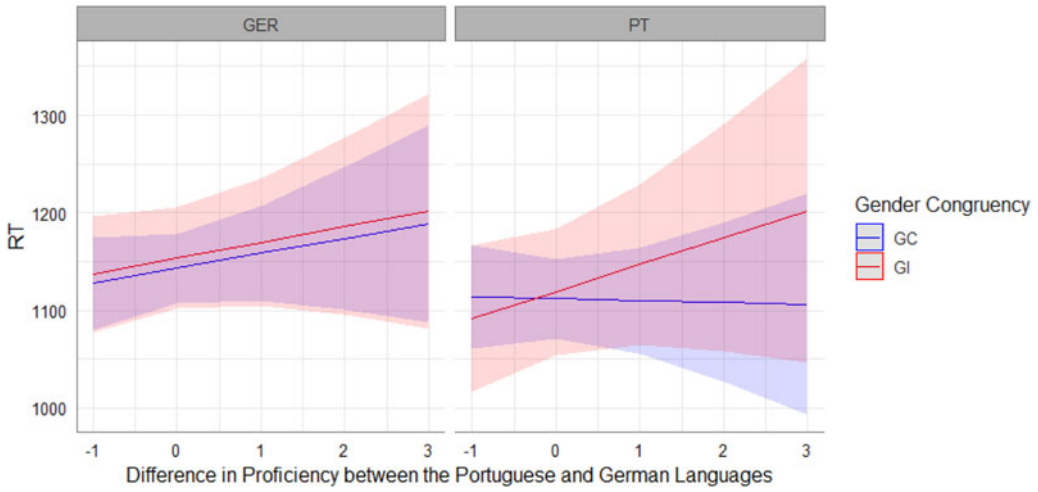

Figure 4. Plot of three-way interaction between Gender Congruency, Language Balance Score, and Target Language
Note. GER = German, PT = Portuguese. GC = Gender Congruent, GI = Gender Incongruent. The higher the difference in proficiency between languages, the higher the imbalance, the higher the effect of gender congruency when producing Portuguese (the higher the interference for heterogeneric nouns and the facilitation for homogeneric nouns). Results in German are not significant.

freedom of the denominator (*p*-values were estimated by the lmerTest package, Kuznetsova et al., 2017).

The results showed a three-way interaction between Gender Congruency, Language Balance Score, and Target Language, *estimate* = 0.02, *SE* = 0.01, *t* = 2.12, *p* = .034, $\chi^2(1) = 4.55$, *p* = .033 (see Table 2 for the results of the linear mixed-effects model and Table 3 for mean RTs and standard errors). This triple interaction indicates that the language balance score influenced the cross-linguistic GCE when producing EP words, *estimate* = 0.02, *SE* = 0.01, *t* = 2.44, *p* = .017, $\chi^2(1) = 5.78$, *p* = .016, but not when producing German words, *estimate* = -0.00, *SE* = 0.01, *t* = 0.56 *p* = .577, $\chi^2(1) = 0.33$, *p* = .566. The results show that, when producing EP words, GCE increased in line with participants' difference in proficiency between languages (see Figure 4), i.e., the higher the imbalance between languages, the higher the effect. A Target Language effect was also observed, *estimate* = -0.02, *SE* = 0.01, *t* = 3.28, *p* = .001, showing that participants were faster at translating words into EP than into German, probably because EP was their L1, the language they learned at home. In contrast, neither an effect of AoA of the German language nor the interaction between that variable and the rest of variables was observed (all *p*s > .05).

Discussion

In the present study, we conducted a forward and a backward translation task with EP and German adult bilinguals. We were interested in testing the cross-linguistic GCE (i.e., facilitation in the processing of homogeneric translations in comparison to heterogeneric translations) in both languages, including as factors within the analyses two usually ignored but relevant variables: AoA and language balance. By following the tenets of the BIA-d model (Grainger et al., 2010) and the Multilink (Dijkstra et al., 2019), as well as previous evidence in other areas of bilingualism (e.g., Abutalebi & Green, 2007; Pivneva et al., 2012; Soares et al., 2019), we proposed that the strength of the links between the L2 lemmas and their gender values varied according to the balance of proficiency between the two languages regardless of the AoA. As a consequence, the dependency of the L2 on the L1 representation would also vary, so that the state of development of the L2 following the BIA-d model also varied depending on this strength. Ultimately, the interference of the L1 on the L2 during gender selection would be more reduced the greater the strength between the L2 lemmas and gender nodes (since there should be less dependency of the L2 on the L1). As the strength of these links is related to frequency of use and exposition, we also hypothesized that a reduction in the use of the L1 would affect the strength of its links and would create a mirroring situation in which the L1 would be more dependent on the L2 and suffer from its interference. Importantly, following previous evidence with translation tasks (Bordag & Pechmann, 2008; Salamoura & Williams, 2007) and recent proposals on the slipperiness of gender effects due to a low degree of activation of gender nodes and their sensitiveness to the time course of lexical access, we also consider the possibility of obtaining null results in the less transparent language, German. Hence, we expected to obtain a cross-linguistic GCE that: (a) was dependent on the language balance, so that the greater the imbalance, the greater the GCE; (b) was observable in both the L1 and the L2 but could be affected by the gender opaqueness of the language; (c) was not dependent on the AoA.

The results were clear-cut: they confirmed the existence of a cross-linguistic GCE with bare nouns in a translation task with early and late bilinguals of EP and German. In line with hypothesis a) our results showed that the GCE increased in parallel with the difference in proficiency between the L1 and L2, so the greater the imbalance, the greater the effect. Note that, indeed, balanced bilinguals did not show the effect. This does not necessarily imply that balanced bilinguals have an autonomous gender system for each language (Costa et al., 2003). Rather, the gender system in bilinguals may be integrated, and following our proposal based on the BIA-d and Multilink models, whether or not interference happens will depend on the strength of the lemma-gender nodes connections and the independence of both languages' representations within the same system.

Furthermore, the triple interaction between Gender Congruency, Language Balance Score, and Target Language showed that, partially in line with hypothesis b) the effect was actually obtained in the

https://doi.org/10.1017/S1366728923000378 Published online by Cambridge University Press

more transparent language (EP) and not in the more opaque language (German) and, in line with hypothesis c), the effect was independent of the AoA, thus it was obtained in early as well as in late bilinguals. Our study hence constitutes supporting evidence to the idea that, indeed, it is the proficiency, and more specifically, the balance between languages that better explains the cross-linguistic influences between languages at the level of processing. Furthermore, our analyses showed that it did not matter whether the imbalance was due to higher proficiency scores in EP or in German: regardless of which language was more dominant, the effect was always visible in EP, never in German. On the one hand, perhaps, once there is imbalance and dependency of one language over another, connections between languages allow for interaction to occur during gender selection even when producing the dominant language. On the other hand, these results are in line with the idea that for gender opaque languages, gender competition during lexical access entails lesser levels of activation that produce smaller and more slippery effects, as shown by Bordag and Pechmann (2008) and Salamoura and Williams (2007).

In conclusion, these results show that gender is selected across languages competitively and that this competition depends mainly on language balance regardless of the AoA or the direction of language dominance. The fact that the effect was restricted to EP corroborates the idea that the resting level of activation of gender nodes and the levels of activation involved in the process of gender selection is higher than that of less transparent languages, like German. We recognize, nevertheless, that more research is necessary to sustain this hypothesis. More specifically, since it seems a quantitative problem on the levels of activation, future studies should examine the cross-linguistic GCE within translation tasks comparing language pairs of different degrees of gender transparency. If we are right, the effect should become stable and greater with transparent pairs, whilst absent with opaque pairs. Yet, with mixed pairs, the effect should be present when the target language has a high degree of transparency and absent (or slippery) when the target language has a low degree of transparency. A more fine-grained analysis is also possible, considering the regularities within the language itself rather than its overall degree of gender transparency. Hence, a factor of transparency congruency mirroring that of gender congruency could be an interesting addition: comparing the effects between transparent translation pairs and opaque pairs, as well as mixed pairs. We also encourage future studies to include the variable of language balance in their design, and to further explore the finding of gender effects in the dominant language, especially if both languages are highly transparent. In that case, the effect would be expected for both the L1 and the L2, and so differences on the size of these effects may be encountered across languages. In this sense, if we were to test the effect of language balance in other L2 languages, it would be interesting to try other type of measurements, such as these based on reaction times in both languages (rather than accuracy), since they can be useful when exploring the differences of naming and translation tasks related to the time course of lexical access (see Casado et al., 2022). Finally, the fact that participants were faster at translating words into EP than into German, independently of their dominant language, yields an interesting result that requires a closer look in future studies. We hypothesized that this may be due to the status of EP as main family language, which is present in the heritage speaker's daily life, despite their higher proficiency in German; however, our explanation is only tentative and calls for more research.

Supplementary material. The supplementary material for this article can be found at https://doi.org/10.1017/S1366728923000378

APPENDIX S1 - Language proficiency, language balance and the DIALANG test

APPENDIX S2 - *All target nouns according to their translation type*

Table S1 - *Means and standard errors of the controlled variables for Portuguese nouns across translation types*

Table S2 - *Means and standard errors of the controlled variables for German nouns across translation types*

Table S3 - *Means and standard errors of the overlap measures*

Acknowledgements. This work was funded by the Foundation for Science and Technology (FCT) through the Portuguese State Budget (UIDP/01662/2020) and the grant UIDB/00305/2020, as well as by the Spanish Ministry of Education and Vocational Training, through the Training program for Academic Staff (Ayudas para la Formación del Profesorado Universitario [FPU16/06983]), and the Spanish Ministry of Science and Innovation [research project PID2019-110583GB-I00]. We thank the Reviewer Alba Casado for her helpful comments especially regarding the data analysis, and all the participants that agreed to be part of our task during the COVID-19 pandemic.

Competing interests. The author(s) declare none.

Data availability. The data that support the findings of this study are openly available at http://doi.org/10.17605/OSF.IO/UE9XH

References

Abutalebi, J., & Green, D. (2007). Bilingual language production: the neurocognition of language representation and control. *Journal of Neurolinguistics*, *20*(30), 242–475. doi: 10.1016/j.jneuroling.2006.10.003

Alderson, C. (2005). *Diagnosing Foreign Language Proficiency: The Interface between Learning and Assessment*. Bloomsbury Publishing PLC.

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

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

Bachman, L. F., & Palmer, A. S. (1989). The construct validation of self-ratings of communicative language ability. *Language Testing*, *6*(1), 14–29. doi: 10.1177/026553228900600104

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, *68*(3), 255–278. doi: 10.1016/j.jml.2012.11.001

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, *67*(1), 1–48. doi: 10.18637/jss.v067.i01

Berken, J. A., Chai, X., Chen, J. K., Gracco, V. L., & Klein, D. (2016). Effects of Early and Late Bilingualism on Resting-State Functional Connectivity. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, *36*(4), 1165–1172. doi: 10.1523/JNEUROSCI.1960-15.2016

Berken, J. A., Gracco, V. L., & Klein, D. (2017). Early bilingualism, language attainment, and brain development. *Neuropsychologia*, *98*, 220–227. doi: 10.1016/j.neuropsychologia.2016.08.031

Birdsong, D. (2015). Dominance in bilingualism: Foundations of measurement, with insights from the study of handedness. In C. Silva-Corvalán & J. Treffers-Daller (Eds.), *Language Dominance in Bilinguals: Issues of Measurement and Operationalization* (pp. 85–105). Cambridge University Press. doi:10.1017/CBO9781107375345.005

Boersma, P., & Weenink, D. (2018). *Praat: doing phonetics by computer* [Computer program]. Retrieved from http://www.praat.org

Bonfieni, M., Branigan, H. P., Pickering, M. J., & Sorace, A. (2019). Language experience modulates bilingual language control: The effect of proficiency, age of acquisition, and exposure on language switching. *Acta Psychologica*, *193*, 160–170. doi: 10.1016/j.actpsy.2018.11.004

Bonvin, A., Brugger, L., & Berthele, R. (2021). Lexical measures as a proxy for bilingual language dominance? *International Review of Applied Linguistics in Language Teaching*. doi: 10.1515/iral-2020-0093

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

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

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

Brysbaert, M., Buchmeier, M., Conrad, M., Jacobs, A. M., Bölte, J., & Böhl, A. (2011). The word frequency effect: A review of recent developments and implications for the choice of frequency estimates in German. *Experimental Psychology*, *58*(5), 412–424. doi: 10.1027/1618-3169/a000123

Bultena, S., Dijkstra, T., & Van Hell, J. (2015). Language switch costs in sentence comprehension depend on language dominance: Evidence from self-paced reading. *Bilingualism: Language and Cognition*, *18*(3), 453–469. doi: 10.1017/S1366728914000145

Burke, L. A., & James, K. E. (2006). Using online surveys for primary research data collection: lessons from the field. *Journal of Innovation and Learning*, *3*(1), 16–30. doi: 10.1504/IJIL.2006.008177

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. (1997). How Many Levels of Processing Are There in Lexical Access? *Cognitive Neuropsychology*, *14*(1), 177–208. doi: 10.1080/026432997381664

Caramazza, A., Miozzo, M., Costa, A., Schiller, N., & Alario, F.-X. (2001). A Cross-linguistic Investigation of Determiner Production. In E. Dupoux (Ed.), *Language, brain, and cognitive development: Essays in honor of Jacques Mehler* (pp. 209–226). MIT Press.

Carroll, S. (1989). Second-language acquisition and the computational paradigm. *Language Learning*, *39*, 535–594. doi: 10.1111/j.1467-1770.1989.tb00902.x

Casado, A., Szewczyk, J., Wolna, A., & Wodniecka, Z. (2022). The relative balance between languages predicts the degree of engagement of global language control. *Cognition*, *226*, 105169. doi: 10.1016/j.cognition.2022.105169

Chuang, Y.-Y., & Baayen, R. H. (2021). Discriminative learning and the lexicon: NDL and LDL. In M. Aronoff (Ed.), *Oxford Research Encyclopedia of Linguistics*. Oxford University Press.

Corbett, G. G. (2013). Number of Genders. In MS Dryer & M Haspelmath (Eds.), *The World Atlas of Language Structures Online*. Max Planck Institute for Evolutionary Anthropology. http://wals.info/chapter/30

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. https://doi.org/10.1017/S1366728999000334

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

Costa, A. S., Comesaña, M., & Soares, A. P. (2021). *PHOR-in-One: A Multilingual Lexical Database With PHonological, ORthographic and Phonographic Word Similarity Estimates in Four Languages* [Manuscript in preparation]. School of Psychology, University of Minho.

de Leeuw, J. R. (2015). jsPsych: a JavaScript library for creating behavioral experiments in a Web browser. *Behavior Research Methods*, *47*(1), 1–12. doi: 10.3758/s13428-014-0458-y

Dijkstra, T., Wahl, A., Buytenhuijs, F., Van Halem, N., Al-Jibouri, Z., De Korte, M., & Rekké, S. (2019). Multilink: A computational model for bilingual word recognition and word translation. *Bilingualism: Language and Cognition*, *22*(4), 657–679. doi: 10.1017/S1366728918000287

Dussias, P. E., & Sagarra, N. (2007). The effect of exposure on syntactic parsing in Spanish-English bilinguals. *Bilingualism: Language and Cognition*, *10*(1), 101–116. doi: 10.1017/S1366728906002847

Egger, E., Hulk, A., & Tsimpli, I. M. (2018). Crosslinguistic influence in the discovery of gender: the case of Greek–Dutch bilingual children. *Bilingualism: Language and Cognition*, *21*(4), 694–709. doi: 10.1017/S1366728917000207

Flores, C. (2015). Understanding heritage language acquisition. Some contributions from the research on heritage speakers of European Portuguese. *Lingua*, *164*, 251–265. doi: 10.1016/j.lingua.2014.09.008

Flores, C. (2020). Attrition and reactivation of a childhood language. The case of returnee heritage speakers. *Language Learning*, *70*, 85–121. doi: 10.1111/lang.12350

Flores, C., Zhou, C., & Eira, C. (2022). "I no longer count in German". On dominance shift in returnee heritage speakers. *Applied Psycholinguistics*, *43*(5), 1019–1043. doi: 10.1017/S0142716422000261

Franceschina, F. (2005). *Fossilised second language grammars: the acquisition of grammatical gender*. John Benjamins. doi: 10.1075/lald.38

Fuchs, Z. (2022). Heritage speakers' ability to use gender in online processing: evidence from eye-tracking. Paper presented at the Heritage Languages Around the World Conference, University of Lisbon, Lisbon, Portugal (18–20 May).

Gagarina, N., & Klassert, A. (2018). Input Dominance and Development of Home Language in Russian-German. Bilinguals. *Frontiers in Communication*, *3*, 40. doi: 10.3389/fcomm.2018.00040

Goethe-Institut. (2010). German placement test. Retrieved from http://www.goethe.de/cgi-bin/einstufungstest/einstufungstest.pl

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

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

Guasch, M., Boada, R., Ferré, P., & Sánchez-Casas, R. (2013). NIM: A Web-based Swiss Army knife to select stimuli for psycholinguistic studies. *Behavior Research Methods*, *45*(3), 765–771. doi: 10.3758/s13428-012-0296-8

Guo, T., Liu, H., Misra, M., & Kroll, J. F. (2011). Local and global inhibition in bilingual word production: fMRI evidence from Chinese-English bilinguals. *NeuroImage*, *56*(4), 2300–2309. doi: 10.1016/j.neuroimage.2011.03.049

Hamann, C., Chilla, S., Gagarina, N., & Abed-Ibrahim, L. (2017). Syntactic complexity and bilingualism: how (a)typical bilinguals deal with complex structures. In E. Di Domenico (Ed.), *Complexity in Acquisition* (pp. 142–178). Cambridge Scholars Publishing.

Hatzidaki, A., Branigan, H. P., & Pickering, M. J. (2011). Co-activation of syntax in bilingual language production. *Cognitive psychology*, *62*(2), 123–150. doi: 10.1016/j.cogpsych.2010.10.002

Hawkins, R. (2009). Statistical Learning and Innate Knowledge in the Development of Second Language Proficiency: Evidence From the Acquisition of Gender Concord. In A. G. Benati (Ed.), *Issues in Second Language Proficiency* (pp. 63–78). Bloomsbury Academic. http://doi.org/10.5040/9781474212236.ch-005

Heister, J., Würzner, K.-M., Bubenzer, J., Pohl, E., Hanneforth, T., Geyken, A., & Kliegl, R. (2011). dlexDB–Eine lexikalische Datenbank für die psychologische und linguistische Forschung [dlexDB–A lexical database for the psychological and linguistic research]. *Psychologische Rundschau*, *62*(1), 10–20. doi: 10.1026/0033-3042/a000029

Hermans, D., Bongaerts, T., De Bot, K., & Schreuder, R. (1998). Producing words in a foreign language: Can speakers prevent interference from their first language? *Bilingualism: Language and Cognition*, *1*(3), 213–229. doi: 10.1017/S1366728998000364

Hyltenstam, K., & Abrahamsson, N. (2003). Maturational constraints in SLA. In C. Doughty & M. H. Long (Eds.), *The Handbook of Second Language Acquisition* (pp. 539–588). Blackwell.

Klassen, R. (2016). The representation of asymmetric grammatical gender systems in the bilingual mental lexicon. *Probus*, *28*(1). doi: 10.1515/probus-2016-0002

Klassen, R., Kolb, N., Hopp, H., & Westergaard, M. (2022). Interactions between lexical and syntactic L1-L2 overlap: Effects of gender congruency on L2 sentence processing in L1 Spanish-L2 German speakers. *Applied Psycholinguistics*, 1–36. doi: 10.1017/S0142716422000236

Klaus, J., Lemhöfer, K., & Schriefers, H. (2018). The second language interferes with picture naming in the first language: Evidence for L2 activation during L1 production. *Language, Cognition and Neuroscience*, *33*(7), 867–877. doi: 10.1080/23273798.2018.1430837

Köpcke, K.-M. (1982). *Untersuchungen zum Genussystem der deutschen Gegenwartssprache* [Investigations on the gender system of the contemporary German language].

Köpcke, K.-M., & Zubin, D. (1983). Die kognitive Organisation der Genuszuweisung zu den einsilbigen Nomen der deutschen Gegenwartssprache [The cognitive organization of gender allocation to monosyllabic nouns in contemporary German]. *Zeitschrift für germanistische Linguistik*, *11*, 166–182.

Kroll, J. F., Michael, E., Tokowicz, N., & Dufour, R. (2002). The development of lexical fluency in a second language. *Second Language Research*, *18*(2), 137–171. doi: 10.1191/0267658302sr201oa

Kupisch, T., Akpınar, D., & Stöhr, A. (2013). Gender assignment and gender agreement in adult bilingual and second language speakers of French. *Linguistic Approaches to Bilingualism*, *3*(2), 150–179. doi: 10.1075/lab.3.2.02kup

Kupisch, T., Geiß, M., Mitrofanova, N., & Westergaard, M. (2018). *Gender Cues in L1 Russian Children Acquiring German as an Early L2.* EuroSLA 28 - Universität Münster.

Kupisch, T., Mitrofanova, N., & Westergaard, M. (2022). Phonological vs. natural gender cues in the acquisition of German by simultaneous and sequential bilinguals (German–Russian). *Journal of Child Language*, *49*(4), 661–683. doi: 10.1017/S0305000921000039

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest Package: Tests in Linear Mixed Effects Models. *Journal of Statistical Software*, *82*(13), 1–26. doi: 10.18637/jss.v082.i13

Laufer, B., & Nation, P. (1999). A vocabulary-size test of controlled productive ability. *Language Testing*, *16*(1), 33–51. doi: 10.1191/026553299672614616

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

Levy, B. J., McVeigh, N. D., Marful, A., & Anderson, M. C. (2007). Inhibiting your native language: the role of retrieval-induced forgetting during second-language acquisition. *Psychological science*, *18*(1), 29–34. doi: 10.1111/j.1467–9280.2007.01844.x

Li, P., Zhang, F., Yu, A., & Zhao, X. (2020). Language History Questionnaire (LHQ3): An enhanced tool for assessing multilingual experience. *Bilingualism: Language and Cognition*, *23*(5), 938–944. doi: 10.1017/S1366728918001153

Lim, J. H., & Christianson, K. (2015). Second language sensitivity to agreement errors: Evidence from eye movements during comprehension and translation. *Applied Psycholinguistics*, *36*(6), 1283–1315. doi: 10.1017/S0142716414000290

Linck, J. A., Kroll, J. F., & Sunderman, G. (2009). Losing access to the native language while immersed in a second language: Evidence for the role of inhibition in second-language learning. *Psychological Science*, *20*(12), 1507–1515. doi: 10.1111/j.1467-9280.2009.02480.x

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

Meisel, J. M. (2008). Child second language acquisition or successive first language acquisition? In B. Haznedar & E. Gavruseva (Eds.), *Current Trends in Child Second Language Acquisition: A Generative Perspective* (pp. 55–82). John Benjamins. doi: 10.1075/lald.46.04mei

Montrul, S. (2016). *The Acquisition of Heritage Languages*. Cambridge University Press, doi: 10.1017/CBO9781139030502

Morales, L., Paolieri, D., & Bajo, M. T. (2011). Grammatical gender inhibition in bilinguals. *Frontiers in Psychology*, *2*. doi: 10.3389/fpsyg.2011.00284

Morales, L., Paolieri, D., Cubelli, R., & Bajo, M. T. (2014). Transfer of Spanish grammatical gender to English: Evidence from immersed and non-immersed bilinguals. *Bilingualism: Language and Cognition*, *17*(4), 700–708. doi: 10.1017/S1366728914000017

Paolieri, D., Cubelli, R., Macizo, P., Bajo, M. T., Lotto, L., & Job, R. (2010). Grammatical gender processing in Italian and Spanish bilinguals, *The Quarterly Journal of Experimental Psychology*, *63*, 1631–1645. doi: 10.1080/17470210903511210

Paolieri, D., Padilla, F., Koreneva, O., Morales, L., & Macizo, P. (2019). Gender congruency effects in Russian–Spanish and Italian–Spanish bilinguals: The role of language proximity and concreteness of words. *Bilingualism: Language and Cognition*, *22*, 112–129. doi: 10.1017/s1366728917000591

Paolieri, D., Demestre, J., Guasch, M., Bajo, M. T., & Ferré, P. (2020). The gender congruency effect in Catalan–Spanish bilinguals: Behavioral and electrophysiological evidence. *Bilingualism: Language and Cognition*, *23*(5), 1045–1055. doi: 10.1017/S1366728920000073

Pivneva, I., Palmer, C., & Titone, D. (2012). Inhibitory control and l2 proficiency modulate bilingual language production: evidence from spontaneous monologue and dialogue speech. *Frontiers in psychology*, *3*, 3–57. doi: 10.3389/fpsyg.2012.00057

Prior, A., MacWhinney, B., & Kroll, J. F. (2007). Translation norms for English and Spanish: The role of lexical variables, word class, and L2 proficiency in negotiating translation ambiguity. *Behavior Research Methods*, *39*, 1029–1038. doi: 10.3758/BF03193001

Ramscar, M., Hendrix, P., Shaoul, C., Milin, P., & Baayen, H. (2014). The myth of cognitive decline: non-linear dynamics of lifelong learning. *Topics in cognitive science*, *6*(1), 5–42. doi: 10.1111/tops.12078

Roelofs, A. (1992). A spreading-activation theory of lemma retrieval in speaking. *Cognition*, *42*(1–3), 107–142. doi: 10.1016/0010-0277(92)90041-F

Sá-Leite, A. R. (2021). Representation and Processing of Grammatical Gender: Analysing the gender congruency effect [Doctoral dissertation, University of Santiago de Compostela].

Sá-Leite, A. R., Fraga, I., & Comesaña, M. (2019). Grammatical gender processing in bilinguals: An analytic review. *Psychonomic Bulletin & Review*, *26*(4), 1148–1173. doi: 10.3758/s13423-019-01596-8

Sá-Leite, A. R., Luna, K., Fraga, I., & Comesaña, M. (2020). The gender congruency effect across languages in bilinguals: A meta-analysis. *Psychonomic Bulletin & Review*, *27*(4), 677-693. doi: 10.3758/s13423-019-01702-w

Sá-Leite, A. R., Luna, K., Tomaz, Â., Fraga, I., & Comesaña, M. (2022). The mechanisms underlying grammatical gender selection in language production: A meta-analysis of the gender congruency effect. *Cognition*, *224*, 105060. doi: 10.1016/j.cognition.2022.105060

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

Schad, D. J., Vasishth, S., Hohenstein, S., & Kliegl, R. (2020). How to capitalize on a priori contrasts in linear (mixed) models: A tutorial. *Journal of Memory and Language*, *110*, 104038. doi: 10.1016/j.jml.2019.104038

Schmid, M. S. (2009). On L1 attrition and the linguistic system. In L. Roberts, G. D. Verónique, A. Nilsson, and M. Tellier (Eds.), *EUROSLA Yearbook: Volume 9* (pp. 212–244). John Benjamins. doi: 10.1075/eurosla.9.11sch

Soares, A. P., Machado, J., Costa, A., Iriarte, Á., Simões, A., de Almeida, J. J., Comesaña, M., & Perea, M. (2015). On the Advantages of Word Frequency and Contextual Diversity Measures Extracted from Subtitles: The Case of Portuguese. *Quarterly Journal of Experimental Psychology*, *68*(4), 680–696. doi: 10.1080/17470218.2014.964271

Soares, A. P., Costa, A. S., Machado, J., Comesaña, M., & Oliveira, H. (2017). The Minho Word Pool: Norms for imageability, concreteness and subjective frequency for 3,800 Portuguese words. *Behavior Research Methods*, *49*(3), 1065–1081. doi: 10.3758/s13428-016-0767-4

Soares, A. P., Iriarte, A., Almeida, J. J., Simões, A., Costa, A., Machado, J., França, P., Comesaña, M., Rauber, A., Rato, A., & Perea, M. (2018). Procura-PALavras (P-PAL): A web-based interface for a new European Portuguese lexical database. *Behavior Research Methods*, *50*(4), 1461–1481. doi: 10.3758/s13428-018-1058-z

Soares, A. P., Oliveira, H., Ferreira, M., Comesaña, M., Macedo, F., Ferré, P., Acuña-Fariña, J. C., Hernández-Cabrera, J., & Fraga, I. (2019). Lexico-syntactic interactions during the processing of temporally ambiguous L2 relative clauses: An eye-tracking study with intermediate and advanced Portuguese-English bilinguals. *Plos One*, *14*(5): e0216779. DOI: 10.1371/journal.pone.0216779

Treffers-Daller, J., & Korybski, T. (2016). Using lexical diversity measures to operationalise language dominance in bilinguals. In C. Silva-Corvalán & J. Treffers Daller (Eds.), *Language Dominance in Bilinguals: Issues of Measurement and Operationalization* (pp. 106–123). Cambridge University Press.

Ulbrich, C., & Ordin, M. (2014). Can L2-English influence L1-German? The case of post-vocalic /r/. *Journal of Phonetics*, *45*, 26–42. doi: 10.1016/j.wocn.2014.02.008

Unsworth, S. (2008). Age and input in the acquisition of grammatical gender in Dutch. *Second Language Research*, *24*(3), 365–395. http://www.jstor.org/stable/43103774

Van Hell, J. G., & Tanner, D. (2012). Second Language Proficiency and Cross-Language Lexical Activation. *Language Learning*, *62*, 148–171. doi: 10.1111/j.1467-9922.2012.00710.x

Wattendorf, E., Festman, J., Westermann, B., Keil, U., Zappatore, D., Franceschini, R., Luedi, G., Radue, E.-W., Münte, T. F., Rager, G., & Nitsch, C. (2014). Early bilingualism influences early and subsequently later acquired languages in cortical regions representing control functions. *International Journal of Bilingualism*, *18*(1), 48–66. doi: 10.1177/1367006912456590

Zuur, A. F., Ieno, E. N., & Elphick, C. S. (2010). A protocol for data exploration to avoid common statistical problems. *Methods in Ecology and Evolution*, *1*(1), 3–14. doi: 10.1111/j.2041-210X.2009.00001.x

https://doi.org/10.1017/S1366728923000378 Published online by Cambridge University Press