The role of lexico-syntactic features in noun phrase production and comprehension: insights from Spanish and Chinese in unilingual and bilingual contexts

Wu, R.

The Role of Lexico-Syntactic Features in Noun Phrase Production and Comprehension

Insights from Spanish and Chinese in Unilingual and Bilingual Contexts

Published by

LOT
Binnengasthuisstraat 9
1012 ZA Amsterdam
The Netherlands

phone: +31 20 525 2461

e-mail: lot@uva.nl
http://www.lotschool.nl

Cover illustration: *Link* — concept, design, and drawing by Ruijie Wu.

ISBN: 978-94-6093-490-2
DOI: https://dx.medra.org/10.48273/LOT0705
NUR: 616

The Role of Lexico-Syntactic Features in Noun Phrase Production and Comprehension

Insights from Spanish and Chinese in Unilingual and Bilingual Contexts

Proefschrift

ter verkrijging van
de graad van doctor aan de Universiteit Leiden,
op gezag van rector magnificus prof.dr.ir. H. Bijl,
volgens besluit van het college voor promoties
te verdedigen op dinsdag 16 december 2025
klokke 14.30 uur

door

Ruixue Wu

geboren 8 november 1993
te Shangqiu, China

Promotor: Prof. dr. N. O. Schiller
Co-promotor: Dr. M. C. Parafita Couto (University of Santiago de Compostela)

Promotiecommissie: Prof. dr. F. K. Ameka
Prof. dr. N. Sebastián Gallés (Pompeu Fabra University)
Prof. dr. A. Muntendam (Florida State University)
Dr. K. R. Bellamy

This thesis is dedicated to my parents.

衷心感谢我的父母，是他们竭尽全力地托举，才让我有机会看见更广阔的世界，拥有追逐梦想的勇气与力量.

Table of Contents

Acknowledgements

Pursuing a PhD is a long and challenging journey, one that requires perseverance, dedication, and hard work, but most importantly, the unwavering support of supervisors, family, friends, and colleagues.

First and foremost, I want to express my deepest gratitude to my parents. Their unconditional love and support have been the foundation of my academic journey. Growing up, I was never the most well-behaved child, and I even went through a rebellious phase in high school. But my parents never gave up on me. With patience, encouragement, and sometimes tough love, they guided me through each stage, from high school to university, to graduate school, and now my PhD. I am deeply grateful to my loving and supportive parents for always trusting my decisions and giving me everything they could to help me succeed.

I would like to sincerely thank my supervisors, Prof. Dr. Niels O. Schiller and Dr. Maria del Carmen Parafita Couto. This dissertation would not have been possible without their guidance and support. Their expertise, patience, and thoughtful feedback have been invaluable in shaping my academic path and growth as a researcher. I am especially grateful to Maria for her warmth and kind words, "I truly enjoy working with you and want to see you succeed," which gives me confidence and strength. I also deeply appreciate Niels for his constant support and uplifting words like "Really great job, Rachel," which motivated me more times than I can count. Their belief in me helped me move forward at every step of this journey.

Special thanks to some very special people in my life. **Yingjia Liu,** my best friend for over eight years, thank you for always being there, supporting me through tough times, and sharing in both my joy and frustration. **Fang Hai,** my dearest bestie, thank you for every thoughtful gift, your warm hospitality whenever I returned to China. My sister, **Ruijie Wu,** thank you for taking care of our parents while I pursued my PhD abroad. Your love and support made it possible for me to chase my dreams without worry. **Shijie Fu**, thank you for being my emotional outlet, carrying all my tears, and staying by my side through dark moments. **Tingting Zheng**, thank you for always listening and offering thoughtful advice whenever I struggled with research or life. **Jiaqi Wang**, thank you for all late-night chats and constant encouragement. **Yaqian Wang,** thank you for being there whenever I felt lost. **Zheyu Shang**, thank you for being warm, kind, and supportive all the time. **Jingtian Shi**, thank you for all the laughs, drinks, and ice creams that brightened my office life. I love you all!

To my dear friends and colleagues at LUCL and Humanities: Sarah Von Grebmer Zu Wolfsthurn and Shaoyu Wang, thank you for your help with lab training, EEG pre-processing, and statistical advice. Special thanks to Yanan Wang, Yufang Wang, Jin Wang, Qi Zhang, and Ziheng Cheng for all group meetings! To the dear PhD council: Bente De Grarve, Hester Groot, Priscilla Lam, Wenwei (Charlie) Xu, Olga Nozdracheva, and Turan Hanci, thank you for organizing fun activities that brought joy and color to my PhD life. Special thanks to Jiang Wu, Zhen Li, Xinyi Wen, Ruoyu Shi, Yiran Ding, Xin Li, Jian Sun, Jiahui Liang, Yingyang Wang, Sarah Stolle, Hadis Tamleh, Saskia Dunn, Sophia Nauta, Matthew Sung, Pascale Eskes, and Rahmatan Idul for being part of my journey.

A warm thank you to everyone at the Center for Brain and Cognition (CBC), Universitat Pompeu Fabra (UPF). I am especially grateful to Prof. Dr. Núria Sebastián Gallés for her warm welcome and insightful suggestions, which made my research possible. Many thanks to Francesc Xavier Mayoral Terron and Sílvia Blanch Rabat for their generous help with the experimental setup and data collection. To my dear friends at CBC, Judit Ciarrusta Monzon, Ibtissam Ghailan, Chiara Santolin, and Ege Ekin Özer, thank you for your company, shared lunchtimes, and helpful advice. I also deeply appreciate the time and effort of all participants who made this PhD project possible.

To my Jiu-Jitsu coach, Hugo Nijhof, and Brazilian Jiu-Jitsu coach, Iwan Hansen, thank you for your encouragement, professionalism, and for creating such a welcoming and supportive environment where I could recharge and find balance amidst my academic work. My heartfelt thanks also go to Hilda Vos, Aurélie Berthot, Pjotr Straver, Dennis de Booij, Baan Schroef, Caroline Schep, Xueting (Mia) Zhang, and all my BJJ and JJ teammates for their passion, friendliness, and the energy they brought to every training session and sparring round.

A special thank you goes to my idol, Junjie Lin (JJ Lin). Your beautiful music has been my emotional companion during difficult times, giving me comfort and strength when I needed it the most.

Finally, I want to thank myself for enduring countless sleepless nights, emotional breakdowns, and self-doubt. Thank you for never giving up, even when things got really hard. I stayed strong, kept going, and made it through. I am proud to say: I did it!

Abbreviations

CS	code-switching
FEM	feminine
MASC	masculine
NEU	neuter
COM	common
SG	singular
PL	plural
ADJ	adjective
N	noun
ART	article
INDEF	indefinite
DEF	definite
CL	classifier
NP	noun phrase
PP	prepositional phrase
CN	Chinese
SP	Spanish

Chapter 1

General Introduction

1.1 Background

Over half of the world's population speaks more than one language, making bilingualism[1] a widespread phenomenon rather than an exception (Grosjean, 1982). This prevalence naturally raises one important question in the field of bilingualism (e.g., Desmet & Duyck, 2007), i.e., how do bilingual speakers manage two distinct languages in their minds? To answer this question, accumulating research has first established that bilinguals undoubtedly differ from monolinguals in certain ways, such as vocabulary size and verbal skills (e.g., Bialystok et al., 2010; Bialystok & Luk, 2012), as well as cognitive patterns and lexical retrieval (e.g., response time and accuracy in picture-naming tasks, word production, and comprehension, see Bialystok et al., 2008; Gollan et al., 2005; Costa & Santesteban, 2004; Hernandez et al., 2000; Gollan et al., 2007; Roberts et al., 2002). From a psycholinguistic perspective, most theories of bilingual lexical access propose that both lexicons in a bilingual's mind are activated in parallel (e.g., Costa et al., 1999a; De Bot, 1992; De Bot & Schreuder, 1993; Dijkstra & Van Heuven, 1998; Green, 1986; Roelofs, 1998). According to these theories, bilinguals undergo a more complex and multi-staged process than monolinguals when producing a word: first, to-be-produced lexical items are activated simultaneously in both used and non-used languages; then, an external production rule determines the target language; third, the highest activated target lexical item in the chosen language is ultimately produced while suppressing those of the non-used languages.

A widely held view is that bilingualism often arises not from deliberate personal choice, but from circumstances such as birthplace, migration background, or the language patterns prevalent within an individual's community (Bialystok et al., 2012). This is especially evident among early Spanish−Chinese bilinguals living in a young but stable immigrant comm-unity in Barcelona, Spain, where consistent heritage language use and rich

[1] In this thesis, *bilingualism* is used as a cover term to refer broadly to individuals who speak two or more languages.

language contact often create a context of frequent use of two languages. This thesis thus focuses on this bilingual population, not only because their language practices offer rich opportunities to study bilingual language processing, but also because they remain notably understudied in existing research. Specifically, we defined ***early Spanish–Chinese bilinguals*** as individuals who acquired Chinese as a heritage language (HL), that is, as their home language learned in a Chinese-speaking household environment, and Spanish through formal educational settings in Spain. Moreover, we included two baseline comparison groups, Mandarin Chinese speakers and Spanish speakers, to establish language-specific patterns and provide essential reference points for identifying strategies or deviations unique to bilinguals. Rather than labeling these participants as monolinguals, we defined them more precisely: ***Mandarin Chinese speakers*** are individuals born and raised in monolingual Chinese-speaking environments in China, and ***Spanish speakers*** are those born and raised in Spanish-speaking countries. This approach acknowledges that even individuals raised in predominantly single-language environments may possess some knowledge of other languages. Notably, these two groups were included only in the unilingual[2] investigations in this thesis.

The central languages of interest in this thesis are Spanish and Mandarin Chinese (hereafter, "Chinese" refers to Mandarin Chinese), as these two languages differ markedly in their typological and grammatical systems, providing a compelling case for exploring bilingual language processing. Specifically, they contrast not only in typology, with Spanish being a fusional language with rich inflectional morphology, while Chinese is isolating with minimal morphology, but also in key lexico-syntactic

---

[2] In this thesis, *unilingual* refers to the use of only one language by bilingual speakers, typically in a task or interaction conducted in a single language. It describes the language mode, not the speaker type. In contrast, *monolingual* refers to individuals who primarily use one language in their daily lives or to tasks involving exclusively monolingual participants. This distinction ensures that the term *unilingual* is reserved for describing the single-language use of bilinguals, while monolingual pertains to both speaker identity and task composition.

features. In this thesis, ***lexico-syntactic features*** are broadly defined as properties of a word that reflect both its lexical characteristics (such as meaning or category, e.g., noun, verb) and its syntactic role or behavior within a sentence (such as agreement). These features help determine how words combine to form phrases and sentences, and how they interact with grammatical rules. Examples of lexico-syntactic features include grammatical gender, classifier systems, and adjective placements (see detailed descriptions in Section 1.2). In Spanish, grammatical gender is an inherent feature of nouns, requiring determiners and adjectives to agree with nouns in gender (Roca, 1989; Harris, 1991). For example, *la*$_{FEM}$ *mesa*$_{FEM}$ "the table", where the feminine determiner *la* agrees in gender with the feminine noun *mesa*. In contrast, Chinese lacks grammatical gender but instead employs a classifier system, where classifiers are obligatorily placed after numerals or demonstratives but before nouns (Li & Thompson, 1981). The selection of classifiers largely depends on the semantic properties of nouns they modify, such as animacy, shape, size, function, and consistency (Li & Thompson, 1981; Myers & Tsay, 2000). For instance, *一棵树*, /yi1ke1shu4/ [one + classifier-ke1 + tree], "one tree", both the classifier *棵* /ke1/ and the noun *树* /shu4/ "tree" refer to plants (Guo, 2002). Another point of contrast lies in the adjective placements: Spanish typically places adjectives after nouns (e.g., *guante amarillo*, [glove yellow], "yellow glove"), whereas Chinese adjectives generally precede nouns (e.g., *黄色的手套*, [yellow glove], "yellow glove"). These structural differences provide a rich context for examining how bilingual speakers navigate the lexico-syntactic features across their languages during production and comprehension.

Taken together, these lexico-syntactic contrasts between Spanish and Chinese create "conflict sites" for bilingual speakers (Poplack & Meechan, 1998), where the grammatical rules of the two languages differ. To understand how bilinguals mentally navigate such lexico-syntactic differences across languages, it is essential to first examine how these features function in each language individually. Unlike the adjective placements, which exist in both languages but are realized differently, grammatical gender and classifiers are language-specific lexico-syntactic features that exist in Spanish and Chinese, respectively, but serve comparable grammatical functions. Among these, grammatical gender stands out as one of the most extensively studied lexico-syntactic features, particularly in monolingual language production, as evidenced by the well-documented ***gender congruency effect*** (e.g., Bürki et al., 2016; Heim et al., 2009; La Heij et al., 1998; Schiller, 2013; Schiller & Caramazza, 2003; Schiller & Costa, 2006; Schriefers & Teruel, 2000; Starreveld & La Heij, 2004). The gender congruency effect refers to a more efficient and accurate processing of gender-congruent nouns (i.e., nouns and determiners or adjectives agree in gender) compared to gender-incongruent nouns (i.e., nouns and determiners or adjectives disagree in gender, see Wang & Schiller, 2019; Sá Leite et al., 2022 and Bürki et al., 2023 for overviews).

According to the influential Levelt, Roelofs, and Meyer (LRM) speech production model (Levelt et al., 1999), there are three main stages to produce a word. First, the to-be-produced word is conceptualised. Then, the concept is lexicalised, which activates and encodes the grammatical properties (e.g, grammatical gender, classifiers, number, etc.) and phonological and phonetic form of the word. Finally, the to-be-produced word is articulated. This production process is mostly examined in a picture-naming task using the picture-word interference (PWI) paradigm (Rosinski et al., 1975), where speakers are asked to name a target picture while ignoring a simultaneously presented distractor word. In this context, the gender nodes for both the target and the distractor are activated. When the target and distractor mismatch in gender (i.e., gender-incongruent condition), the gender feature of the distractor word competes with that of the target picture, resulting in a longer time to name the picture. Conversely, when the target and the distractor have an agreement in gender (gender-congruent condition), they activate the same gender node, where no competition occurs and results in shorter naming latencies. This difference in processing, reflected by faster naming latencies in the gender-congruent condition, is referred to as the gender congruency effect. This effect has been extensively documented in Germanic languages (see Wang & Schiller, 2019; Sá Leite et al., 2022 and Bürki et al., 2023 for an overview), while controversial results have been observed in Romance languages, where this effect has been reported to either exist, be absent, or even reverse, especially in Spanish NP production (see Section 1.2.1 for detailed descriptions). This motivates the present thesis to first investigate whether Spanish speakers show a gender congruency effect in Spanish NP production. Establishing this baseline is essential for understanding how grammatical gender is processed in unilingual contexts, which in turn provides a foundation for examining how bilingual speakers process grammatical gender in bilingual speech contexts.

Another comparable lexico-syntactic feature is the numeral classifier system in Chinese, which has received much less investigation, especially regarding its processing in bilingual speech production. Like grammatical gender, classifiers are processed through the same stages of word production proposed in the LRM model (Levelt et al., 1999), from conceptualisation to lexical encoding of grammatical properties and finally to articulation (see Wang & Schiller, in press and Qian, in press for overviews). In monolingual speech production, a phenomenon known as the ***classifier congruency effect*** has been consistently documented using the PWI paradigm (e.g., Wang et al., 2006; Huang & Schiller, 2021). This effect shows that speakers name pictures faster when the distractor and target picture match regarding their classifiers (classifier-congruent condition) than when there is a mismatch (classifier-incongruent condition). Specifically, studies using the PWI-based picture-naming task have demonstrated that Chinese speakers produce classifier-NPs more quickly under classifier-congruent conditions than under classifier-incongruent conditions, as the classifier of the distractor competes with that of the target under classifier-incongruent conditions, leading to longer naming latencies (e.g., Wang et al., 2006; Huang & Schiller, 2021). While this effect has been well-documented among monolingual Chinese speakers, it is still unclear whether bilingual speakers exhibit a similar classifier congruency effect when processing classifiers in unilingual Chinese. Addressing this gap, the present thesis also aims to investigate how Spanish−Chinese bilinguals process classifiers in such unilingual contexts.

Before we delve into how bilinguals manage grammatical gender, classifiers, and adjective placements across their two languages, it is first essential to clarify what we mean by ***bilingual*** in this thesis. Following Grosjean (1992, p. 51), bilinguals are defined as "…those people who need and use two (or more) languages in their everyday lives." In previous research, bilingual speakers are often suggested to activate both their languages in parallel during speech production and comprehension (e.g., De Bot, 1992; De Bot & Schreuder, 1993; Dijkstra & Van Heuven, 1998; Green, 1986), making it unlikely for one language to be fully "switched off", even when in single-language contexts (Desmet & Duyck, 2007). As a result of such parallel activation, bilingual speakers do not simply manage two separate linguistic systems in isolation; rather, they often alternate between them and integrate elements from both languages within a single conversation, a phenomenon known as code-switching (CS) (Poplack, 1980; Deuchar, 2012). Interestingly, it has also been observed that bilinguals actively make choices about when, where, and with whom to use one language or to mix the two in code-switching (Parafita Couto et al., 2023). Thus, studying code-switched speech often offers a unique window to our understanding of language, as it can reveal patterns that might remain hidden when examining a single language (Parafita Couto et al., 2023). In this way, combining the investigation of code-switched speech with an analysis of how bilinguals process and manage lexico-syntactic features across their two languages provides a more comprehensive perspective than studies limited to monolingual speakers or contexts. This thesis therefore shifts the focus away from monolingual contexts, aiming to use bilingualism instead as a gateway into the mechanisms underlying these lexico-syntactic features, both in unilingual and bilingual language modes. In doing so, it allows us to observe not only how these features are processed in a single-language mode but also how they interact across two languages, thereby offering insights into a relatively comprehensive understanding of bilingual language processing.

Building on this aim, this thesis then extends the investigation by examining how grammatical gender, classifiers, and adjective placement are managed and processed in code-switched speech. It is widely accepted that code-switching does not involve a random mixture of two languages but follows specific rules (e.g., Poplack, 1980; Deuchar, 2012; Bullock & Toribio, 2009). Accordingly, previous studies have first explored how grammatical gender is processed and assigned in code-switched speech, particularly in mixed NPs involving gender-marked determiners with non-gender-marked nouns (cf. Bellamy & Parafita Couto, 2022). Specific to language combinations involving Spanish (e.g., Spanish−English), previous studies have identified three main strategies for gender assignment (cf. Bellamy & Parafita Couto, 2022; also see Section 1.2.1 for detailed descriptions). This includes assigning gender based on either translation equivalents or morpho-phonological properties of nouns, or by defaulting to masculine, which serves as the default gender in Spanish (see Bellamy & Parafita Couto, 2022 for an overview). By contrast, to date, no studies have systematically examined how classifiers are assigned in code-switched speech. Given the similar noun-based nature of both grammatical gender and classifier systems, it is plausible that comparable assignment strategies apply to classifiers. Thus, we propose that classifiers may be assigned using translation equivalents, morpho-phonological cues, or defaulting to the general classifier *ge*, which is used as default in Chinese (Myers & Tsay, 2000, see detailed descriptions in Section 1.2.2). In doing so, this thesis systematically investigates how grammatical gender and classifiers are assigned and processed in mixed NPs, thereby shedding light on the strategies Spanish−Chinese bilinguals employ when navigating these two lexico-syntactic features across their two languages.

Additionally, adjective placement, as a unique "conflict site", emerges in bilingual contexts where the conflict depends on whether adjectives come before or after nouns in the language involved. This issue has been extensively explored in several studies using different methodologies across language combinations (e.g., Balam & Parafita Couto, 2019; Stadthagen-González et al., 2019; Parafita Couto & Gullberg, 2019; Pablos et al., 2019; Vanden Wyngaerd, 2017; Parafita Couto et al., 2015; Parafita Couto et al., 2017; Vaughan-Evans et al., 2020; Van Osch et al., 2023; also see overview in Appendix 6.A in Chapter 6). These studies examined such adjective placement conflicts based on predictions of two theoretical frameworks, including the Matrix Language Frame (MLF; Myers-Scotton, 1993, 2002) and the Minimalist Program (MP)-based approaches (Cantone & MacSwan, 2009). These models capture different aspects of syntactic constraints and differ in predictions: the MLF (Myers-Scotton, 1993, 2002) proposes that code-switched sentences involve a matrix language (ML) that provides the morphosyntactic frame and an embedded language (EL) that supplies lexical items, and thus predicts that the word order of ML determines adjective positions. In contrast, MP-based approaches (Cantone & MacSwan, 2009) argue that code-switching should be governed by the same grammatical principles as monolingual syntax, and thus grammatical features are determined by the properties of individual lexical items. Thus, regarding adjective placement in code-switching, Cantone and MacSwan (2009) follow Cinque's (1994, 1999) proposal that adjectives are universally placed before nouns (see Section 1.2.3 for detailed descriptions). Based on this, they reach the descriptive generalization that adjective placement should be determined by the language of the adjective itself. Notably, although these two models capture different aspects of adjective-noun patterns and differ in assumptions, their predictions sometimes converge or diverge, i.e., adjective positions may align with both models, with only one, or with neither.

To date, previous studies have examined predictions of these two models and yielded inconclusive results regarding fully explaining grammatical patterns in adjective placement in code-switched speech (see Section 1.2.3 for detailed descriptions). Specifically, while some studies, such as those on English−Welsh (e.g., Parafita Couto et al., 2015), Spanish−English (e.g., Balam & Parafita Couto, 2019), Papiamento−Dutch (e.g., Parafita Couto & Gullberg, 2019), and Spanish−Dutch/Papiamento−Dutch (e.g, Van Osch et al., 2023), found evidence supporting predictions of both models, others like French−Dutch (e.g., Vanden Wyngaerd, 2017), Welsh−English (e.g., Parafita Couto et al., 2017; Vaughan-Evans et al., 2020), Spanish−English (e.g., Stadthagen-González et al., 2019), and Papiamento−Dutch (e.g., Pablos et al., 2019) suggested that adjective-noun patterns could be explained by either the MLF or MP, or neither of them (see Appendix 6.A in Chapter 6). Although these findings are inconclusive, the preference for noun insertion has been consistently observed in previous studies (e.g., Parafita Couto et al., 2015; Vanden Wyngaerd, 2017; Van Osch et al., 2023). This thesis therefore seeks to build on these findings by systematically examining how Spanish−Chinese bilinguals process syntactic constraints proposed by these two models in placing adjectives within mixed adjective-noun nominal constructions. In doing so, it aims to deepen our understanding of bilingual syntactic integration while also exploring whether these bilingual speakers exhibit a preference for noun insertion as part of their integration strategy.

Taken together, the combined investigation of grammatical gender, classifiers, and adjective placement is crucial for a comprehensive understanding of bilingual language processing. These three lexico-syntactic features represent distinct but parallel grammatical systems that vary significantly across Spanish and Chinese. Grammatical gender is exclusive to Spanish, while classifiers are unique to Chinese; however, both function as noun-based lexico-syntactic features that require selection based on the morpho-phonological or semantic properties of the noun. Despite their language-specific nature, they offer a valuable basis for examining how bilinguals manage comparable grammatical functions across different linguistic systems. For instance, analyzing bilinguals' processing of grammatical gender in unilingual Spanish and classifiers in unilingual Chinese enables us to investigate whether similar strategies are employed across these structurally distinct systems. Moreover, exploring these features in bilingual Spanish−Chinese contexts sheds light on the extent to which shared processing mechanisms underlie cross-linguistic integration. However, given that grammatical gender and classifiers are language-specific lexico-syntactic features that exist independently in each language, they cannot directly inform us about how the two linguistic systems are integrated or interact in the bilingual mind. In contrast, adjectives are present in both Spanish and Chinese but differ in their syntactic placement. This shared yet structurally divergent feature provides a unique opportunity to examine how bilinguals resolve syntactic conflicts when both languages encode the same lexico-syntactic feature differently across their languages. While Spanish adjectives encode grammatical gender, which is absent in Chinese adjectives, we focus here on the shared feature, namely the linear order of adjectives and nouns, rather than the language-specific feature of grammatical gender. Moreover, since grammatical gender does not affect the placement of adjectives, it is not considered in our discussion of adjective placements. As such, adjective placement complements the study of grammatical gender and classifiers by enabling us to explore an aspect that is not fully captured by language-specific features alone. Collectively, the examination of grammatical gender, classifiers, and adjective placement offers a more holistic understanding of how lexico-syntactic features are processed and integrated in bilingual speech.

The overarching aim of this dissertation is to uncover how grammatical constraints, task types/demands, and underlying cognitive processing mechanisms guide bilinguals' production and comprehension of lexico-syntactic features across languages, in both unilingual and bilingual contexts. This is accomplished through four core areas of exploration: first, we examine the gender congruency effect in unilingual Spanish NP production using data from Spanish speakers and establish a baseline for how grammatical gender is processed in unilingual contexts. Second, we explore the classifier congruency effect in unilingual Chinese NP production, drawing on data from both Mandarin Chinese speakers and early Spanish−Chinese bilinguals. This comparison helps determine whether bilinguals process classifiers in a way similar to Mandarin Chinese speakers and provides a basis for understanding how bilinguals employ classifiers in unilingual Chinese. Together, these first two investigations provide initial insights into how comparable lexico-syntactic features across languages are processed and produced in unilingual contexts. Third, we analyze the strategies bilinguals use to assign grammatical gender and classifiers in mixed NPs through multi-task approaches, aiming to understand how bilinguals navigate and assign lexico-syntactic features in bilingual contexts, particularly during code-switching. Fourth, we investigate how bilinguals resolve syntactic conflicts that arise from the different word order preferences of adjectives in Spanish and Chinese, aiming to understand how bilinguals integrate their two language in bilingual contexts.

The remainder of this section is structured to provide a detailed description of these lexico-syntactic features in both unilingual and bilingual contexts. First, section 1.2 offers an overview of key lexico-syntactic features across languages in both unilingual and bilingual contexts. Specifically, subsection 1.2.1 reviews grammatical gender, with a focus on Spanish in particular. Subsection 1.2.2 introduces the classifier system, especially as realized in Mandarin Chinese. Subsection 1.2.3 discusses adjective placement patterns in Spanish and Chinese. Together, these sections lay the groundwork for investigating how bilingual speakers manage, process, and produce these features in both unilingual and bilingual contexts. Furthermore, section 1.3 introduces the importance of open science and bilingualism as discovery science in scientific research. Finally, section 1.4 lists the main examination of each chapter in the thesis.

1.2 Overview of lexico-syntactic features across languages in unilingual and bilingual contexts

To establish a solid foundation for understanding how the lexico-syntactic features are processed and produced in both unilingual and bilingual contexts, this section begins with a broad overview of the main properties of lexico-syntactic features across languages. We then narrow our focus to two languages of interest, Spanish and Chinese, by introducing the core characteristics of their respective lexico-syntactic systems and how these features are processed and produced in both unilingual and bilingual contexts.

1.2.1 Grammatical gender: an overview of Indo-European languages with a focus on Spanish

In many languages, morphosyntactic classification systems are employed to classify their nominal lexicon, with two of the most widespread features being grammatical gender in Indo-European languages and lexico-syntactic numeral classifiers in East Asian languages (Seifart, 2010; Aikhenvald, 2000; Parafita Couto et al., in press). In Indo-European languages, i.e., those that still have a gender system, grammatical gender is considered a lexico-syntactic feature to classify nouns and plays an integral role in computing agreement within the NPs (Corbett, 1991; Schriefers & Jescheniak, 1999). In gender-marking languages, human and animate nouns are primarily assigned grammatical gender categories that are related to biological sex (e.g., *maestro* "male teacher"; *maestra* "female teacher"). Likewise, inanimate nouns, which refer to non-living objects, are also marked exclusively for grammatical gender, although this marking is not deducible from the noun's meaning (e.g., the noun *mesa* "table" is feminine; *abrigo* "coat" is masculine; Corbett, 2013). In other words, the distinction of grammatical gender to inanimate nouns has no conceptual basis and is therefore considered relatively arbitrary and invariable, as each noun is assigned to only one gender form (e.g., feminine or masculine in Spanish; Corbett, 1991). Moreover, nouns denoting the same referent may receive different grammatical gender assignment across languages (Corbett, 1991), as nouns often receive distinct gender classifications depending on the language. For instance, the word *car* is masculine in Spanish (*el*$_{MASC}$ *coche*$_{MASC}$) but feminine in French (*la*$_{FEM}$ *voiture*$_{FEM}$) and neuter in German (*das*$_{NEU}$ *Auto*$_{NEU}$). Despite this cross-linguistic variability, gender is fundamentally a system of agreement classes, with agreement serving as its defining characteristic (Hockett, 1958; Corbett, 1991).

In gender-marking languages, grammatical gender is an inherent property of nouns, typically manifested through agreement with associated words such as determiners and adjectives, which must align with the gender of the noun they modify (Hockett, 1958; Corbett, 1991;). According to the LRM speech production model (Levelt et al., 1999), grammatical gender assignment generally occurs at the lexical and syntactic levels. At the lexical level, the grammatical gender of a noun is activated and retrieved as part of its syntactic feature, with each noun assigned to one specific gender category (e.g., masculine or feminine in Spanish). When multiple lexical candidates are activated during speech planning, their gender features are encoded and compete for selection at the lemma level within the mental lexicon. At the syntactic level, gender agreement is established between the noun and other elements within NPs, such as determiners and adjectives, with these elements reflecting gender agreement with the head noun. Accordingly, gender agreement between nouns and their accompanying determiners or adjectives within NPs is easily observed in many Indo-European languages, such as Romance languages (e.g., Spanish, Italian, French) and Germanic languages (e.g., German, Dutch, see Wang & Schiller, 2019 and Sá Leite et al., 2022 for overviews).

A well-established phenomenon that illustrates the role of gender agreement in NP processing is the ***gender congruency effect***, which reveals how the gender agreement between nouns and their modifiers (i.e., nouns and modifiers agree or disagree in gender) significantly affects the speed and accuracy of NP production (see Wang & Schiller, 2019; Sá Leite et al., 2022; and Bürki et al., 2023 for an overview). Previous studies have used the PWI paradigm (Rosinski et al., 1975) to explore the gender congruency effect, where participants name the target picture while ignoring the simultaneously presented distractor word. Studies have consistently found faster naming latencies when distractors and target nouns were gender-congruent (i.e., both share the same grammatical gender) compared to gender-incongruent pairings (i.e., nouns and distractors have distinct grammatical genders) in NP production in Germanic languages, such as German (e.g., Bürki et al., 2016; Heim et al., 2009; Schiller, 2013; Schiller & Caramazza, 2003; Schiller & Costa, 2006; Schriefers & Teruel, 2000) and Dutch (e.g., La Heij et al., 1998; Schiller, 2013; Schiller & Caramazza, 2003, 2006; Starreveld & La Heij, 2004).

In contrast to the consistent gender congruency effects in Germanic languages, findings from Romance languages, particularly Spanish, have been more conflicting and less conclusive. For instance, Spanish, one of the primary languages of investigation in this thesis, employs a binary gender system in which all nouns possess an inherent grammatical gender, either masculine or feminine (Roca, 1989; Harris, 1991). This gender distinction is often predictable due to the relatively transparent morpho-phonological structure of nouns in Spanish, i.e., *-o* are predominantly masculine, *-a* are often feminine, and *-e* can be either gender (Roca, 1989; Harris, 1991). Within Spanish NPs, determiners and adjectives do not carry inherent gender themselves, but instead exhibit agreement by aligning with the gender of the noun they modify (e.g., *la manzana roja* [the_{FEM} $\text{apple}_{\text{FEM}}$ red_{FEM}] "the red apple", Hopp, 2016). As such, gender agreement is relatively transparent and easily observed in Spanish NPs. Despite this structural transparency, empirical findings on the gender congruency effect in Spanish remain inconsistent and contradictory. For instance, Paolieri et al. (2010) reported a reversed gender congruency effect in Spanish bare noun production, while other studies found no such effect, as in Finocchiaro et al. (2011) and O'Rourke (2007) (see detailed descriptions in Chapter 2 of this thesis). Further inconsistencies arise in Spanish NP production, where Von Grebmer zu Wolfsthurn et al. (2021) observed a cross-language gender congruency effect among German second language (L2) learners of Spanish, whereas Costa et al. (1999b) reported no such effect in native Spanish speakers. This variability highlights the need for a more nuanced understanding of how grammatical gender is processed and produced in Spanish NP production, thereby motivating the current investigation into unilingual Spanish NP production by Spanish speakers in Chapter 2.

However, how is grammatical gender processed in bilingual contexts, particularly in situations where code-switching frequently occurs? In previous research on code-switching, it has been found that bilinguals tend to frequently code-switch between gender-marked determiners and gender-marked or non-gender-marked nouns within mixed NPs in naturalistic data (e.g., Parafita Couto & Gullberg, 2019; Deuchar, 2005, 2006). This raises great interest in how grammatical gender is assigned for nouns, especially for non-gender-marked nouns, within mixed NPs. By investigating this phenomenon, three main gender assignment strategies have been identified and well-documented using different methodologies across language combinations (e.g., Spanish−English, Spanish−Basque, Purepecha−Spanish, see overview in Appendix 3.A of Chapter 3). This includes the default strategy, the translation equivalent strategy, and the shape-based strategy (see Bellamy & Parafita Couto, 2022 for an overview).

The first well-documented strategy is the default strategy, which refers to a tendency to default one gender to most inserted nouns from the other language, irrespective of their translation equivalents, semantic properties, or morpho-phonological features (cf. Bellamy & Parafita Couto, 2022). For example, Spanish−English bilinguals in Miami, Florida, default Spanish masculine gender to a transparent feminine human-referent English noun and produce *un renaissance woman* "the$_{\text{MASC}}$ renaissance woman", rather than *una renaissance woman* "the$_{\text{FEM}}$ renaissance woman" (Valdés Kroff, 2016). The second strategy is the translation equivalent strategy, also known as analogical criterion (Liceras et al., 2008) or analogical gender (Montes-Alcalá & Lapidus Shin, 2011), in which speakers tend to assign gender to most other-language inserted nouns based on the gender of their translation equivalents in the matrix or recipient language in mixed NPs (cf. Bellamy & Parafita Couto, 2022). For instance, Spanish-English bilinguals associate the English word *deer* with its masculine Spanish translation equivalent *venado*$_{\textit{MASC}}$, producing a mixed NP such as *el deer* "the$_{\text{MASC}}$ deer" (Jake et al., 2002). The third is the shape-based strategy, which also refers to "phonological gender criterion" (Poplack et al.,1982) or "phonological strategy" (Munarriz-Ibarrola et al., 2022). This strategy has primarily been observed among bilinguals with alphabetic languages that share similar morpho-phonological properties (cf. Bellamy & Parafita Couto, 2022). In such cases, bilinguals tend to rely on the phonological or morpho-phonological cues of nouns from the donor language when assigning gender in the recipient language (cf. Bellamy & Parafita Couto, 2022). For instance, Basque−Spanish bilinguals produce *la liburua* "the$_{\text{FEM}}$ book" where they associate the Basque definite marker *-a* with the Spanish feminine marker *-a*, although Basque does not feature a gender system (Parafita Couto et al., 2016).

In the previous studies, these three strategies have been consistently documented in various language combinations (see Bellamy & Parafita Couto, 2022), some of which involve a pairing where one language has grammatical gender while the other does not (e.g., Spanish−English in Balam et al., 2021 and Liceras et al., 2008; Spanish−Basque in Parafita Couto et al., 2016; or Spanish−Purepecha in Bellamy et al., 2018; also see Bellamy & Parafita Couto, 2022 for an overview), and other combinations involve languages in which one or both have two or more gender categories (e.g., Spanish−German in González-Vilbazo, 2005). Building on this background, we have already gained a broader understanding of how bilinguals employ strategies to assign grammatical gender in different language combinations in code-switching contexts. Despite this progress, some language pairs remain understudied. One such language pair is Spanish−Chinese, which, to date, has received little attention and remains largely unexplored, making it a valuable case for examining how the presence or absence of grammatical gender across bilinguals' languages shapes gender assignment strategies. As such, we aim to examine how Spanish−Chinese bilinguals assign Spanish grammatical gender to Chinese nouns in Spanish−Chinese mixed NPs, and, accordingly, to determine which gender assignment strategies they would employ in Chapter 3.

1.2.2 Classifiers: an overview of East Asian languages with a focus on Chinese

Unlike Indo-European languages, which often feature grammatical gender systems, most East Asian languages do not categorize nouns based on grammatical gender; instead, they use lexico-syntactic numeral classifier systems (Aikhenvald, 2000; Parafita Couto et al., in press). These classifier systems in most East Asian languages typically comprise extensive sets of classifiers, often ranging from dozens to hundreds, many of which overlap in form or meaning with open-class nouns (see Seifart, 2010; Bisang, in press; Parafita Couto et al., in press). In such systems, the numeral classifier functions to categorize count nouns into distinct semantic groups, with each group corresponding to a specific classifier (Gil, 2005). These classifiers are obligatorily used in numeral constructions, where they appear adjacent to numerals, but not necessarily to the nouns (Greenberg, 1972). However, the position of classifiers is often not consistent across East Asian languages. For example, in Chinese, classifiers typically precede the noun, as in *一只鸟*, /yi1zhi1niao3/ [one + classifier-zhi1+ bird], "one bird", which follows the grammatically acceptable structure of [numeral + classifier + noun]. In contrast, Japanese and Korean allow classifiers to appear either before or after the noun (see Qian, in press for an overview). For example, in Japanese, both *三羽の鳥*, /san wa no tori/ [three + classifier-wa + genitive case marker + bird], "three birds", where the classifier precedes the noun, and *鳥三羽*, /tori san wa/ [bird + three + classifier-wa], "three birds", where the classifier follows the noun, are grammatically acceptable (see Qian, in press for an overview). Importantly, the selection of classifiers largely depends on the semantic properties of nouns, in which classifiers largely align with the semantic properties of the noun they modify (Seifart, 2010).

Chinese serves as a primary example of the numeral classifier language, in which a numeral classifier system is used for categorizing nouns. In Chinese NPs, classifiers are obligatorily positioned between numerals or demonstratives and nouns (i.e., demonstrative/numeral + classifier + noun, Li & Thompson, 1981). The selection of classifiers within Chinese NPs largely relies on semantic properties of nouns, such as animacy, shape, size, function, and consistency, though this transparency does not apply uniformly to all nouns (Li & Thompson, 1981; Myers & Tsay, 2000). For example, in the phrase *一把勺子*, /yi1ba3shao2zi0/ [one + classifier-ba3 + spoon], "one spoon", both the classifier *把* /ba3/ and the noun *勺子* /shao2zi0/ "spoon" refer to things that typically have a handle (Guo, 2002). This relative transparency of classifier choice has raised considerable interest in research on classifier processing, particularly in how speakers access and retrieve classifiers during speech production.

Following the LRM speech production model (Levelt et al., 1999), previous research suggested that numeral classifiers, as a lexico-syntactic feature, undergo a selection process that is similar to that of grammatical gender in gender-marking languages (e.g., Wang et al., 2019; also see Wang & Schiller, in press and Qian, in press for an overview). One phenomenon that supports this parallel is an analogous effect observed in Chinese, i.e., the ***classifier congruency effect***, which reflects how congruent or incongruent classifier-noun pairings influence the speed and accuracy of Chinese NP production. Specifically, previous studies using the PWI paradigm found that Mandarin Chinese speakers tend to produce classifier-NPs more quickly and accurately when the classifier is congruent with the noun, as opposed to when there is incongruency in NP production (e.g., Wang et al., 2006; Huang & Schiller, 2021) but not in bare noun production (e.g., Wang et al., 2006; Wang et al., 2019). While the classifier congruency effect has been documented in Chinese speakers, little is known about how bilingual speakers, particularly those whose other language does not employ a classifier system, process and retrieve classifiers during speech production. To address this gap, Chapter 4 of this thesis investigates the classifier congruency effect in the Chinese NP production of early Spanish−Chinese bilinguals, offering insights into how lexico-syntactic features are processed in bilinguals' minds.

In addition to exploring the phenomenon of classifiers in unilingual Chinese contexts, this bilingual population also provides a special opportunity to examine how classifiers are processed and assigned within mixed NPs during code-switching. However, to date, limited studies have specifically investigated this aspect among Spanish−Chinese bilinguals. Given that both grammatical gender and classifier function as noun-based categorization systems, and that Section 1.2.1 and Chapter 3 of this thesis have already discussed gender assignment strategies across bilingual communities and within the Spanish−Chinese bilingual group, it is thus equally important to examine how classifier assignment unfolds in code-switched contexts. This investigation helps build a more complete picture of how bilingual speakers navigate parallel lexico-syntactic features across their two languages. To set the stage for this investigation, we thus draw on the strategies of gender assignment in Section 1.2.1 to hypothesize analogous strategies for classifier assignment within mixed NPs in Chapter 5.

First, we propose a comparable default strategy for classifier assignment in mixed NPs. Comparable to Spanish, which features a default masculine gender, Chinese also possesses the general classifier *ge* (个 /ge4/) that is often used as a default choice for most Chinese nouns by Mandarin Chinese speakers in Mainland China, particularly with nouns lacking specific categories or involving unfamiliar concepts (see Myers & Tsay, 2000 and Erbaugh, 2006). By analogy, this suggests the plausibility of a potential default strategy that applies to classifier assignment in mixed NPs, in which the general classifier *ge* would be the default for most other-language inserted nouns. Thus, we posited a default strategy for classifier assignment in mixed NPs. Second, according to the rationale of the translation equivalent strategy in gender assignment, we also hypothesize a comparable translation equivalent strategy for classifier assignment in mixed NPs. By employing such a strategy, bilingual speakers might draw on the classifier typically associated with the Chinese translation equivalents of other-language inserted nouns.

However, unlike the first two strategies described above, we do not expect the shape-based strategy to appear for classifier assignment in Spanish−Chinese bilingual contexts. This is largely due to the typological and orthographic differences between the two languages. Specifically, this strategy has been predominantly observed in bilinguals with two alphabetic languages, where phonological or morpho-phonological cues, such as phonological or morphological endings of nouns, can serve as reliable gender cues across languages (cf. Bellamy & Parafita Couto, 2022). In contrast, Chinese is a logographic language that lacks both inflectional morphology and transparent phonological endings in typology, orthography, and morpho-phonological form. Moreover, classifier assignment in Chinese is not typically driven by the phonological or morphological shape of the noun; instead, it largely relies on the semantic properties of nouns. Furthermore, Spanish lacks a classifier system. Therefore, the shape-based strategy is unlikely to be used for classifier assignment in Spanish−Chinese bilingual contexts. Together, by analogically proposing two classifier assignment strategies, we aim to examine how bilinguals process the assignment of Chinese classifiers in Chinese−Spanish code-switched NPs in Chapter 5.

To sum up, adopting this dual perspective by examining both grammatical gender and classifier in unilingual and bilingual contexts is particularly valuable, as it sheds light on how early Spanish−Chinese bilinguals navigate two distinct but systematic mechanisms in mixed NPs involving different lexico-syntactic features. Studying both features provides us with a broader understanding of how bilinguals manage typologically different grammatical systems and whether they rely on similar or distinct cognitive processes when processing these diverse lexico-syntactic features.

1.2.3 Adjective placement in Spanish and Chinese

We have established how grammatical gender and classifier systems, language-specific features, are featured and processed in Spanish and Chinese, respectively. It is also crucial to turn to another key aspect of lexico-syntactic features: adjective placement, a shared feature that is present in both languages but differs in its position. Specifically, Spanish predominantly employs a postnominal adjective order, placing the adjective after the noun in most nominal constructions (Terker, 1985). For example, *flor roja*, [flower red], "red flower". This surface postnominal order in Spanish (i.e., Romance language) is generally explained by overt noun movement across the adjective, triggered by a strong EPP (Extended Projection Principle) feature in the agreement phrase (AgrP), as described by Cinque (1994, 1999) within the universal base structure. In this account, adjectives are base-generated prenominally, but the noun raises past them to satisfy language-specific syntactic requirements, resulting in the surface postnominal order typical of Spanish. Cantone and MacSwan (2009) further propose that, in code-switched speech, the adjective placement should be determined by the language of the adjective itself. Moreover, considering such features of adjectives in code-switched speech, in which the adjective placement is determined by the lexical item (i.e., the adjective itself, proposed by the MP model) while at the same time shaped by language-specific syntactic rules (e.g., to check strong agreement features before pronunciation), we thus treated adjective placement, particularly adjective-noun word order, as a lexico-syntactic feature in this thesis. Nevertheless, it is important to note that the following paragraphs also address alternative explanations for adjective positioning offered by the MLF model (Myers-Scotton, 1993, 2002).

Unlike Spanish, Chinese employs two common adjective-noun structures, adjective *de* noun (A *de* N) and adjective plus noun (A N), with the adjective consistently preceding the noun in both structures (Li & Thompson, 1981; Paul, 2010). The "A N" form typically categorizes the noun as a subtype (e.g., *黄豆*, /huang2dou4/ [yellow bean], "soybean"), while the "A *de* N" structure uses *de* (i.e., *的* /de0/) as a linker to specify or clarify features of the noun (e.g., *红色的花*, /hong2se4de0hua1/ [red flower], "red flower") (Li & Thompson, 1981; Paul, 2010). In general, the "A *de* N" structure is the more frequent pattern for adjectival modification than the "A N" structure in Chinese (Li & Thompson, 1981), and it is therefore the primary focus of this thesis.

In recent years, adjective placement in code-switched speech has been well-studied and discussed in several studies (e.g., Cantone and MacSwan, 2009; Myers-Scotton, 1993, 2002; Balam & Parafita Couto, 2019; Stadthagen-González et al., 2019; Parafita Couto & Gullberg, 2019; Pablos et al., 2019; Vanden Wyngaerd, 2017; Parafita Couto et al., 2015; Parafita Couto et al., 2017; Vaughan-Evans et al., 2020; Van Osch et al., 2023), particularly because it provides an informative lens into how bilingual speakers manage structures that exist in both languages but are governed by distinct syntactic rules. Such structures represent what Poplack and Meechan (1998) term "conflict site", where the grammatical rules of two languages differ. Adjective placement thus provides an ideal case for investigating how bilinguals manage lexico-syntactic integration across Spanish and Chinese.

Two influential theoretical frameworks, the Matrix Language Frame (MLF, Myers-Scotton, 1993, 2002) and the Minimalist Program (MP)-based approaches (MacSwan, 1999), have been proposed to account for the adjective placement conflict in bilingual speech. Specifically, the MLF (Myers-Scotton, 1993, 2002) suggests that in code-switching, one language, the ML, provides the morphosyntactic structure, while the other language, the embedded language (EL), contributes the lexical items. As such, in adjective-noun code-switched phrases, the MLF (Myers-Scotton, 1993, 2002) predicts that the adjective placement should follow the word order of the ML. On the other hand, the MP-based approaches (Cantone & MacSwan, 2009) suggest that the grammatical features in code-switched phrases should be determined by the lexical items themselves. Moreover, following Cinque's (1994, 1999, 2005) universal base structure, Cantone and MacSwan (2009) reach the generalized description that the adjective placement is determined by the language of the adjectives. It is important to note that although these two models focus on different aspects and have distinct assumptions, their predictions about adjective placement can overlap or differ. For example, some sentences can be explained by both models (MLF+/MP+), some only by one of them (e.g., MLF+/MP− or MLF−/MP+), and others by neither (MLF−/MP−).

While the predictions of these models have been extensively examined across various language combinations, such as Spanish−English (e.g., Balam & Parafita Couto, 2019; Stadthagen-González et al., 2019), Papiamento−Dutch (e.g., Parafita Couto & Gullberg, 2019; Pablos et al., 2019), French−Dutch (e.g., Vanden Wyngaerd, 2017), Welsh−English (e.g., Parafita Couto et al., 2015; Parafita Couto et al., 2017; Vaughan-Evans et al., 2020), and Spanish−Dutch/Papiamento−Dutch (e.g., Van Osch et al., 2023), no definitive consensus has yet been reached in fully explaining grammatical patterns in adjective-noun code-switching. This lack of conclusive findings underscores the importance of further research with understudied language combinations to deepen our understanding of adjective-noun code-switched patterns. The Spanish−Chinese language combination is especially underexplored in this context, which could provide an ideal opportunity to address this gap and offer new insights into how bilinguals resolve syntactic conflicts involving adjective-noun order. Accordingly, Chapter 6 of this thesis aims not only to examine how bilinguals process grammatical constraints proposed by these models when placing adjectives in mixed adjective-noun constructions, but also to uncover novel aspects of code-switching behavior in this distinctive linguistic context.

1.3 Open science and bilingualism as discovery science

Beyond the core linguistic and psycholinguistic issues discussed so far, it is also important to situate this thesis within broader scientific practices. In particular, we turn to the emerging phenomenon of ***Open Science***, which has become increasingly influential in shaping contemporary scientific practices (Vicente-Saez & Martinez-Fuentes, 2018). Specifically, Open Science provides a framework for greater transparency, accessibility, shareability, and collaboration within the scientific community. Those goals are supported through practices, such as open data, open code, open access, open peer review, and pre-registration (Vicente-Saez & Martinez-Fuentes, 2018). The absence of such practices has been considered a major factor that caused the ***replication crisis*** in linguistics, particularly in experimentally oriented fields (Sönning & Werner, 2021; Ioannidis, 2005). For instance, a lack of transparency in methodology and data analysis, and the unavailability of original data and materials, have made it difficult to verify published results and to reproduce scholarly work (Sönning & Werner, 2021). To address these challenges, recent research has encouraged measures such as full transparency of data analysis and full accessibility of materials and code (Paquot & Callies, 2020), and the pre-registration of data collection and analysis plans (Chambers, 2013). This thesis aims to continuously contribute to these Open Science efforts in several ways. First, we ensure full transparency in our experimental designs, procedures, and in reporting participants' linguistic backgrounds in all main chapters. Second, we have made all experimental materials, data analyses, and statistical code for all chapters openly available and easily accessible through the Open Science Framework (OSF, Foster & Deardorff, 2017). Finally, we provide a detailed rationale for our statistical analytical choices and describe every step of our data analysis procedures thoroughly. By doing so, all data for each chapter in this thesis are transparent, accessible, and readily replicable.

Apart from concerns about Open Science, there has been growing attention on the replication crisis in bilingualism research, particularly regarding the reproducibility and replicability of published findings (Navarro-Torres et al., 2021). Previous research has pointed out that factors such as statistical power and sample size (e.g., Brysbaert, 2021), sampling and methodological noise (e.g., García-Pentón et al., 2016a, 2016b), and publication bias (Bialystok et al., 2015) undoubtedly affect how reproducible and interpretable bilingualism research is. However, it is the prescribed solutions to these factors, such as employing large samples (e.g., Brysbaert, 2021) or enforcing strict methodological uniformity (i.e., employing exactly the same methodologies as those used in the studies being replicated, García-Pentón et al., 2016a, 2016b), that have the strongest influence on shaping the current replication challenges in bilingualism research (Navarro-Torres et al., 2021). These prescribed solutions can be overly simplistic and fail to reflect the true nature of science as a process of discovery, one that involves exploration, adaptation, interpretation, and evolution, thereby easily misleading the direction of replication research (Navarro-Torres et al., 2021). The remedy for this issue is the generation of ***variety*** by applying diverse ideas, methods, and scientific practices, as it could help identify reliable signals (e.g., appropriate characterization, research practices, and research tools) from inconsistent observations to ensure incremental improvement in science (Navarro-Torres et al., 2021). This is evident, for example, in the study on grammatical gender assignment strategies in mixed NPs (i.e., Chapter 3 of this thesis), where findings and interpretations regarding the strategies employed by bilinguals have varied across studies, different methods, bilingual populations, social networks, and community characteristics (see Chapter 3 for detailed descriptions). In such cases, simply following rigid prescriptions, such as repeating a single study's methodologies or expanding the sample size, cannot guarantee meaningful and successful replication. Instead, identifying convergent patterns across diverse studies and maintaining *variety*, such as exploring new hypotheses, employing multi-method approaches, and interpreting findings within the complex variables of bilingual contexts based on rich characterizations of data, is key to reaching reliable, replicable conclusions (i.e., reliable signals).

To sum up, bilingualism is complex, and we are still far from fully defining the limits and conditions of previous findings (Navarro-Torres et al., 2021). For replication efforts on bilingualism to be truly meaningful, it is essential to situate them within the broader view of science as an ongoing process of discovery, where incremental insights and methodological diversity are essential in advancing research on bilingualism. In line with these principles, the present thesis continues this effort by combining rich participant characterizations, accessible and transparent materials, multi-method approaches, and incremental understanding to support more reliable and replicable research on Spanish−Chinese bilingual language processing. Central to this endeavor is the systematic use of a multi-task approach, employing diverse methodologies as a response to the complexity of bilingualism in the context of this thesis. It reflects an effort to extract reliable signals from inconsistent observations by capturing a relatively broad and comprehensive range of variables relevant to bilingual language processing. This approach underpins the bilingual investigation carried out in this thesis.

1.4 The current thesis

Taken together, the questions outlined above bring us back to the central question: how do bilingual speakers manage two distinct languages in their minds? Despite considerable research, several critical issues still remain to be addressed in order to answer this question: first, we aim to determine whether grammatical gender is competitively selected during Spanish NP production and further identify whether this competitive selection is reflected in the grammatical gender effect; second, we explore how Spanish grammatical gender is processed and assigned to Chinese nouns in mixed Spanish−Chinese NPs by Spanish−Chinese bilinguals and determine which gender assignment strategies are employed; third, we investigate whether classifiers are activated and competitively selected during Chinese NP production by both Mandarin Chinese speakers and Spanish−Chinese bilinguals and whether such selection processes manifest as a classifier congruency effect; fourth, we examine how classifiers are processed and assigned to Spanish nouns in mixed Chinese−Spanish NPs by Spanish−Chinese bilinguals and identify the corresponding classifier assignment strategies they used in mixed NPs; fifth, we explore how Spanish−Chinese bilinguals resolve syntactic conflicts in adjective placement across their two languages in mixed adjective-noun constructions, aiming to not only examine how grammatical constraints influence bilinguals' processing of adjective placements but also uncover insights into code-switching behaviors in this unique linguistic context. Overall, our core research questions center on a unifying theme: how are lexico-syntactic features, namely grammatical gender, classifiers, and adjective placement, processed and produced in bilingual and/or unilingual contexts across Spanish and Chinese?

In **Chapter 2** of this thesis, we first explore the effects of grammatical gender on Spanish NP production by Spanish speakers. Specifically, we examine the *gender congruency effect* on Spanish determiner-NP production by measuring Spanish speakers' naming latencies. Previous studies exploring the presence of the gender congruency effect in Spanish language production obtained controversial results, in which this effect was reported as either existing (e.g., Von Grebmer zu Wolfsthurn et al., 2021) or non-existent (e.g, Costa et al., 1999b; Finocchiaro et al., 2011; O'Rourke's, 2007), or even a reverse effect was observed (e.g., Paolieri et al., 2010). To tackle this critical issue, we employ a PWI paradigm to address the following questions: first, whether the selection process of grammatical gender is competitive in Spanish; second, if the process is competitive, whether this competition of grammatical gender surfaces as a variation in naming latencies is reflected by the gender congruency effect in Spanish. In this, we seek to characterize how Spanish speakers process grammatical gender in unilingual Spanish determiner-NP production. This study is important, as establishing the gender congruency effect in Spanish is not just a "baseline" but a critical behavioral evidence for challenging existing inconsistencies in Romance language literature. Additionally, this study offers valuable insights into how Spanish speakers process grammatical gender in Spanish NP production within unilingual Spanish contexts, laying a foundational understanding for investigating lexico-syntactic features not only in bilingual contexts but also in typologically different languages that share comparable lexico-syntactic features (e.g., classifier in Chinese).

In **Chapter 3**, we shift the focus from unilingual to bilingual contexts to investigate how grammatical gender is processed in mixed determiner-NPs by bilingual speakers. Yet, little research has explored the processing and assignment of grammatical gender in bilingual contexts involving Spanish and Chinese. Therefore, we employ a multi-task approach to test how early Spanish−Chinese bilinguals process and assign Spanish grammatical gender to Chinese nouns in mixed Spanish−Chinese NPs. In other words, we examine which grammatical gender categories (e.g., masculine or feminine) bilinguals tend to assign to Chinese nouns through their choice of Spanish determiners. Next, we identify the corresponding gender assignment strategies that bilingual speakers prefer when selecting grammatical gender in mixed NPs. The critical contribution of this study lies in being the first to examine how grammatical gender is assigned in mixed Spanish−Chinese NPs by Spanish−Chinese bilinguals. It documents that Spanish−Chinese bilinguals do not rigidly adhere to one linguistic rule or strategy, but instead flexibly adapt their language processing and production based on task demands and sociolinguistic context. As such, this study adds a crucial piece to the broader picture of how grammatical gender, as one of the lexico-syntactic features, is processed in bilingual contexts.

In **Chapter 4**, we turn our attention to the effects of classifiers in Chinese, a comparable lexico-syntactic feature that exists in a typologically different language from Spanish. Previous research on Chinese NP production is scarce, particularly regarding the role of Chinese classifiers in both Mandarin Chinese speakers and early Spanish−Chinese bilinguals with electrophysiological measures. Therefore, we investigate how Chinese classifiers are processed during Chinese NP production in overt picture-naming tasks by both early Spanish−Chinese bilinguals and Mandarin Chinese speakers. Specifically, we examine whether the *classifier congruency effect* occurs in Chinese NP production in both groups. To do so, we measure naming latencies and probe electrophysiological correlates, with particular focus on the modulation of the N400-like component elicited by classifier violations, during the production of Chinese NPs in four manipulated conditions. This study contributes to crucial behavioral and neurocognitive evidence for the robustness of the classifier congruency effect in bilingual language production, advancing our understanding of how bilinguals process classifiers in ways that closely resemble the patterns observed in monolingual Chinese speakers in previous studies. This finding is important to the thesis because it extends the investigation of lexico-syntactic processing beyond grammatical gender, showing that bilinguals also demonstrate sensitivity to classifier-NPs in production and underscoring the parallel cognitive processing mechanism in both gender and classifier systems. As such, this study provides a more complete picture of how bilinguals manage different lexico-syntactic systems across their two languages.

In **Chapter 5**, we extend the investigation by examining how bilingual speakers assign classifiers during the processing and production of code-switched classifier-NPs. Building on the fundamental understanding of classifier processing in Chinese NP production within unilingual contexts established in Chapter 4, this chapter expands the research to bilingual contexts to examine how bilinguals assign classifiers when processing and producing mixed Chinese−Spanish NPs. To do so, we employ a multi-task experimental design parallel to that used in Chapter 3 to determine which classifier assignment strategies bilinguals will use in mixed NPs. This approach allows us to examine whether bilinguals apply similar strategies when assigning these distinct lexico-syntactic features in mixed NPs within bilingual contexts. This study contributes to the first demonstration that bilinguals flexibly employ default and translation equivalent strategies for classifier assignment based on task demands and sociolinguistic context. It reveals parallel patterns with gender assignment strategies observed in the same bilingual community in Chapter 3, highlighting bilingual flexibility and adaptability in navigating multiple linguistic systems. Additionally, this particular study completed a piece of the puzzle to this thesis in understanding how lexico-syntactic features are processed within NPs across unilingual and bilingual contexts.

In **Chapter 6**, we present a study on adjective-noun order in mixed Spanish−Chinese nominal constructions, conducted with the same group of early Spanish−Chinese bilingual speakers examined in Chapter 5. Building on the previous four chapters, which provided a relatively comprehensive understanding of how grammatical gender and classifiers are processed in unilingual and bilingual contexts, we shift our focus to adjective placement. This is because grammatical gender and classifiers are language-specific features unique to Spanish and Chinese, respectively, and thus offer limited insight into how the two systems interact. In contrast, adjective placement represents a shared structure with contrasting word orders, which allows us to explore how bilinguals navigate structural differences between their languages. Thus, we examine how early Spanish−Chinese bilinguals resolve the conflict of adjective placement in mixed Spanish−Chinese nominal constructions. Specifically, we probe how syntactic constraints, proposed by two theoretical models (i.e., the MLF model and MP-based approaches), shape adjective-noun patterns in these code-switched constructions. Moreover, we examine whether noun insertion, which is favored by bilinguals in other communities, is similarly preferred by Spanish−Chinese bilinguals. This study is conducted in both unilingual and bilingual contexts to provide a comprehensive understanding of whether bilinguals process adjective-noun constructions in similar ways across language contexts. This allows us to further explore how bilinguals navigate linguistic interaction in two languages with divergent grammatical systems. This study addresses syntactic differences with an underexplored bilingual population and language pair, contributing to the growing body of research on adjective placement in bilingual speech. Importantly, this study offers the final analytical piece in understanding how lexico-syntactic features are processed and integrated across Spanish and Chinese in both unilingual and bilingual contexts, completing the broader investigation undertaken in this thesis.

Chapter 7 provides an integrated discussion of the research findings from each chapter, emphasizing their theoretical implications. Furthermore, we identify potential limitations of the different studies and outline directions for future research to provide a more comprehensive understanding of how lexico-syntactic features are processed in both unilingual and bilingual contexts.

Chapter 2

Gender Congruency Effects in Spanish: Behavioral Evidence from Noun Phrase Production

---

A version of this chapter has been published in *Brain Sciences* as: Wu, R., & Schiller, N. O. (2023). Gender congruency effects in Spanish: behavioral evidence from noun phrase production. *Brain Sciences*, *13*(4), 696.

34 *The Role of Lexico-Syntactic Features in Noun Phrase Production and Comprehension*

Abstract

Grammatical gender as a lexico-syntactic feature has been well explored, and the gender congruency effect has been observed in many languages (e.g., Dutch, German, Croatian, Czech, etc.). Yet, so far, this effect has not been found in Romance languages such as Italian, French, and Spanish. It has been argued that the absence of the effect in Romance languages is due to the fact that the gender-marking definite article is not exclusively dependent on the grammatical gender of the head noun, but also on its onset phonology (e.g., *lo zucchero* "the sugar" in Italian, not *il zucchero*, *il* being the default masculine determiner in Italian, while *lo* is often used when masculine nouns begin with z, s+consonant, gn, ps, or similar clusters). For Spanish, this argument has also been made because feminine words starting with a stressed /a/ take the masculine article (e.g., *el água* is "the water", not *la água*). However, the number of words belonging to that set is rather small in Spanish, and it may be questionable whether or not this feature can be taken as an argument for the absence of a gender congruency effect in Spanish. In this study, we investigated the gender congruency effect in Spanish noun phrase production. We measured 30 Spanish speakers' naming latencies in four conditions via the picture-word interference paradigm by manipulating gender congruency (i.e., gender-congruent vs. gender-incongruent) and semantic relatedness (i.e., semantically related vs. semantically unrelated). The results revealed significantly longer naming latencies in gender-incongruent and semantically related conditions compared to gender-congruent and semantically unrelated conditions. This result suggests that grammatical gender as a lexico-syntactic feature in Spanish is used to competitively select determiners in Spanish speakers' noun phrases. Our findings provide an important behavioral piece of evidence for the gender congruency effect in Romance languages.

Keywords: language production; lexico-syntactic features; gender congruency effect; PWI; Spanish

2.1 Introduction

The study of grammatical gender processing has been a topic of interest for psycholinguists for decades. In gender-marking languages, such as Romance languages (e.g., Spanish, Italian, etc.) and Germanic languages (e.g., German, Dutch, etc.), grammatical gender processing is a crucial part of successfully producing a determiner noun phrase (NP) (e.g., in Spanish, *la manzana* "the$_{\text{FEM}}$ apple$_{\text{FEM}}$"). According to Levelt et al.'s speech production model (Levelt et al., 1999), speakers need to encode the to-be-produced word by conceptualizing the message first. Then, by lexicalizing the concept, the word's grammatical properties, such as its syntactic features (e.g., grammatical gender, number, case, etc.), are activated and eventually retrieved, and its corresponding phonological and phonetic forms are encoded. Finally, the phonetic motor representation of the word is articulated.

In a picture-word interference task, speakers are presented with a target picture and a distractor word. When speakers produce a determiner-NP in gender-marking languages, the concept and form of the target picture and the distractor, respectively, are activated first. Then, the syntactic representations of the target and distractor (via their word form), including the syntactic features (e.g., grammatical gender, number, case, etc.), are activated. In this case, all nodes that are conceptually related to the target are activated to different degrees, with the lexical node that is conceptually related to the target being the most highly activated. The most highly activated lexical node is then selected for production, including the retrieval of the phonological features for further word-form encoding. Finally, the determiner-NP is produced. According to this model, the gender feature of the to-be-produced word will be activated but only selected if it is needed for further production (e.g., for determiner-NP production, but not for bare noun naming, when the gender feature is not strictly needed for production).

The processing of grammatical gender in this speech production model is supported by numerous experimental studies of NP production in gender-marking languages such as German (e.g., Bürki et al., 2016; Heim et al., 2009; Schiller, 2013; Schiller & Caramazza, 2003; Schiller & Costa, 2006; Schriefers & Teruel, 2000) and Dutch (La Heij et al., 1998; Schiller, 2013; Schiller & Caramazza, 2003, 2006; Starreveld & La Heij, 2004; for a detailed description, see the next section). However, conflicts were observed in experimental studies in Romance languages (e.g., Spanish in Costa et al., 1999b; Italian in Caramazza et al., 2001; Miozzo et al., 2002; Miozzo and Caramazza, 1999; Catalan in Costa et al., 1999b; French in Alario and Caramazza, 2002, and Portuguese in Sá Leite et al., 2021), in which the correct grammatical gender was selected and produced in NP production, but no effect of this selection process was found. Therefore, the question arises as to why this selection/competition of grammatical gender is not reflected by a gender congruency effect in Romance languages. This study tackles this question experimentally.

2.1.1 The gender congruency effect

Gender agreement, generally represented by agreement between the noun and the determiner or adjective in the noun phrase (Hopp, 2016), is a key feature of gender-marking languages such as Romance languages (e.g., Spanish) as well as Germanic languages (e.g., German and Dutch). Nouns in these languages are assigned a gender (e.g., in Spanish, masculine or feminine), which is marked on associated determiners and adjectives, for example, in Spanish, *la manzana roja* (literally: the_{FEM} $\text{apple}_{\text{FEM}}$ red_{FEM}). In this example, the form of the determiner is *la* because *manzana* is a feminine noun. In other words, the determiners match the gender of the noun they accompany. The gender congruency effect, which entails faster and more accurate processing in cases of a match between the gender of nouns and their associated determiners or adjectives, has been studied extensively in Romance languages (e.g., Cubelli et al., 2005; Paolieri et al., 2011; Sá Leite et al., 2021; Von Grebmer zu Wolfsthurn et al., 2021), as well as in German (e.g., Bürki et al., 2016; Heim et al., 2009; Schiller, 2013;

Schiller & Caramazza, 2003; Schiller & Costa, 2006; Schriefers & Teruel, 2000), Dutch (e.g., La Heij et al., 1998; Schiller, 2013; Schiller & Caramazza, 2003, 2006; Starreveld & La Heij, 2004) and some other gender-marking languages (for an overview, see Wang & Schiller, 2019 and Sá Leite et al., 2022; for a recent meta-analysis, see Bürki et al., 2023).

The gender congruency effect in language production has been investigated in experimental studies using the picture-word interference (PWI) paradigm (Rosinski et al., 1975). In this experimental paradigm, participants are asked to name a picture while ignoring a distractor word presented shortly before, at the same time, or shortly after picture onset. It has been found that the reaction time to name the picture is affected by the relationship between the distractor and the target picture. In the study of Schriefers (1993), the PWI task was initially employed to investigate how grammatical gender (i.e., in Dutch, common and neuter) is processed by native Dutch speakers. He manipulated the gender congruency between target pictures and distractors, i.e., creating gender-congruent conditions (e.g., a target picture of *boek* "book$_{NEU}$" with the distractor *dak*, "roof$_{NEU}$") and gender-incongruent conditions (e.g., a target picture of *boek* "book$_{NEU}$" with the distractor *tafel* "table$_{COM}$"). Participants were presented with a target picture along with a gender-congruent or -incongruent distractor at the same time and asked to name the picture using a noun phrase while ignoring the distractor. Faster naming latencies were obtained in the gender-congruent condition than in the gender-incongruent condition, coined as the gender congruency effect. Schriefers (1993) interpreted the gender congruency effect as the result of grammatical gender features of targets and distractors competing for selection in participants' noun phrase production in gender-incongruent conditions.

Experimental research has shown a consistently faster response time for gender-congruent conditions than for gender-incongruent conditions in noun phrase production in German (e.g., Bürki et al., 2016; Heim et al., 2009; Schiller, 2013; Schiller & Caramazza, 2003; Schiller & Costa, 2006; Schriefers & Teruel, 2000) and Dutch (e.g., La Heij et al., 1998; Schiller, 2013; Schiller & Caramazza, 2003, 2006; Starreveld & La Heij, 2004). Bürki et al. (2016), for instance, conducted a picture naming task in German using the PWI paradigm by manipulating two factors, i.e., gender congruency and phonological congruency. Participants were asked to name the pictures using noun phrases and ignore the distractors. As the grammatical gender of the target picture is selected in competition with distractors during NP production (determiner + noun or determiner + adjective + noun), variations in the naming response times were found to depend on the gender and phonological congruency status. Both the gender-congruent condition and the phonologically congruent condition were faster than the corresponding incongruent conditions. The consistent gender congruency effect was found in many studies in the NP language production of German and Dutch (for an overview, see Wang & Schiller, 2019 and Sá Leite et al., 2022).

Nevertheless, conflicts have been observed in the attempts to replicate the gender congruency effect in Romance languages. The gender congruency effect in Italian was successfully replicated in the production of bare nouns (e.g., Paolieri et al., 2010, 2011 and Cubelli et al., 2005), but not in the production of noun phrases (e.g., *il gatto* “the cat”) in Cubelli et al.’s (2005) research. In Cubelli et al.’s (2005) study, a gender congruency effect with an unexpected direction was found in Italian bare noun production. Longer naming latencies were observed in the gender-congruent condition than in the gender-incongruent condition. This effect has been successfully replicated in three experiments with different materials (e.g., Paolieri et al., 2010, 2011). However, Finocchiaro et al. (2011) reported the absence of a gender congruency effect in their experimental work on Italian, Spanish, and French using bare noun naming. They attempted to replicate the study of Cubelli et al. (2005) by testing native Italian speakers on bare noun production. However, no gender congruency effect was found with either transparent or opaque distractors in the two experiments. Similarly, naming latencies in their Spanish and French bare noun production experiments were not affected by the gender of a distractor word presented with the target picture. On the contrary, Alario and Caramazza (2002) demonstrated significantly faster response times for gender-congruent conditions than for incongruent conditions in French NP production (e.g., determiner + noun and determiner + noun + adjective).

Moreover, O'Rourke's (2007) and Finocchiaro et al.'s (2011) replication studies did not result in a finding of gender congruency effects in Spanish bare noun production, but Paolieri et al. (2010) did. Furthermore, Paolieri et al. (2010) found a reversed gender congruency effect, i.e., participants responded faster when naming target pictures in Italian (e.g., *pera* "pear$_{\text{FEM}}$") with gender-incongruent distractors (e.g., *cervo* "deer$_{\text{MASC}}$") than with gender-congruent distractors (e.g., *calza* "sock$_{\text{FEM}}$"). Similarly, they also observed longer naming latencies in Spanish for target pictures (e.g., *mono* "monkey$_{\text{MASC}}$") with gender-matched distractors (e.g., *grifo* "tap$_{\text{MASC}}$") than with gender-unmatched distractors (e.g., *cartera* "wallet$_{\text{FEM}}$"). Additionally, Von Grebmer zu Wolfsthurn et al. (2021) observed a cross-language gender congruency effect in Spanish NP production with German speakers who were learning Spanish as a second language. Based on these contradictory findings in Romance languages, it is at least questionable whether or not the selection process of grammatical gender is competitive, and if so, whether or not this competitive process surfaces as a variation in naming latencies. In particular, the answer to the question of whether or not the congruency status between the grammatical gender of the targets and distractors has a significant effect on naming latencies remains unclear.

2.1.2 The grammatical gender feature in Spanish

Spanish, like many other Romance languages, has a gender system that distinguishes between masculine and feminine gender for nouns and their associated determiners and adjectives. Specifically, Spanish has a two-gender system including two-gender features for nouns (masculine and feminine), with the determiners and adjectives exhibiting gender agreement according to the lexical properties of the following nouns in NPs (e.g., determiner + noun and determiner + noun + adjective) (O'Rourke, 2007; White et al., 2004). The distribution of feminine and masculine gender values is approximately balanced in Spanish (Bull, 1984; Eddington, 2002). However, it has been argued that masculine and feminine gender may be represented differentially in Spanish, with masculine being the default gender, and feminine taking a more marked position in the language (Beatty-Martínez & Dussias, 2019).

As a lexico-syntactic feature, grammatical gender in Spanish applies to all nouns, and the grammatical gender agreement of determiners is obligatory within NPs, e.g., *la camisa* "the$_{\text{FEM}}$ shirt$_{\text{FEM}}$" (Wang & Schiller, 2019). Many nouns are morphologically and/or phonologically marked by grammatical gender (Wang & Schiller, 2019), and the selection of determiners and adjectives depends on the phonological and morphological forms of nouns (Miozzo & Caramazza, 1999; Sá Leite et al., 2020). Specifically, nouns ending in *-o* often have masculine gender (99.9%) and those ending in *-a* generally have feminine gender (96.3%). Additionally, a small number of nouns, i.e., those ending in *-e*, have either feminine or masculine gender, e.g., *el tigre* "the$_{\text{MASC}}$ tiger$_{\text{MASC}}$" and *la llave* "the$_{\text{FEM}}$ key$_{\text{FEM}}$", with 89.4% of all *-e* words being masculine (Grüter et al., 2012; O'Rourke, 2007; Teschner & Russell, 1984). Similarly, a small number of nouns ending in consonants (e.g., *-z*, *-l*, *-s*, etc.) are opaque (O'Rourke, 2007). In general, there are about twice as many transparent nouns as there are opaque nouns in Spanish (Harris, 1991).

According to these transparent endings, the selection of determiners can mainly rely on the morphological feature of nouns, in which the corresponding feminine determiners (e.g., *la* "the$_{\text{SG}}$" and *las* "the$_{\text{PL}}$") are assigned to nouns ending in *-a* (e.g., *la guitarra* "the$_{\text{FEM}}$ guitar$_{\text{FEM}}$"). Similarly, the masculine determiners (e.g., *el* "the$_{\text{SG}}$" and *los* "the$_{\text{PL}}$") are involved in nouns ending in *-o* (e.g., *el gato* "the$_{\text{MASC}}$ cat$_{\text{MASC}}$"). However, there are less than 0.5% exceptions to this transparent gender marking of nouns, including words where the correspondence between their gender and their ending is not transparent (Costa et al., 1999b). For instance, the feminine determiners cannot be assigned to nouns beginning with a stressed /a/ (e.g., *el agua* "the$_{\text{MASC}}$ water$_{\text{MASC}}$") (Costa et al., 1999b). This means that only when the phonological information about the nouns is available can the correct form of the determiner be selected. Whether or not the contradictory findings of the gender congruency effect in Spanish are due to the fact that the gender-marking determiners are not exclusively dependent on the grammatical gender of the head noun, but also on its onset phonology, invites more debate.

2.1.3 The current study: noun phrase production in Spanish

This study presents behavioral evidence from a determiner-noun phrase production task using a picture-word interference (PWI) paradigm (Glaser, 1992; Rosinski et al., 1975) to further explore the gender congruency effect in Spanish. We set gender congruency (i.e., gender-congruent vs. gender-incongruent) and semantic relatedness (i.e., semantically related vs. semantically unrelated) as the two main factors. The goal of the present study is to investigate whether or not the grammatical gender of determiners is competitively selected in the production of noun phrases by Spanish speakers in a well-controlled experiment with new targets and pictures. If grammatical gender is selected competitively, the next question is whether or not this competition is reflected in the effect of reaction times. Therefore, we addressed the main research question: is there a gender congruency effect in Spanish NP production?

Hypotheses

The lexical selection by competition theory (Levelt et al., 1999; Roelofs, 2003) refers to the process of competitively selecting the target word from all the activated non-target words when producing a word. Speakers will take more time to select the target word in their language production when more non-target words are highly activated. Based on this theory, we expect the effects of semantic relatedness on naming latencies to be present in the picture naming task. That is, a slower reaction time is expected when the target picture and the distractor belong to the same semantic category (Glaser & Düngelhoff, 1984; Schriefers et al., 1990). On the contrary, we predict faster naming latencies in semantically unrelated conditions, in which the target object belongs to a different semantic category than the distractor. For the gender congruency condition, we predict a significant gender congruency effect on naming latencies. Specifically, we predict that the gender-congruent condition shortens speakers' reaction times for naming target objects. In contrast, speakers' naming latencies are prolonged due to the gender difference between the target and the distractor.

Levelt et al.'s (1999) model of speech production claims that the content of the target picture is conceptualized first at the conceptual level, and then the grammatical gender as a syntactic feature is activated and selected at the lemma level, both occurring in a sequential order in the picture naming task. Since the selection of the grammatical gender of determiners in NPs is dominated by properties of the noun in Spanish (Costa et al., 1999b), the process of grammatical gender is supposed to be independent of semantic relatedness in word production. As a result, an interaction between the gender congruency effect and the semantic interference is not expected.

2.2 Materials and methods

2.2.1 Participants

Thirty healthy, right-handed Spanish speakers ($Mean_{age}$ = 26.08 years and SD_{age} = 4.85 years; nineteen females) participated in this experiment. Participants did not report any history of neurological or language disorders. Before testing, informed consent was obtained from all participants. They read an information sheet and signed a consent form, which was approved by the Ethics Committee of the Faculty of Humanities at

Leiden University. Upon termination of all tasks, participants were paid for their participation.

2.2.2 Materials

Twenty black-and-white line pictures were obtained from Severens' picture database (Severens et al., 2005) based on two criteria: pictures had to refer to a familiar and concrete concept, and pictures had to have easily recognizable features. Each picture was assigned four distractors, which were manipulated in four conditions according to their gender congruency (i.e., gender-congruent vs. gender-incongruent) and semantic relatedness (i.e., semantically related vs. semantically unrelated) to the target picture (see Appendix 2.A). As a result, a total of 80 combinations of target pictures and distractor pairs were generated. The word frequency of the distractors was controlled on the basis of Corpus del Español (Davies, 2016) across the four conditions, with $F(3, 76) = 1.358$ and $p = 0.262$. Similarly, the visual complexity of the distractors was controlled by balancing the number of letters across the four conditions, with $F(3, 76) = 1.925$ and $p = 0.133$. Targets and distractors were neither phonologically nor orthographically related.

2.2.3 Design and procedure

The experiment was designed as a 2 by 2 fully factorial within-subjects design with two main factors: gender congruency (G) and semantic relatedness (S). The factor gender congruency included two levels, i.e., gender-congruent (G+) and gender-incongruent (G−), based on gender congruency or gender incongruency between the target picture and the distractors. The factor semantic relatedness was divided into two levels, i.e., semantically related (S+) or semantically unrelated (S−), depending on whether or not the target pictures and the distractors belonged to the same semantic category. As a result, four conditions were generated for each target picture: G+S+, G+S−, G−S+, and G−S− (see Table 2.2.1).

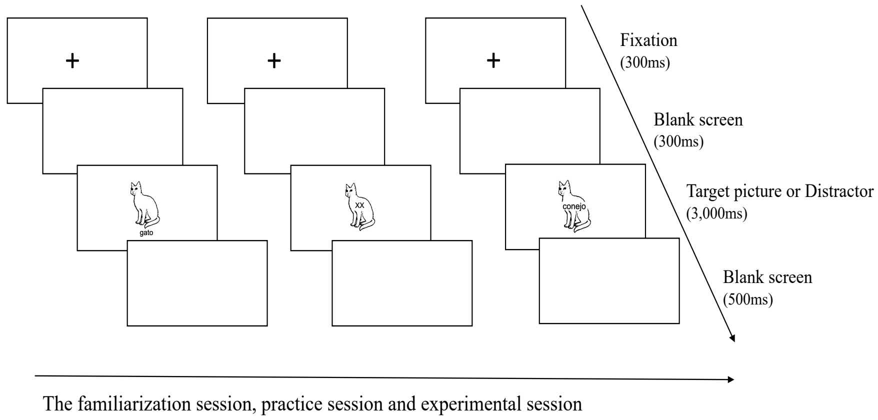
Table 2.2.1. Sample of stimuli in the experimental session for the PWI task.

	Condition			
Target picture **PIG** [CERDO]	G+S+	G+S−	G−S+	G−S−
Grammatical gender el (m.)				
				
	dog	ice	cow	wood
Distractors	perro	hielo	vaca	madera
Grammatical gender of distractors	el (m.)	el (m.)	la (f.)	la (f.)

The picture naming task was programmed in E-prime 2.0 (Schneider et al., 2002) and designed based on the PWI paradigm. In order to counterbalance the effect of the order, we used the Windows program Mix (Van Casteren & Davis, 2006) to generate a pseudo-random order of trials according to two criteria: no two trials within the same condition or associated with the same target picture were allowed to appear consecutively, and trials of the same grammatical gender could be presented no more than twice in a row. As a result, 40 blocks were generated in E-prime 2.0. Furthermore, we also used the by-subjects order design for the programming of the task, so that the order of the blocks was randomized across participants.

The whole experiment was divided into three sessions: a familiarization session, a practice session, and an experimental session, lasting 20 minutes in total. Participants were first presented with the familiarization session, in which they were instructed to learn the exact name under the target picture for 3,000 ms. After the presentation of all 20 target pictures, participants were asked to practice naming the same pictures in a practice session, in which the target pictures were presented with a meaningless "XX" string in the center of the screen. In this session, each picture was presented for 3,000 ms. The correct name of the target was provided if participants produced an incorrect name. In the experimental session, participants were expected to name the target picture as fast and accurately as possible with a Spanish noun phrase (e.g., *el gato* "the$_{\text{MASC}}$ cat$_{\text{MASC}}$") while ignoring the distractor word. For each trial, the typical procedure began with a fixation cross presented in the center of the screen for 300 ms, followed by a blank screen for 300 ms. This was followed by the display of the target picture and distractor word for 3,000 ms, during which participants' vocal responses were recorded. At the end of each trial, a blank screen was displayed for 500 ms (see Figure 2.2.1).

Figure 2.2.1. Illustration of the familiarization session, practice session, and experimental session for the picture naming task.



2.3 Results

Naming latencies were calculated and extracted from all recorded data of 30 participants using Praat (Boersma & Weenink, 2019) (see Table

2.3.1). Of all the 2,400 recorded data points, 4.6% were removed from further behavioral data analysis due to the presence of (a) incorrect responses, no responses, or delayed responses (3.3%) and (b) outliers, i.e., naming latencies exceeding 3 SDs around the average responding time of participants (1.3%). Next, we employed the generalized linear mixed model (GLMM) and the *glmer* function with a *gamma* distribution to analyze our behavioral data in R Studio version 4.2.2 (R Core Team, 2020) using the *lme4* package (Bates et al., 2015).

Table 2.3.1. Mean picture naming latencies by condition.

Condition	**Naming latencies (ms)**	
	Mean	**SD**
Gender-congruent / Semantically Related (G+S+)	714	194
Gender-congruent / Semantically Unrelated (G+S−)	673	149
Gender-incongruent / Semantically Related (G−S+)	725	184
Gender-incongruent / Semantically Unrelated (G−S−)	691	165

To avoid the risk of increasing the Type I error rate, the analysis of the behavioral data was modeled using a top-down model selection procedure (Barr, 2013), in which the model starts with the theoretically maximal model. In order to fit the model to our data, we included gender congruency and semantic relatedness as two fixed factors, and item and participant as two random factors. To control for potential confounders, we added a distractor category as a covariate to our statistical analysis. Further, we included a target picture category as a random slope for the item factor. We generated the best-fit model for our data by taking the following steps: first, we removed the non-significant random effect of interaction between gender congruency and semantic relatedness for each participant, and the interaction of the fixed effects of gender congruency and semantic relatedness in the case of a singular fit; second, the random slopes of the target category for the items, and the distractor category were removed as they did not significantly improve the model fit and resulted in non-convergence; third, the correlation between gender congruency and semantic relatedness for the participant factor, and the random intercept and slope of gender congruency for the participant factor were excluded on the basis of akaike's information criterion (AIC) (Akaike, 1974), the Bayesian information criterion (BIC) (Neath & Cavanaugh, 2012), and the log-likelihood ratio; finally, the best-fit model for our data was fitted with gender congruency and semantic relatedness as fixed effects, and random intercepts for the participant and target item, as well as a by-participant random slope for semantic relatedness. The best-fit model (see Table 2.3.2) demonstrated significantly shorter naming latencies in gender-congruent conditions compared to gender-incongruent conditions with $\beta = -13.526$, $SE = 4.64$, $t = -2.91$, and $p = 0.004$. Moreover, participants responded significantly slower to semantically related trials than to semantically unrelated trials, with $\beta = 38.41$, $SE = 10.75$, $t = 3.57$, and $p < 0.001$ (see Figure 2.3.1).

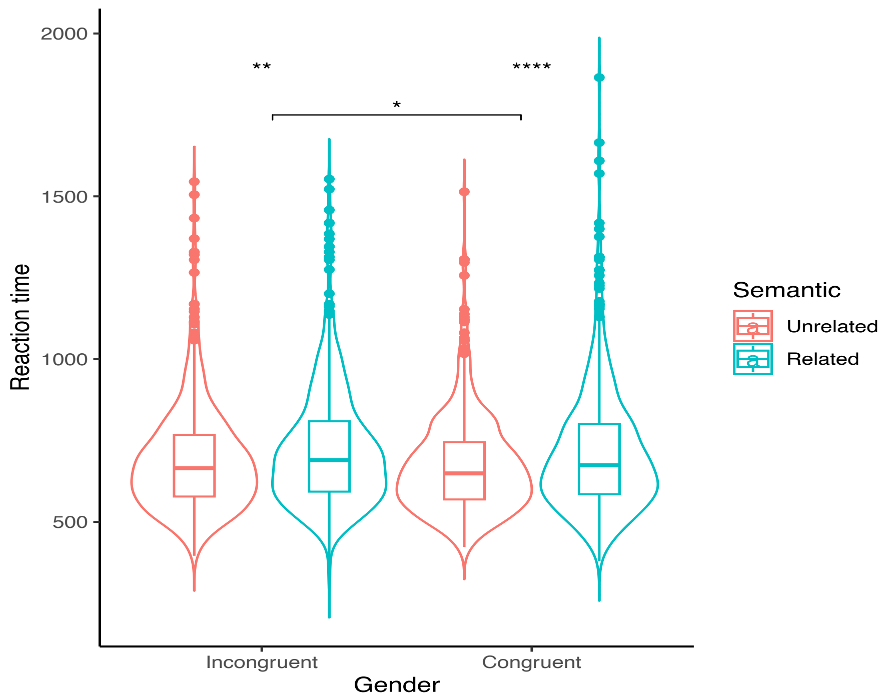
Table 2.3.2. General mixed effects model of best fit for RTs with gender congruency and semantic relatedness as two predictors, including estimates, confidence intervals and *p*-values. The result demonstrated that both gender congruency and semantic relatedness had a significant impact on naming latencies. "***" indicates $p < 0.001$ and "**" indicates $p < 0.01$.

Formula: naming latency ~ gender congruency (congruent vs. incongruent) + semantic relatedness (related vs. unrelated) + (semantic relatedness | participant) + (1 | item)

Predictors	**RTs**			
	Estimate	*95% CI*	*Statistic*	*Pr(>|z|)*
(Intercept)	714.882	683.193 – 746.570	44.239	<0.001***
Gender [Congruent]	− 13.526	− 22.634 – − 4.419	− 2.912	0.004**
Semantic [Related]	38.409	17.320 – 59.498	3.571	<0.001***
Random Effects				
σ^2	0.04			
$\tau_{00\ Participant}$	2460.61			
$\tau_{00\ Item}$	578.38			

τ_{11} participant. SemanticRelated	872.19
ρ_{01} participant	− 0.20
ICC	1.00
N participant	30
N Item	20
Observations	2288
Marginal R^2/Conditional R^2	0.115/1.000

Figure 2.3.1. Picture naming latencies by condition. Both gender congruency and semantic relatedness showed a significant effect on naming latencies (i.e., $p < 0.0001$, indicated as "****"; $p < 0.01$, indicated as "**" and $p < 0.05$, indicated as "*" respectively in the figure).



2.4 Discussion

In this study, we investigated the potential impact of gender congruency and semantic relatedness on Spanish NP production using the picture-word interference paradigm. The gender congruency and semantic relatedness were manipulated between the target picture and the distractors. The gender congruency effect was examined by comparing the naming latencies between gender-matched and gender-unmatched determiner-NPs. We predicted shorter naming latencies for gender-congruent NPs compared to incongruent NPs. Critically, this would indicate that the target gender value was activated and selected at the lemma level. We also studied the semantic effect with semantically related or unrelated NPs, and we expected longer response times for unrelated trials compared to related trials. The presence of such a semantic interference effect would imply that the concept of the target picture and the distractor actively competed for selection during *lemma level* processing.

In line with our expectations, we found that participants' naming latencies were significantly affected by the semantic relatedness of a distractor word and the target picture. To be more precise, the participants showed longer naming latencies when naming a target picture with a distractor of the same semantic category. In other words, the semantic interference effect (Glaser, 1992; La Heij, 1988; Schriefers et al., 1990) was found in naming semantically related targets and distractors. Our behavioral results are important because the presence of the semantic interference effect is suggestive of the lexical selection process during semantic processing. It implies that the semantically related distractors represent competitors of the target, resulting in a more difficult selection process than in unrelated conditions. When the target (e.g., CHAIR) is presented with a semantically related distractor (e.g., table), more than one lemma (e.g., chair, table, bed, stool) of the target (via activating its semantic category) is activated, and the distractor only activates one lemma. In the semantically related conditions, one of the competitors of the target (e.g., chair) receives additional activation from the distractor (e.g., table), leading to more competition during the lexical selection process. These behavioral findings are in line with numerous previous studies that report semantic interference (Belke et al., 2005; La Heij, 1988; Levelt et al., 1999; Roelofs, 2003).

Meanwhile, we explored the effect of gender congruency on access to the target word in the presence of a distractor by manipulating the gender congruency between the target and the distractors. A significant gender congruency effect was observed in the naming latencies of the NP naming task. Our behavioral results suggest that participants were significantly faster at naming the target with gender-congruent distractors than with gender-incongruent distractors. These findings are important for three reasons: first, the significantly faster naming latencies in the gender-congruent conditions suggest the presence of a gender congruency effect during gender processing; second, the gender congruency effect implies that the competition occurred during the process of gender feature selection, in which the different gender values for the target picture and the distractor were activated in the gender-incongruent condition and the gender feature of distractors competed for selection at the lemma level and interfered with the naming of the target NP; third, the statistically significant longer reaction time in gender-incongruent conditions also suggests that the competition can be reflected by a reaction time effect in Spanish NP production.

It is important to mention that no significant gender congruency effect was found in previous studies on Spanish NP naming (e.g., Costa et al., 1999b) and some of the bare noun naming studies (e.g., O'Rourke, 2007 and Finocchiaro et al., 2011). However, the results of the gender congruency effect are reversed for the study of Paolieri et al. (2010), where longer naming latencies were found for gender-matched target-distractor pairs than for gender-non-matched pairs. Compared to all aspects of the previous studies, the gender congruency effect was possibly found in the current study for the following reasons.

First of all, in the study by Paolieri et al. (2010), twenty native Spanish participants were asked to name 28 target pictures using bare nouns in Spanish, and a reliable reversed gender congruency effect was obtained in their bare noun naming tasks. They concluded that their findings confirm the assumption that grammatical gender is not only a syntactic feature, but also an inherent lexical property of nouns that can be selected in bare noun production. As is consistent with this assumption, in our experiment, participants were asked to perform a noun phrase naming task, in which a determiner-noun phrase had to be produced. In determiner-noun phrase production, the correct determiner-noun phrase can only be produced by selecting the grammatical gender of nouns and determiners. Therefore, it was essential for participants to explicitly produce the correct determiner when given the target gender value. In this process, there was competition in the selection of the syntactic feature of the target.

However, the gender congruency effect in the study by Paolieri et al. (2010) was reversed, with the longer naming latencies being found in the gender-congruent condition. This finding is unexpected within the current language production models, e.g., the LRM model (Levelt et al., 1999) and the WEAVER++ model (Levelt, 1992). More specifically, according to the LRM model (Levelt et al., 1999), when the target picture and a distractor are presented to participants, grammatical gender as an inherent lexical property of nouns is activated. In a gender-incongruent condition, different gender nodes are activated for the distractor and the target, and the gender feature of the distractor interferes with the naming of the target. In this case, the competition occurs when the target syntactic feature is selected with the interference from the distractor. In contrast, in a gender-congruent condition, the same gender node is activated for the distractor and the target, where no competition occurs. As a result, it takes more time to produce an NP with the correct determiner in the gender-incongruent condition than in the gender-congruent condition, as the target gender node competes for selection with non-target activated gender nodes in the gender-incongruent condition. In this case, the effect on naming latencies of gender congruency is easy to observe. Our results are consistent with the results of gender processing in current models.

Second, it is necessary to pay attention to the study by Costa et al. (1999b), who conducted three experiments asking participants to perform three Spanish NP (i.e., determiner + noun) naming tasks and found no gender congruency effect. They concluded that the gender congruency effects may have occurred during early lexical selection processes but were rendered invisible by the selection and retrieval of the phonological form of a word. However, this explanation is rather unlikely for Spanish, since only a rather small number of nouns, i.e., less than 50 cases of all lexical items (less than 0.5%) (*Real Academia Española: Diccionario de La Lengua Española*, 2014), are consistent with this conclusion. In general, noun endings are either transparent (i.e., *-o* for masculine and *-a* for feminine) or opaque (i.e., *-e* for either masculine or feminine, or consonants, e.g., *-z*, *-l*, *-s*, etc.) (O'Rourke, 2007). In many cases, the correct choice of determiner in Spanish can be made on the basis of the transparent property (i.e., the phonological information) of the nouns. Only in a small proportion of cases (less than 0.5%) of all lexical items does the selection of determiners depend on the phonological context (Costa et al., 1999b). Given these facts, the conclusion of Costa et al. (1999b) that there is no gender congruency effect in Spanish NP production becomes less likely. On the other hand, the conclusion of Costa et al. (1999b) can also be interpreted in the following way: it is possible for the gender congruency effect to be found without selecting and retrieving the phonological form of a word. In our list of stimuli, there were no exceptions included (i.e., masculine determiner for nouns starting with a stressed /a/) for both targets and distractors, which meant that the selection of determiners could be based on the lexical information of the nouns. In this case, the finding of a gender congruency effect in our study seems plausible.

Third, compared to the study by O'Rourke (2007), in which the distractors were presented in auditory form, we adopted a different experimental procedure with the presentation of the distractor being in printed form. Such a difference in display form has been shown to influence the timing of distractor effects (Damian & Martin, 1999; Finocchiaro et al., 2011). Given this difference, the discrepancy between O'Rourke (2007) and the current study does not necessarily imply contradictory conclusions. Additionally, in terms of the structure of the grammatical gender of nouns, Spanish is similar to Dutch in that phonological information is hardly necessary to choose the correct determiners (Costa et al., 1999b). The few exceptional cases may be hard wired, i.e., learned and stored as phrases, e.g., *el agua*, "the water", or *el agua fria*, "the cold water". In this case, the findings of the study by O'Rourke (2007) and Finocchiaro et al. (2011) may be explained by the Dutch case, where the gender congruency effect was not found in bare noun production but was found in noun phrase production (La Heij et al., 1998). This process of producing NPs can be interpreted as an amplification of the potential gender congruency effect that is not detected in the more straightforward bare noun task, which does not require the explicit selection of the grammatical gender for determiner selection.

Additionally, it is worth mentioning the statistical analysis we applied to our data, which supports the robustness of the effects we observed in the present study. First, unlike the previous studies (e.g., Cubelli et al., 2005), which defined outliers as a naming latency exceeding two standard deviations below or above the average responding time of participants (i.e., mean +/−2 SDs), we defined outliers by removing the naming latencies that lay three standard deviations around participants' mean response times (i.e., mean +/−3 SDs). This resulted in a few data points being identified as outliers and excluded from further data analysis. In general, the method of using mean +/−2 SDs is considered a restrictive method that excludes outliers, resulting in the loss of more reliable data (Finocchiaro et al., 2011; Berger & Kiefer, 2021). Second, the implementation of a single-trial generalized linear mixed-effect model (GLMM) was a novelty of our statistical analysis. Compared to the traditional statistical analysis of naming latencies using ANOVAs (e.g., Paolieri et al., 2010; Finocchiaro et al., 2011), the single-trial GLMM is more sensitive and specific to single-trial analysis, and does not assume an underlying distribution or an equal number of observations for each participant or condition. Instead, it captures variance as explained by each participant and by each item, as well as the experimental manipulations of interest (Linde-Domingo et al., 2019). When modeling behavioral data with unequal numbers of conditions or participants, it therefore has superior explanatory power to traditional ANOVAs. In other words, it also provided a more robust result for our behavioral data analysis.

2.5 Conclusion

In this study, we explored whether or not grammatical gender, as a lexico-syntactic feature in Spanish, is used to competitively select determiners in the production of noun phrases by Spanish speakers in a well-controlled experiment. We employed the picture-word interference paradigm to examine the naming latencies of 30 participants for multiple objects in four conditions in which gender congruency, i.e., gender-congruent and gender-incongruent, and semantic relatedness, i.e., semantically related and semantically unrelated, were manipulated. We found reliable effects of grammatical gender congruency and semantic relatedness in noun phrase production in Spanish. More specifically, statistically significant shorter naming latencies were found for gender-congruent target-distractor pairs than for gender-incongruent pairs. Similarly, we found longer naming latencies in semantically related trials than in unrelated trials. These results provide crucial evidence supporting the notion that grammatical gender in determiner-NPs is competitively selected and that this competition is reflected in speakers' naming latencies. Furthermore, our findings provide an important behavioral piece of evidence for the gender congruency effect in Romance languages.

Author Contributions Statement

Ruixue Wu: Conceptualization, methodology, software, formal analysis, data curation, writing—original draft preparation, writing—review and editing. **Niels O. Schiller**: Conceptualization, methodology, writing—review and editing, supervision, funding acquisition. All authors have read and agreed to the published version of the manuscript.

Acknowledgements

We are grateful to all members of the LUCL Experimental Linguistics Lab, particularly Sarah von Grebmer zu Wolfsthurn, for providing advice on statistical analyses and programming. We also thank the Statistical Consultation Service of Leiden University for advice on statistical analyses. Finally, we would like to thank our anonymous reviewers and all our participants for their contributions. This work was partly funded by the China Scholarship Council (CSC) No. 202107720015.

Data Availability Statement

The data supporting the findings of this study are openly available in Open Science Framework (OSF) at https://osf.io/9vjdp/?view_only=cd0afe62286b4a4f9a648ff6b307bb32 (accessed on 5 April 2023) (view-only link).

Appendix

Appendix 2.A Table of stimuli used in the picture naming task.

Target Picture	Gender	Distractor Type			
		G+S+	**G+S−**	**G−S+**	**G−S−**
gato (cat)	Masculine	conejo (rabbit)	sol (sun)	rata (rat)	miel (honey)
caballo (horse)	Masculine	burro (donkey)	sello (stamp)	mula (mule)	flor (flower)
cuervo (crow)	Masculine	papagayo (parrot)	monte (mountain)	paloma (dove)	toalla (towel)
guitarra (guitar)	Feminine	trompeta (trumpet)	isla (island)	violin (violin)	zapato (shoe)
mesa (table)	Feminine	silla (chair)	araña (spider)	armario (closet)	gorro (hat)
estantería (shelving)	Feminine	mecedora (rocking chair)	bandera (flag)	sofá (sofa)	cabello (hair)
cama (bed)	Feminine	alacena (cupboard)	garra (claw)	sillón (armchair)	libro (book)
cerdo (pig)	Masculine	perro (dog)	hielo (ice)	vaca (cow)	madera (wood)
tigre (tiger)	Masculine	león (lion)	avión (airplane)	serpiente (snake)	oreja (ear)
brazo (arm)	Masculine	pie (foot)	pan (bread)	pierna (leg)	arena (sand)
boca (mouth)	Feminine	nariz (nose)	abeja (bee)	ojo (eye)	árbol (tree)
tenedor (fork)	Masculine	cuchillo (knife)	bolso (handbag)	cuchara (spoon)	pintura (painting)
tomate (tomato)	Masculine	chile (chili)	diente (tooth)	patata (potato)	piedra (stone)
abrigo (coat)	Masculine	jersey (sweater)	huevo (egg)	camiseta (shirt)	estrella (star)
manzana (apple)	Feminine	pera (pear)	bota (boot)	plátano (banana)	pescado (fish)
autobús (bus)	Masculine	tren (train)	mono (monkey)	bicicleta (bike)	llave (key)
dedo (finger)	Masculine	codo (elbow)	ratón (mouse)	palma (palm)	nave (ship)
zorro	Masculine	lobo	papel	hiena	lengua

(fox)		(wolf)	(paper)	(hyena)	(tongue)
flecha	Feminine	pistola	roca	puñal	oso
(arrow)		(gun)	(rock)	(dagger)	(bear)
maíz	Masculine	trigo	gallo	soja	carta
(corn)		(wheat)	(rooster)	(soy)	(letter)

Chapter 3

Variation in Gender Assignment Strategies in Mixed Spanish−Chinese Noun Phrases: Insights from a Multilingual Community in Barcelona

---

A version of this chapter has been published in *International Journal of Bilingualism* as: Wu, R., Parafita Couto, M. C., & Schiller, N. O. (2025). Variation in gender assignment strategies in mixed Spanish−Chinese noun phrases: insights from a multilingual community in Barcelona. *International Journal of Bilingualism*, published online. https://doi.org/10.1177/13670069251363811

60 *The Role of Lexico-Syntactic Features in Noun Phrase Production and Comprehension*

Abstract

This study investigates gender assignment strategies in mixed Spanish−Chinese noun phrases (NPs) among early Spanish−Chinese bilinguals in Barcelona, Spain. It explores whether bilinguals rely on a default masculine or use the gender of Spanish translation equivalents across production and comprehension tasks. A multi-task approach was employed, including a director-matcher task, a sentence repetition task, and a two-alternative forced-choice judgment task. Forty-two early sequential Spanish−Chinese bilinguals participated in the study. The analysis focused on the use of masculine and feminine Spanish determiners with Chinese nouns, categorized according to the gender of their Spanish translation equivalents. One-way repeated ANOVA was conducted to examine whether gender choices significantly differed when Chinese nouns, either with feminine or masculine Spanish equivalents, were paired with different Spanish determiners in each task. Paired-samples t-tests and non-parametric Wilcoxon signed-rank tests were conducted to determine whether masculine gender is significantly preferred over feminine gender in the three tasks. Results revealed task-dependent variability in gender assignment. In production, bilinguals used both masculine and feminine determiners for Chinese nouns with feminine equivalents, reflecting both default masculine and translation equivalent strategies. For nouns with masculine equivalents, masculine determiners predominated. In comprehension, participants favored the translation equivalent strategy, using determiners that matched the gender of Spanish translation equivalents. This study is the first to examine gender assignment in mixed Spanish−Chinese NPs. The study highlights the flexible nature of bilingual language processing, demonstrating that gender assignment strategies can change in response to different task demands. It emphasizes the importance of capturing the range of bilingual behaviors in both production and comprehension. However, the study focuses on a specific bilingual community, limiting generalizability. Future studies should examine gender assignment over time and in other Spanish−Chinese bilingual populations with different sociolinguistic backgrounds.

Keywords: code-switching; mixed noun phrases; grammatical gender; gender assignment strategies; Spanish; Chinese; bilingualism

3.1 Introduction

Code-switching (CS) is commonly described as the practice of switching back and forth between languages within an utterance (Poplack, 1980; Deuchar, 2012), and CS data can provide valuable insights that extend our understanding of language beyond monolingual perspectives (López, 2020). In the study of CS, researchers have examined grammatical gender assignment in mixed noun phrases (NPs), where a determiner from one language is paired with a noun from another (cf. Bellamy & Parafita Couto, 2022). This interest arises from frequent occurrences of mixed NPs (Jake et al., 2002) and differences in how languages mark gender. A mixed NP typically consists of at least two elements: a gender-marked determiner from one language and a noun, either gender-marked or non-gender-marked, from another (e.g., *la house* "the$_{\text{FEM}}$ house", Spanish−English; *el gürtel* "the$_{\text{MASC}}$ belt$_{\text{MASC}}$", Spanish−German in López, 2020), or a non-gender-marked determiner from one language with a gender-marked noun from another (e.g., *the casa* "the house$_{\text{FEM}}$", English−Spanish in Blokzijl et al., 2017). In gender-marked languages such as Romance and Germanic languages, grammatical gender is morphosyntactically represented by the agreement between gender-marked nouns and other elements (determiner, adjective) in the NPs (e.g., Spanish: *la camisa blanca* "the$_{\text{FEM}}$ shirt$_{\text{FEM}}$ white$_{\text{FEM}}$") (Corbett, 2012). In contrast, languages such as Chinese and English have no grammatical gender distinction for nouns.

To set the stage for this investigation, we first outline the structure of the NP in Spanish and Chinese, followed by an examination of the distinctive characteristics that emerge in mixed Spanish−Chinese NPs.

3.1.1 Spanish NPs

In Spanish, NPs typically consist of a determiner followed by a noun, which may or may not be accompanied by an adjective (determiner + noun +/− adjective). All nouns in Spanish have an inherent grammatical gender, either masculine or feminine (Roca, 1989; Harris, 1991). Determiners and adjectives do not have inherent gender themselves, but they must agree in gender with nouns they modify (Harris, 1991; see examples 1a and 1b below).

(1) a. **la** camis**a** roj**a**
the.FEM coat.FEM red.FEM
"the red coat"

b. **el** huev**o** rot**o**
the.MASC egg.MASC broken.MASC
"the broken egg"

Agreement in Spanish is easily observed because nouns typically feature morphologically or phonologically transparent suffixes, although this is not always the case. Specifically, nouns ending in *-o* are predominantly masculine (99.9%), while those ending in *-a* are often feminine (96.3%) (Teschner & Russell, 1984). A subset of nouns ending in *-e* can be either gender (see 2a, 2b), and those ending in consonants (e.g., *-l*, *-n*, *-z*) are opaque and less predictable. Spanish has roughly twice as many transparent nouns as opaque ones (Harris, 1991), making endings reliable predictors of grammatical gender.

(2) a. **el** pi**e**
the.MASC foot.MASC
"the foot"

b. **la** clas**e**
the.FEM classroom.FEM
"the classroom"

Masculine gender is considered the unmarked default, while feminine is marked and exclusive in Spanish (Roca, 1989; Harris, 1991). Spanish speakers often assign masculine gender to unknown nouns or those with irregular endings, like consonants, partly because masculine nouns generally allow more varied phonological endings (Beatty-Martínez & Dussias, 2019).

3.1.2 Chinese NPs

Unlike Spanish, Chinese lacks grammatical gender and employs a nominal classifier system. In Mandarin Chinese, NPs may contain a demonstrative, numeral, or quantifier, followed by a classifier and a noun (e.g., *一张桌子*, /yi1zhang1zhuo1zi0[3]/ [one + classifier-zhang1+ table], "one table"), in which not all are required but classifiers are obligatory and cannot be omitted (Li & Thompson, 1981). In Chinese NPs, the choice of classifiers is largely based on the semantic properties of nouns, such as animacy, shape, function, and size (Myers & Tsay, 2000). Unlike Spanish, which uses an alphabetic writing system based on the Latin script, where each letter generally corresponds to a specific sound, Mandarin Chinese employs a logographic writing system, where each character typically represents a syllable and a unit of meaning, rather than a single sound (Wang, 1973). Chinese characters represent morphemes and correspond to Pinyin syllables, which reflect the language's phonology (Wang, 1973). A Pinyin syllable consists of an initial consonant and a final vowel or an optional nasal consonant (e.g., /n/ or /ng/) (Li & Thompson, 1981). Final vowels are usually compound (e.g., /ai/, /ei/, /ao/) but sometimes simple (e.g., /a/, /o/, /e/). In writing, Chinese is typically presented using only characters without accompanying Pinyin, which marks a significant distinction from alphabetic Spanish.

3.1.3 Mixed Spanish−Chinese NPs

Mixed Spanish−Chinese NPs contain at least two elements, a gender-marked Spanish determiner and a non-gender-marked Chinese noun (e.g., *el/un 太阳*, "the$_{\text{MASC}}$/a$_{\text{MASC}}$ sun"). The question that arises is: what gramm-

[3] Mandarin Chinese Pinyin features four phonemic tones: Tone 1 (high-level), Tone 2 (high-rising), Tone 3 (low-dipping), and Tone 4 (high-falling) (Chao, 1948). In this paper, tone numbers indicate the specific tones in Pinyin.

atical gender will Spanish−Chinese bilinguals assign to Chinese nouns in mixed NPs? This issue is interesting given that Spanish requires gender agreement for determiners and adjectives, while Chinese lacks grammatical gender altogether. Research on other mixed-language contexts, such as Spanish−English, has identified several strategies for gender assignment, including using the gender of the Spanish translation equivalent, relying on the morphological shape of the noun, or defaulting to masculine gender (see Section 3.2). However, it remains to be seen which strategy Spanish−Chinese bilinguals will prefer, given the linguistic differences between Spanish and Chinese. Understanding the structure of mixed Spanish−Chinese NPs sets the stage for exploring how bilinguals assign grammatical gender in such contexts. The following section reviews relevant research on gender assignment strategies in mixed NPs across different language pairs.

3.2 Research background

3.2.1 Gender assignment strategies in mixed NPs

In the context of code-switching, three primary gender assignment strategies have been documented in mixed NPs across various language combinations, namely the translation equivalent strategy, the default strategy, and the shape-based strategy (see Bellamy & Parafita Couto, 2022). These strategies are observed in two types of language combinations: those where only one language has grammatical gender, and those where both languages have gender. Spanish−Chinese mixed NPs fall into the first category, as Spanish has grammatical gender while Chinese does not. Building on this background, we now turn to the specific strategies that have been documented for gender assignment in mixed NPs, with particular attention to language pairs where only one language marks grammatical gender, as is the case for Spanish−Chinese.

3.2.1.1 Translation equivalent strategy

The translation equivalent strategy, also referred to as the analogical gender criterion (Montes-Alcalá & Lapidus Shin, 2011; Jake et al., 2002), involves assigning gender to the inserted noun based on the gender of its translation equivalent in the recipient language (Bellamy & Parafita Couto, 2022). For example, in Spanish−English mixed NPs, bilinguals have been observed to pair a Spanish feminine determiner with an English noun if its Spanish translation is feminine (Liceras et al., 2008). This strategy has also been reported in other bilingual contexts, such as Tsova-Tush−Georgian (Bellamy & Wichers Schreur, 2022) and Spanish−Basque (Munarriz-Ibarrola et al., 2022), particularly among speakers who learned the gender-marked language earlier. In the Spanish−Chinese context, this strategy would involve assigning the gender of the Spanish translation equivalent to the Chinese noun (e.g., *la 吉他*, "the$_{FEM}$ guitar", following the feminine gender of *guitarra* in Spanish).

3.2.1.2 Default strategy

Regarding the default strategy, speakers tend to default to one gender for most inserted nouns, regardless of their semantic and morpho-phonological properties or translation equivalents (Bellamy & Parafita Couto, 2022). This approach is prevalent in Spanish−English bilingual speech, particularly in communities with frequent code-switching (e.g., Balam, 2016; see Bellamy & Parafita Couto, 2022 for an overview). For instance, Spanish−English bilinguals in Belize assigned masculine gender to 99.6% of English nouns in mixed DPs (Balam, 2016). Likewise, Balam et al. (2021) reported a strong preference for masculine gender among Spanish−English bilingual children in the Miami corpus, with Valdés Kroff (2016) noting 93.7% of English nouns paired with Spanish masculine determiners in the Bangor Miami corpus, independently of their translation equivalent, in the same community. Additionally, Spanish−English bilinguals with habitual CS practices in Southern Arizona (USA) favored masculine gender for English inanimate nouns in a forced-choice elicitation task (Cruz, 2023) and the Corpus del Español en el Sur de Arizona (CESA) corpus (Cruz, 2021).

In addition to the masculine default, Parafita Couto et al. (2016) reported a preference for the feminine default among Spanish−Basque bilinguals in an auditory judgment task. To sum up, the default strategy is particularly prevalent among frequent code-switchers and simultaneous bilinguals (Bellamy & Parafita Couto, 2022).

3.2.1.3 Shape-based strategy

The shape-based strategy, also referred to as "phonological gender criterion" (Poplack et al.,1982) or "phonological strategy" (Munarriz-Ibarrola et al., 2022), involves assigning gender based on morpho-phonological properties of the inserted noun, such as its final sound or letter. This strategy is more common in language pairs with similar alphabetic systems and overlapping morpho-phonological patterns, as observed among Spanish−Basque and Purepecha−Spanish bilinguals (Parafita Couto et al., 2016; Bellamy et al., 2018). Bilinguals, such as Purepecha−Spanish or Spanish−Basque, tend to associate the phonological shape or morpho-phonological cues from words in the donor language with gender markers in the recipient language when assigning gender to donor language nouns (cf. Bellamy & Parafita Couto, 2022). For instance, Spanish−Basque bilinguals prefer to pair the Spanish feminine determiner *la* with Basque nouns ending in *-a* (Basque lacks gender), such as *la arrautza* "the$_{\text{FEM}}$ egg", where *huevo* "egg$_{\text{MASC}}$" is masculine in Spanish (Parafita Couto et al., 2016). Similarly, Purepecha−Spanish bilinguals preferred matching Purepecha nouns ending in *-a* (Purepecha lacks gender) with the feminine determiner *la* in a forced-choice acceptability judgment task, even when their Spanish equivalents were masculine (Bellamy et al., 2018).

Similarly, Montes-Alcalá and Lapidus Shin (2011) tested Spanish−English mixed NPs in written and oral corpora, and they concluded that orthographic similarities, which largely reflect the phonological shape of words in many languages, influence gender assignment in written CS. Additionally, Rekun and Meir (2024) found that Russian−Hebrew bilinguals, where Russian features three genders and Hebrew possesses two genders, used the shape-based strategy in adjective-noun phrases during online acceptability judgment tasks. Both heritage speakers and immigrant speakers preferred this strategy, though other strategies were also observed. However, due to the logographic nature of Chinese characters and the limited correspondence between Pinyin endings and Spanish gender markers, the shape-based strategy is unlikely to influence gender assignment in this population. Thus, testing this strategy is reserved for future research, once the roles of other strategies are better understood.

To investigate these strategies empirically, researchers have employed a range of experimental methodologies. The next section reviews the main production and comprehension tasks used to study gender assignment in mixed NPs.

3.2.2 Production and comprehension tasks

Studies on gender assignment in mixed NPs have used production and comprehension tasks. Production tasks often involve corpus-based analyses of spontaneous speech and naturalistic data, reflecting actual language use (Gullberg et al., 2009). However, these methods are resource-intensive with limited evidence for specific questions (Gullberg et al., 2009). To mitigate these challenges, researchers have also used semi-structured interviews (Otheguy & Lapidus, 2003), interactive tasks like director-matcher and map tasks (Gullberg et al., 2009; Beatty-Martínez & Dussias, 2017), and sentence repetition tasks. Comprehension tasks, such as Likert scale evaluations (Liceras et al., 2008; Parafita Couto & Stadthagen-González, 2019) and forced-choice judgment tasks (Bellamy et al., 2018), assess responses to written or auditory stimuli. Together, these methods effectively elicit speakers' grammatical knowledge (Bellamy & Parafita Couto, 2022).

While these methodological approaches provide important insights, it is also crucial to consider how extra-linguistic factors, such as task type, bilingual profile, and community context, influence gender assignment strategies. The following subsections address these variables in greater detail.

3.2.3 Extra-linguistic factors in gender assignment strategies

3.2.3.1 Task type

Previous studies reveal variations in gender assignment strategies across task types, even within the same bilingual group. For instance, Bellamy et al. (2018) found that early Purepecha−Spanish bilinguals favored the masculine default strategy in a director-matcher task. However, they used mixed strategies in a forced-choice judgment task, assigning feminine gender to Purepecha nouns ending in *-a* and masculine to those ending in *-i*/*-u*. Such variation may stem from orthographic or morpho-phonological similarities between languages, which may amplify the influence of phonological and orthographic cues in written tasks, as observed in Basque–Spanish bilinguals, where noun endings influenced gender assignment in mixed NPs (Parafita Couto et al., 2016).

3.2.3.2 Bilingual profile

Bilingual profiles and language experience influence gender assignment strategies. Munarriz-Ibarrola et al. (2022) found that Spanish–Basque bilinguals with distinct bilingual profiles used different strategies in a forced-switch director-matcher task. Bilinguals who first acquired Spanish (a gender-marked language) preferred the analogical criterion (translation equivalent strategy), while those who first learned Basque (a non-gender-marked language) primarily used a phonological criterion (shape-based strategy).

3.2.3.3 Community characteristics

The characteristics of bilingual communities also affect gender assignment strategies among bilinguals. For example, in the well-established Spanish–English community in Miami, bilinguals prefer the masculine default (Balam et al., 2021). Conversely, in communities where speakers are more exposed to monolingual Spanish than in Miami, like El Paso and Granada, Spanish–English bilinguals use both masculine and feminine determiners for English nouns with feminine equivalents, though masculine determiners dominate overall (Królikowska et al., 2019). Similarly, in New Mexico, Spanish–English bilinguals show a mix of masculine default and translation equivalent strategies (Cisneros et al., 2023). These findings highlight the impact of community characteristics and social network diversity on gender assignment in mixed NPs.

In sum, gender assignment strategies in mixed NPs vary based on task type, bilingual profile, and community characteristics (see Appendix 3.A). Taken together, these findings highlight the complexity and variability of gender assignment in mixed NPs. Building on this foundation, the present study investigates gender assignment strategies among early Spanish–Chinese bilinguals in Barcelona, employing a multi-task approach to capture the influence of both linguistic and extra-linguistic factors.

3.3 The current study

3.3.1 Research question and hypothesis

This study examines how early Spanish–Chinese bilinguals, heritage Chinese speakers in Barcelona, Spain, assign gender in mixed NPs across various tasks. Drawing on research suggesting task type influences gender assignment (e.g., Bellamy et al., 2018), we use a multi-task approach: a director-matcher task, a judgment task, and a repetition task. The study addresses a main question:

RQ: Which gender assignment strategies (default or translation equivalent) do early Spanish–Chinese bilinguals use when assigning gender to Chinese nouns in mixed Spanish–Chinese NPs in production and comprehension tasks, and to what extent will gender assignment be consistent across tasks?

Hypothesis: Building on prior research showing that gender assignment strategies in mixed NPs vary depending on linguistic, cognitive, and sociolinguistic factors, we hypothesize that early Spanish–Chinese bilinguals in Barcelona will not rely exclusively on a single gender assignment strategy. Instead, we expect their choices to vary across production and comprehension tasks, reflecting the combined influence of task type, individual bilingual experience, and community characteristics. Specifically, we predict greater use of the translation equivalent strategy in the comprehension task, where metalinguistic awareness is heightened, and more balanced or default-masculine use in production tasks, where cognitive demands differ. Moreover, we anticipate that the diverse and relatively young Spanish–Chinese bilingual community in Barcelona, characterized by social network variability and exposure to multiple language norms (monolingual Spanish and bilingual Spanish–Catalan interactions are common in the community), will contribute to flexible and mixed gender assignment patterns. These patterns may differ from those observed in more established or homogeneous bilingual communities.

3.4 Materials and methods

3.4.1 Participants

Forty-two early Spanish–Chinese bilinguals (M_{age} = 20.24 years and SD_{age} = 1.74, twelve males) participated in three tasks. Participants completed a background questionnaire based on the Bilingual Code-Switching Profile (BCSP) (Olson, 2022) to assess their language profiles. Based on the results of BCSP, participants reported that they were all raised in Barcelona, Spain, with most of them born in Spanish-speaking countries

(31 in Spain, 1 in Panama) and the rest in China (10) (see Table 3.4.1). They are early sequential bilinguals, acquiring Chinese first after birth in Chinese-speaking households and Spanish later through formal educational settings in Spain (M_{age} = 4.55 years, SD_{age} = 3.87). Of the 31 participants born in Spain, 20 were sent to China at birth and returned to Spain during childhood (M_{age} = 5.2 years, SD_{age} = 3.33). The 10 participants born in China moved to Spain at a young age (M_{age} = 7.7 years, SD_{age} = 2.06). Moreover, they reported their proficiency in Spanish and Chinese (see Figure 3.4.1), and frequency use of Chinese ($M_{frequency}$ = 37%, $SD_{frequency}$ = 0.177) and Spanish ($M_{frequency}$ = 40%, $SD_{frequency}$ = 0.152) on a daily basis, as well as their attitudes towards mixing languages (see Figure 3.4.2). Specifically, most participants felt confident in extended conversation in Chinese (71%, n = 30) and Spanish (76%, n = 32), whereas only a small number of them felt confident in basic conversation in Chinese (10%, n = 4) and Spanish (2%, n = 1). Regarding their attitudes towards mixing Chinese and Spanish within the same conversation, most of them (62%, n = 26) believed that language mixing should not be avoided, whereas only a small number of them (12%, n = 5) felt it should be. Additionally, most participants (62%, n = 26) reported that they do not keep Chinese and Spanish separate in their daily conversation, indicating a natural and frequent integration of both languages in everyday use. However, participants' responses regarding Spanish–Chinese code-switching frequency were inconsistent across different but related questions in BCSP, leading to variability and limited reliability. As a result, we did not report a specific or quantified measure of their overall daily code-switching frequency. Instead, we provided more stable indicators based on reported code-switching with friends ($M_{frequency}$ = 21%, $SD_{frequency}$ = 0.311) and siblings ($M_{frequency}$ = 33%, $SD_{frequency}$ = 0.425), though these figures should be interpreted with caution, as they are not fully representative of broader language use. Participants were compensated for their participation upon task completion. All relevant materials, including the questionnaires, target pictures, sentences, and task results, are available on the Open Science Framework at **OSF**.

Table 3.4.1. Linguistic characteristics of early Spanish–Chinese bilingual participants (N = 42) for Experiments 1, 2, and 3.

Experiments 1, 2 and 3	Mean (SD)/Distribution
Number of female/male participants	30/12
Mean age in years (SD; range)	20.24 (1.75;18-26)
Number of participants born in Spain/Panama	31/1
Number of participants born in China	10
Age of Chinese acquisition	After birth
Age of Spanish acquisition	4.55 years old (SD = 3.87)
Daily use of Chinese (frequency)	37% (SD = 0.177)
Daily use of Spanish (frequency)	40% (SD = 0.152)
Daily use of Catalan (frequency)	13.8% (SD = 0.108)
Daily use of English (frequency)	6.8% (SD = 0.067)
Frequency of Spanish–Chinese code-switching with siblings	33% (SD = 0.425)
Frequency of Spanish–Chinese code-switching with friends	21% (SD = 0.311)

Figure 3.4.1. Participants' self-reported proficiency in Spanish and Chinese.

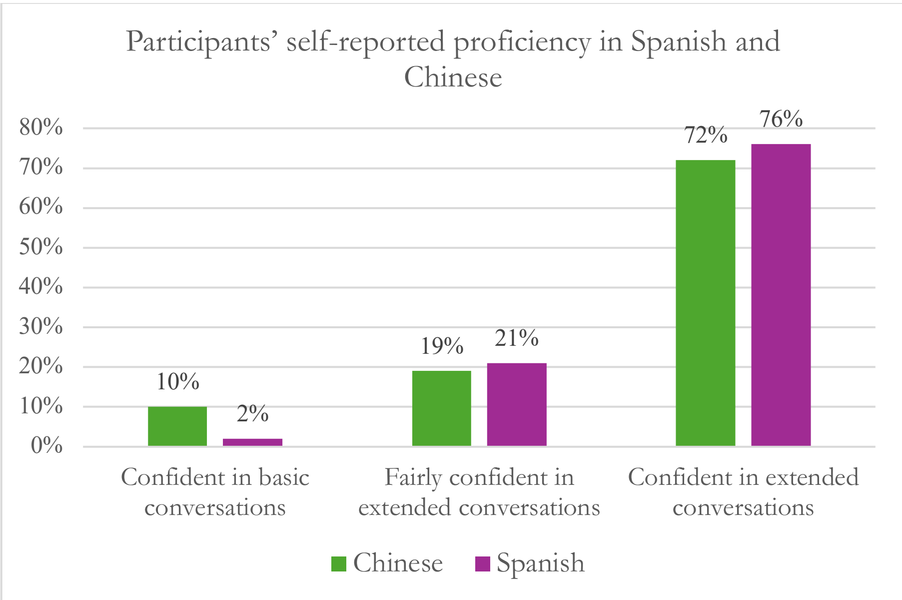
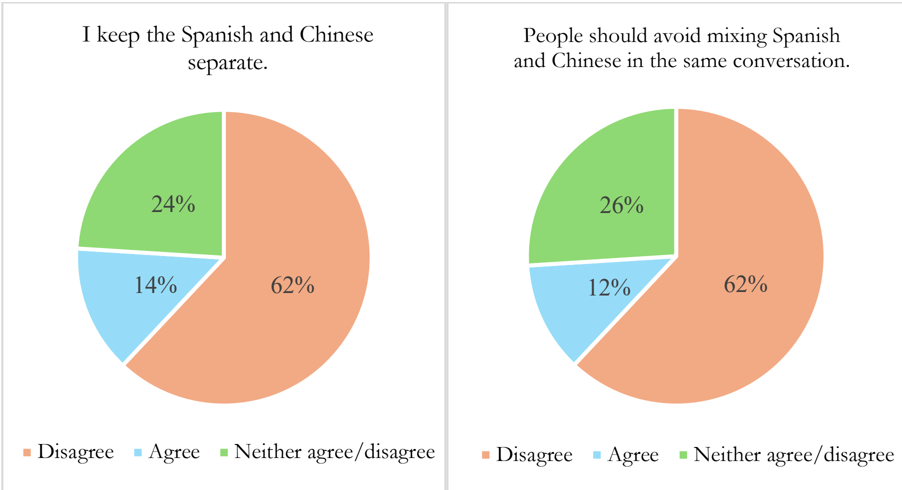


Figure 3.4.2. Participants' attitudes towards mixing Spanish and Chinese languages (should be avoided).



3.4.1.1 Spanish-Chinese bilingual community in Spain

Data from the Instituto Nacional de Estadística (INE, 2016) shows that Spain is home to 171,508 Chinese nationals, making them the second-largest non-EU immigrant group (Robles-Llana, 2018). About 70% of these migrants come from Qingtian, Zhejiang, a region with a history of overseas migration (Beltrán Antolín, 2006). Many first-generation migrants have limited education, work in family businesses (e.g., restaurants), and maintain traditional Chinese cultural values, remaining socially and culturally separate from broader Spanish society (Beltrán Antolín, 2008). Consequently, second-generation heritage speakers often act as translators and mediators between the Chinese community and the local population (Robles-Llana, 2018).

In Barcelona, the Chinese population grew from 13,416 in 2003 to 56,017 in 2020, with children under 16 increasing from 2,412 to 12,285 (He, 2024). This rapid increase contributed to the emergence of Spanish–Chinese bilingual communities, which are relatively young compared to more established bilingual groups, such as the Spanish–English communities in Miami. Moreover, most second-generation immigrants acquired Chinese from birth in the home environment and learned Spanish through immersion in the broader Spanish society, resulting in high proficiency in Chinese and Spanish. Additionally, they were raised in a multilingual environment and interacted with various linguistic groups, including Spanish and Chinese monolinguals, Spanish–Catalan bilinguals, and Spanish–English bilinguals. Research (e.g., Bellamy & Parafita Couto, 2022) suggests that such community characteristics likely influence gender assignment strategies in mixed NPs, a factor considered in our hypotheses.

3.4.2 Production tasks

3.4.2.1 Experiment 1: Director-Matcher task

Materials

Thirty-two colored line drawings were selected from the Multipic database (Duñabeitia et al., 2018) based on the following criteria: pictures represented familiar, concrete concepts with easily recognizable features; each picture had one corresponding noun in both Spanish and Chinese; and a low proportion of animate items was included to avoid potential influence of biological gender on gender assignment (Balam, 2016). Finally, 16 pictures corresponding to masculine nouns and 16 to feminine nouns in Spanish were selected, ensuring balance for Spanish grammatical gender (i.e., target pictures are available at https://osf.io/5c4se/?view_only=d67944cc3fa9413eb2bfcf819f5c3f0e).

Procedure

The experiment consisted of two sessions: familiarization and experimental. During familiarization, participants learned the Spanish and Chinese names of each picture (3,000 ms). After viewing 32 pictures, participants were paired for a forced-switch director-matcher task ("toy task"). In this task, participant "A" (the director) described each picture to participant "B" (the matcher), who then placed pictures on a grid. A hardboard separated the grids to prevent participants from seeing each other's boards (see Figure 3.4.3). Both participants worked with the same grid of 32 blank spaces and the same set of 32 pictures.

Figure 3.4.3. Illustration of the Director-Matcher task.

This task began with auditory instructions in CS mode. Participants described picture details (name, color, position) in Spanish while naming the objects in Chinese. This required participants to code-switch and assign grammatical gender to Chinese nouns (Bellamy et al., 2018). In this procedure, the director described each picture and its location on the grid to the matcher, who then placed the pictures accordingly. After all the pictures were placed, the matcher confirmed their positions with the director. Sample constructions are as follows, with Chinese nouns in italics.

1. un/el/una/la	*牛奶*	blanca/o
ART.INDEF/DEF.MASC/FEM	milk	white.FEM/MASC

"A/The white milk" (cf. una/la leche)

2. un/el/una/la	*太阳*	amarillo/a
ART.INDEF/DEF.MASC/FEM	sun	yellow.MASC/FEM

"A/The yellow sun" (cf. un/el sol)

3.4.2.2 Experiment 2: Repetition task

Materials

In this task, 32 pairs of code-switched sentences were created, each containing a mixed NPs (Spanish determiner + Chinese noun) based on the Chinese names of the 32 pictures from Experiment 1. First, 32 Spanish sentences with unilingual Spanish NPs were selected. The Spanish nouns were then replaced with their Chinese equivalents to create 32 code-switched sentences. Each mixed NP was constructed in two versions: one with gender match (i.e., where the Spanish determiner and the Spanish equivalent of the Chinese noun matched in gender) and one with gender mismatch. This generated 32 pairs of sentences contrasting the gender of determiners in mixed NPs.

In total, 64 trials were created: half with Spanish determiners gender-matched with Spanish equivalents of Chinese nouns, and the other half with a gender mismatch. Four singular Spanish determiners (el_{MASC}, la_{FEM}, un_{MASC}, una_{FEM}) were used, as Mandarin Chinese, a non-inflectional language, lacks the plural inflectional morphology on nouns and appears in bare form without plural markers (Yang, 1998). Mixed NPs were placed at the beginning (3), middle (4), and end (5) of sentences to control for position effects, with the number of masculine and feminine determiners counterbalanced (see Appendix 3.B). Sample sentences with mixed NPs in italics and bold:

3.*Un/una 香蕉* realmente es un remedio natural para muchas enfermedades.

"***A banana*** is really a natural remedy for many illnesses."

4. Las botellas de vino que vi sobre ***la/el 桌子***, ¿las has comprado tú?
"The bottles of wine that I saw on ***the table***, did you buy them?"

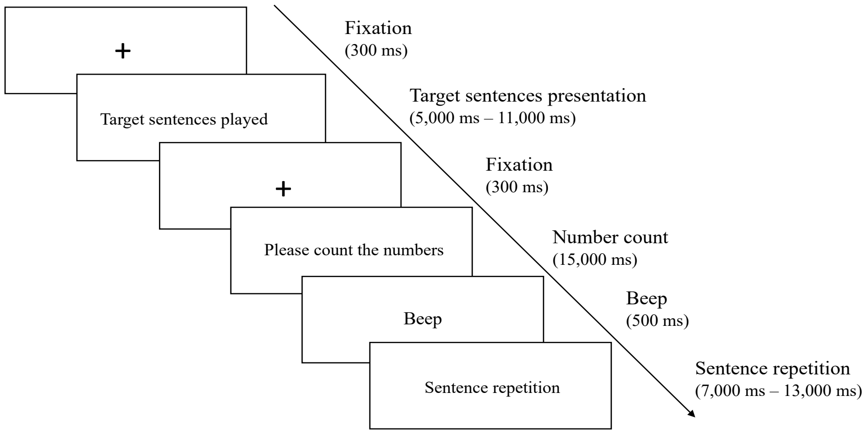
5. En la esquina, un joven estaba sentado solo, tocando suavemente ***una/un 吉他***.
"In the corner, one youngster sat alone, softly playing ***a guitar***."

Procedure

The repetition task, programmed in E-prime 2.0 (Schneider et al., 2002), used a by-subject order design to randomize orders of sentences across participants. Each participant, seated in a soundproof booth, listened to and repeated 64 code-switched sentences. Sentences were played once, and participants reproduced the content of the sentences as accurately as possible. A fifteen-second delay with an intervening counting task was introduced between each sentence and its repetition. This delay helped elicit responses based on participants' underlying grammar if the repetition exceeded the capacity of their short-term memory (Bullock & Toribio, 2009; Gullberg et al., 2009). In this task, participants were required to repeat sentences as accurately as possible after hearing them. The underlying rationale is that when input sentences exceed listeners' capacity of short-term memory, they rely on their internal grammatical knowledge to reconstruct the sentence (Gullberg et al., 2009). If a particular grammatical element is not represented in participants' internal grammar, it is likely to be modified or replaced during repetition (Vinther, 2002). Consequently, participants may either produce accurate repetitions or introduce changes to sentences that reflect their own underlying grammar (Gullberg et al., 2009).

Each trial followed this procedure: a fixation cross appeared at the center of the screen (300 ms), followed by the play of a code-switched sentence. After the sentence ended, another fixation cross appeared (500 ms). Participants then counted in their preferred language (Chinese or Spanish, 15 s) within a given range (e.g., one to ten or ten to one). A 500 ms beep prompted them to repeat the sentences (7,000 − 13,000 ms), with their vocal responses recorded automatically (see Figure 3.4.4).

Figure 3.4.4. Illustration of the experimental session in the repetition task.



3.4.3 Comprehension task

3.4.3.1 Experiment 3: two-alternative forced-choice judgment task

Materials

Critical sentences: Following procedures similar to Experiment 2, 64 critical sentences were generated, forming 32 pairs of comparison sentences. Each pair differed based on whether the gender of Spanish determiners matched the gender of the Spanish equivalents of Chinese nouns. Four Spanish singular determiners (el_{MASC}, la_{FEM}, un_{MASC}, una_{FEM}) were used. Half of the mixed NPs had gender-matched determiners, and half were mismatched. The position of mixed NPs within sentences (beginning, middle, end) and the proportions of each determiner (el/un/la/una) were counterbalanced across sentences (see Appendix 3.C).

Filler sentences: 64 filler sentence pairs were created, each differing in adjectives, verbs, or Chinese classifiers, resulting in 128 filler trials.

Procedure

This task was administered online via Qualtrics (Qualtrics, LLC, Provo, UT, USA) using a by-subjects order design. 192 trials were randomized for each participant, including 64 critical (contrasting determiners) and 128 filler trials (contrasting adjectives, verbs, or classifiers). Participants read each pair and chose the sentence they found more natural when speaking to other bilinguals. Choices should be made before proceeding to the next, and previous trials cannot be revisited.

3.5 Results and analysis

Data from the production tasks were transcribed, and mixed NPs with a Spanish determiner and a Chinese noun/adjective were extracted. Data from two production tasks and the comprehension task were analyzed descriptively and statistically using one-way repeated measures ANOVA in RStudio (v4.2.2; R Core Team, 2020). This analysis examined whether the structure of mixed NPs, specifically in terms of gender congruency, significantly influenced gender assignment choices in each task. The structure of mixed NPs, the independent variable (IV), had four levels: feminine determiners with Chinese nouns having feminine or masculine equivalents and masculine determiners with Chinese nouns having masculine or feminine equivalents. The dependent variable (DV) was the number of gender assignment responses for each NP structure. Results showed the structure of mixed NPs significantly affected gender assignment in all three tasks: director-matcher task ($F(3, 123) = 81.73, p < .001, \eta^2 = 0.656$), repetition task ($F(3, 123) = 61.55, p < .001, \eta^2 = 0.57$) and judgment task ($F(3, 123) = 85.47, p < .001, \eta^2 = 0.676$). These findings highlight significant differences in gender assignment choices across NP structures.

3.5.1 Production tasks

3.5.1.1 Experiment 1: Director-Matcher task

In the director-matcher task, 1,344 trials were generated. Of these, 91 trials (6.77%) were excluded due to the production of single nouns, like *黄瓜* "cucumber", or monolingual Spanish or Chinese NPs. This left 1,253 trials for further analysis. Example (6) illustrates the sentence that participants produced in the director-matcher task, where the Spanish determiner or adjective was used with a Chinese noun.

6. Y luego la segunda fila empieza por **un** *鸡蛋* que está abierto.
And then the second row start by **ART.INDEF.MASC *huevo*** that is open.
"And then the second row starts with an egg that is open."

(participant 9)

Example (7) shows the use of the masculine default strategy, even when the Spanish equivalent is feminine.

7. el *嘴* abierto
ART.DEF.MASC mouth open.MASC
"The open mouth" (cf. la boca abierta)

(participant 26)

Table 3.5.1 presents the distribution of mixed NPs. Participants preferred to use masculine determiners to nouns with masculine equivalents (597 tokens, 94.61%), compared to feminine determiners (34 tokens, 5.39%) ($p < .001$, $t = -29.51$). Conversely, for feminine equivalents, feminine determiners (338 trials/54.34%) were similar to masculine determiners (284 trials/45.66%) ($p = 1$, $t = 0.91$).

Table 3.5.1. Director-matcher task: mixed NPs.

	Masculine determiner	**Feminine determiner**	**Total (Nouns)**
Masculine translation equivalent	597 (94.61%)	34 (5.39%)	631
Feminine translation equivalent	284 (45.66%)	338 (54.34%)	622
Total	881	372	1,253

3.5.1.2 Experiment 2: Repetition task

In the repetition task, 2,411 trials were obtained, with 277 (10.30%) excluded for the same reason as in the director-matcher task. Examples (8−11) in Appendix 3.D illustrate how participants revised mixed NPs during their repetitions. Table 3.5.2 provides the distribution of revised and unrevised mixed NPs. Out of 794 revised trials, 583 (73.42%) originally had feminine gender, which was changed to masculine, regardless of nouns' gender equivalents. Only 211 (26.58%) with masculine gender were revised to feminine. For unrevised trials (1,617), most (997 trials, 61.66%) had masculine gender, while 620 (38.34%) had feminine gender.

Table 3.5.2. Repetition task: revised and unrevised mixed NPs.

Gender (original)	**Mixed NPs**	**Original trials (after exclusion)**	**Revised NPs**	**Unrevised NPs**
Feminine	Feminine determiners + feminine translation equivalent	602	204 (25.69%)	398 (24.61%)
	Feminine determiner + masculine translation equivalent	601	379 (47.73%)	222 (13.73%)
Masculine	Masculine determiner + masculine translation equivalent	585	48 (6.05%)	537 (33.21%)
	Masculine determiner + feminine translation equivalent	623	163 (20.53%)	460 (28.45%)
	Total	2,411	794	1,617

Table 3.5.3 shows that participants preferred to use masculine determiners with nouns having masculine equivalents (77.23%) over feminine determiners (22.77%) ($p < .001$, $t = -13.50$). For feminine equivalents,

masculine determiners (664 trials/54.21%) were similar to feminine determiners (561 trials, 45.79%) (p = 0.81, t = −1.52).

Table 3.5.3. Repetition task: translation equivalence.

	Masculine determiner	**Feminine determiner**	**Total (Nouns)**
Masculine translation equivalent	916 (77.23%)	270 (22.77%)	1,186
Feminine translation equivalent	664 (54.21%)	561 (45.79%)	1,225
Total	1,580	831	2,411

3.5.2 Comprehension task: two-alternative forced-choice judgment task

In the judgment task, participants selected the more natural-sounding sentence from pairs differing in the gender of determiners, resulting in 1,344 responses. Table 3.5.4 provides an overview of the distribution of gender preferences.

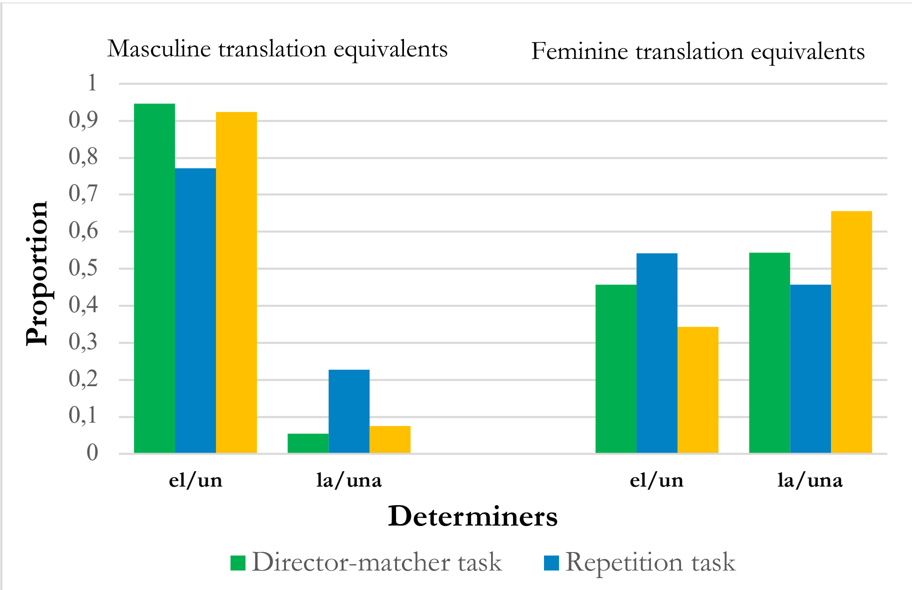
Table 3.5.4. Judgment task: distribution of gender preferences.

	Masculine determiner	**Feminine determiner**	**Total (Nouns)**
Masculine translation equivalent	621 (92.41%)	51 (7.59%)	672
Feminine translation equivalent	231 (34.38%)	441 (65.62)	672
Total	852	492	1,344

Table 3.5.4 shows that participants preferred masculine determiners (92.41%) for nouns with masculine equivalents, compared to feminine determiners (7.59%) (p < .001, t = −29.51). For nouns with feminine equivalents, feminine determiners (65.62%) were favored over masculine determiners (34.38%) (p = .011, t = 3.34).

Overall, bilinguals preferred the masculine gender for masculine equivalents across three tasks. Conversely, the choice of feminine or masculine gender was similar for feminine equivalents, especially in the director-matcher and repetition tasks (see Figure 3.5.1). Moreover, participants showed a significant preference for the masculine determiner compared to the feminine determiner in the director-matcher task with $V = 694.5$, $p < .001$[4], the repetition task with $t(41) = 7.21$, $p < .001$, *95% CI* [12.66, 22.52], and the judgment task with $V = 649$, $p < .001$. Overall, a paired-samples t-test revealed that participants produced significantly more masculine than feminine determiners overall in the three tasks, $t(41) = 9.77$, $p < .001$, *95% CI* [30.20, 45.94].

Figure 3.5.1. Distribution of mixed NPs across three tasks.



[4] We conducted different statistical analyses to examine gender choices across tasks, selecting methods based on data characteristics such as normality. Specifically, we used the non-parametric Wilcoxon signed-rank test for the director-matcher and judgment tasks due to non-normal data distributions, and a paired-samples t-test for the repetition task, where the data met normality assumptions.

3.6 Discussion

This study provides the first systematic investigation of gender assignment strategies in mixed Spanish−Chinese NPs among early Spanish−Chinese bilinguals in Barcelona. By employing a multi-task design, including director-matcher, sentence repetition, and forced-choice judgment tasks, we were able to capture both production and comprehension processes, revealing nuanced patterns in bilingual gender assignment. Findings highlight the interplay between task type and gender assignment strategies, as well as the role of community diversity in shaping these bilingual practices.

Variation in gender assignment strategies across tasks

Our findings illustrate how gender assignment in mixed NPs is highly sensitive to task demands. In the production tasks (director-matcher, sentence repetition), participants alternated between the default masculine and the translation equivalent strategy, particularly when assigning gender to Chinese nouns with feminine Spanish equivalents. This variability suggests that, in spontaneous or less monitored language use, bilinguals do not rigidly adhere to a single strategy but flexibly draw on both the default gender and the semantic associations provided by translation equivalents. For Chinese nouns whose Spanish equivalents are masculine, the masculine determiner overwhelmingly predominated, aligning also with the unmarked status of masculine gender in Spanish and consistent with previous research in other bilingual contexts.

In comprehension (judgment task), however, participants showed a strong preference for the translation equivalent strategy, more consistently matching the gender of the Spanish determiner to that of the Spanish translation of the Chinese noun. This pattern indicates that when bilinguals are prompted to reflect on grammaticality or make explicit judgments, they are more likely to draw on metalinguistic knowledge and analogical reasoning, rather than defaulting to the unmarked masculine.

Notably, the design of the comprehension task, which presented bilinguals with sentence pairs contrasting Spanish determiners (gender-matched vs. gender-mismatched) alongside Chinese nouns, could have amplified participants' metalinguistic awareness and focus on ensuring gender equivalence. This heightened attention facilitated the consistent application of the translation equivalent strategy, resulting in a strong preference for gender-matched mixed NPs in the comprehension task (see Bellamy et al., 2018 for similar findings). The contrast between production and comprehension tasks highlights the importance of methodological triangulation: relying on a single task type would have obscured the full range of bilingual behavior observed in this community. The task-dependent variability observed here supports dynamic models of bilingual language processing, which posit that bilinguals flexibly deploy grammatical strategies based on contextual and cognitive demands (Beatty-Martínez & Dussias, 2019). The interplay between default and translation equivalent strategies underscores the adaptability of bilinguals.

Influence of diversity of social network in the community

The varied gender assignment strategies observed among Spanish−Chinese bilinguals in Barcelona contrast with the more uniform patterns documented in other bilingual communities, such as Spanish−English bilinguals in Belize and Miami (Balam, 2016; Valdés Kroff, 2016). In those well-established communities, a strong preference for the masculine default is evident, likely reflecting the shared linguistic norms of a large, stable, and relatively homogeneous bilingual population. In contrast, the Spanish−Chinese bilinguals in Barcelona exhibit greater variability, alternating between the default masculine and translation equivalent strategies. This pattern mirrors findings from Królikowska et al. (2019) and Cisneros et al. (2023) who documented similar variability among Spanish−English bilinguals in El Paso, Granada, and New Mexico. These studies attribute such flexibility to the diverse social networks present in these communities, where not everybody speaks the same language.

Indeed, the Spanish−Chinese bilinguals tested in Barcelona interact with a linguistically diverse environment, including Spanish and Chinese monolinguals, Catalan−Spanish bilinguals, Chinese−Catalan−Spanish multilinguals, and others with different language backgrounds (e.g., English). Although they reported using unilingual Spanish (40%) and Chinese (37%) in daily conversation, the remainder of their daily used languages (23%) may involve various languages (e.g., Catalan or English) or Spanish−Chinese code-switching with peers and siblings. Such exposure to varied language practices likely fosters a high degree of adaptability in their linguistic behavior. Specifically, this diversity seems to play a crucial role in shaping the observed flexibility in gender assignment strategies, as individuals navigate and accommodate multiple linguistic systems and norms. While both Spanish−English bilinguals in El Paso, Granada, and New Mexico and Spanish−Chinese bilinguals in Barcelona show an overall preference for masculine gender, the variability in strategy use underscores the significant role of community diversity and sociolinguistic context in shaping bilingual language behavior.

However, we refrain from drawing definitive conclusions about the direct impact of community diversity on the observed gender assignment strategies, primarily due to the absence of a systematic and reliable corpus of code-switching data specifically targeting Spanish determiner-Chinese noun contexts within this community. While our experimental findings suggest that social network diversity may foster flexibility in gender assignment, the lack of naturalistic, corpus-based evidence limits our ability to generalize these results to everyday language use. As highlighted by Beatty-Martínez et al. (2018), the structure and dynamics of bilingual communities play a crucial role in shaping code-switching patterns. Without a dedicated corpus capturing spontaneous interactions and mixed NP constructions, it remains challenging to fully evaluate how community-level linguistic diversity influences real-world code-switching and gender assignment. Therefore, future research should prioritize the development of such corpora to provide more systematic and ecologically valid insights into the interplay between social context and grammatical choices in bilingual communities.

Moreover, the relatively recent establishment of the Spanish–Chinese bilingual community in Barcelona, comprised largely of second-generation migrants with an average age of 21 years, likely shapes the observed patterns of gender assignment. As a young and still-evolving community, its linguistic norms are not yet as entrenched as those found in more established bilingual populations. This developmental stage may contribute to the greater variability and flexibility seen in gender assignment strategies, as individuals navigate and negotiate linguistic practices within a dynamic and heterogeneous environment. As the community continues to grow and its social networks become more stable, it is plausible that shared linguistic conventions will emerge and solidify over time. With increased intergenerational transmission and the gradual formation of community-wide norms, gender assignment strategies may begin to converge toward the more uniform patterns observed in older, well-established bilingual communities, such as the Spanish–English bilinguals in Miami. However, without a longitudinal corpus documenting naturalistic code-switching and mixed NP usage as the community matures, it remains difficult to predict the trajectory of these linguistic patterns with certainty. The absence of such a corpus underscores the importance of future research that systematically tracks language use and code-switching practices in this community over time. Doing so would not only clarify how community age and stability influence grammatical choices, but also provide valuable insights into the processes by which new bilingual norms are established and maintained.

Task design and cognitive processing

The task design critically shaped gender assignment strategies. In comprehension, the written mode eliminated the influence of phonological and orthographic cues, which are factors known to affect gender assignment in other bilingual contexts, such as Basque–Spanish (Parafita

Couto et al., 2016). However, given the absence of grammatical gender and the logographic nature of Chinese, Chinese characters were not supposed to provide any grammatical gender cues or cross-linguistic influences on gender assignment in mixed NPs. As such, we predicted that no shape-based strategies were expected to be observed. This prediction was supported by the findings: no evidence of shape-based strategies, involving phonological or morphological cues from Chinese characters or Pinyin endings, was observed. For instance, Chinese Pinyin endings such as *gua* in 黄瓜 (/huanggua/ "cucumber") do not align with Spanish feminine gender-marker *-a,* when its translation equivalent, *pepino*, is masculine. These findings suggest that shape-based strategies are irrelevant in this specific Spanish−Chinese bilingual context, likely due to limited overlap in the phonological and orthographic systems of the two languages.

When bilinguals produce mixed Spanish−Chinese NPs, the target noun activates lexical and syntactic features from both languages. This activation enables three primary gender assignment strategies: (1) adopting the gender of the Spanish equivalent; (2) defaulting to masculine, or (3) using morpho-phonological cues. For nouns with feminine translation equivalents, bilinguals alternated between assigning feminine gender (translation equivalent strategy) and masculine gender (default strategy). For nouns with masculine translation equivalents, the alignment of default and translation-equivalent gender consistently led to masculine assignment.

Limitations and future research

While the current study offers novel insights into gender assignment strategies among Spanish−Chinese bilinguals, several limitations should be acknowledged. First, the study focuses on a specific bilingual population in Barcelona, which may limit the generalizability of the findings to other Spanish−Chinese bilingual communities. Sociolinguistic factors such as frequency of code-switching, community norms, and exposure patterns may vary across regions, and future studies should explore whether similar patterns emerge in other contexts. Second, the study relies primarily on (semi-)experimental tasks, which, while controlled, may not fully capture spontaneous language use. The absence of naturalistic data limits our ability to assess how these gender assignment strategies unfold in everyday bilingual interactions. Third, variability in participants' self-reported code-switching behaviors, partly due to differing interpretations of what constitutes code-switching, may have introduced inconsistencies in our analysis of bilingual language practices. Future research would benefit from incorporating longitudinal and corpus-based methods to track gender assignment over time and across more diverse sociolinguistic settings. Such data would not only enhance the ecological validity of research in this area but also shed light on how community dynamics and linguistic norms evolve as bilingual communities mature. In sum, addressing these limitations will be crucial for advancing our understanding of grammatical gender assignment in mixed-language contexts and for uncovering the broader mechanisms underlying bilingual language processing.

3.7 Conclusion

In this study, we examined gender assignment strategies in mixed NPs among 42 early Spanish−Chinese bilinguals in Barcelona, Spain, focusing on gender assignment strategies used by bilinguals when assigning gender to Chinese nouns. Using a multi-task approach, we found that participants frequently assigned both masculine and feminine genders to nouns with feminine equivalents in production tasks. These patterns align with findings from other bilingual communities with similar sociolinguistic characteristics, such as reduced opportunities for code-switching and heterogeneous social interactions. The Spanish−Chinese bilingual community in Barcelona exemplifies how diversity in individual language strategies can foster shared norms of flexibility and adaptability, suggesting that linguistic variability itself may serve as a unifying feature within the community. This highlights a unique aspect of bilingualism, where diversity in language use is not merely tolerated but actively shapes community cohesion and linguistic practice. Overall, our findings advance the understanding of gender assignment in bilingual contexts by highlighting the importance of investigating communities with diverse linguistic and social experiences. The results demonstrate how language systems adapt to both social and cognitive demands, offering valuable insights into the mechanisms underlying bilingual language processing. While the scope and context of this study necessarily limit the breadth of our evidence, these findings lay important groundwork for future research, planting the seed for a deeper exploration of how community diversity and linguistic variability shape bilingual language use.

Acknowledgements

We are grateful to the participants of this study for their time and commitment in completing the experiments, as well as to the members of the Center for Brain and Cognition at Pompeu Fabra University (UPF) for their support. We especially thank Núria Sebastián Gallés, the director of the Speech Acquisition and Perception (SAP) research group at UPF, for her warm welcome and generous help and suggestions that made this research possible. We also extend our appreciation to Francesc Xavier Mayoral Terron, the Secretary of the Department of Technology (UPF), and Silvia Blanch Rabat, the Research Support of the Department of Technology (UPF), for their generous help with the setup of experimental equipment and continuous guidance during data collection. We also thank all members of the SAP research group and the participants of the 6th Variation and Language Processing Conference (VALP6) for their valuable feedback and suggestions during the presentation of this work. Finally, we would like to thank our anonymous reviewers for their contributions.

Data Availability Statement

The data supporting the findings of this study are openly available in Open Science Framework at https://osf.io/5c4se/?view_only=d67944cc3fa9413eb2bfcf819f5c3f0e (accessed on 24 May 2023) (view-only link).

Appendix

Appendix 3.A Overview of studies supporting different gender assignment strategies, including different language pairs, and data collec-tion tasks *(for a thorough overview, see Bellamy & Parafita Couto, 2022)* in Chapter 3.

Studies	Language pairs	Bilingual profile	Tasks	Gender assignment strategy
Parafita Couto et al. (2016)	Spanish–Basque	• Spanish–Basque bilinguals in the Basque Country	• Naturalistic data • Director-matcher task	• Shape-based strategy
			• Auditory judgment task	• Feminine default strategy (due to phonological cues)
Munarriz-Ibarrola et al. (2022)	Spanish–Basque	• Sequential Spanish–Basque bilinguals (Basque as their first language)	• Forced-switch elicitation task (director-matcher task)	• Shape-based strategy
		• Sequential/simultaneous Spanish–Basque bilinguals (Spanish as (one of) their first language)	• Forced-switch elicitation task (director-matcher task)	• Translation equivalent strategy
Liceras et al. (2008)	Spanish–English	• Spanish (L1)–English (L2) Bilinguals	• Acceptability judgment task	• Translation equivalent strategy
		• L1 French–L2 English–L3 Spanish speakers • L1 English–L2 Spanish bilinguals	• Acceptability judgment task • Naturalistic production data	• Masculine default strategy
Balam (2016)	Spanish–English	• Adult Spanish–English bilinguals in the Belize Country	• Naturalistic production data (Belize corpus)	• Masculine default strategy
Balam et al. (2021)	Spanish–English	• Spanish–English bilingual children in Miami (USA)	• Naturalistic production data (Miami corpus)	• Masculine default strategy
Valdés Kroff (2016)	Spanish–English	• Spanish–English bilinguals in Miami (USA)	• Naturalistic production data (Bangor Miami corpus)	• Masculine default strategy
		• Young adult	• Unconstrained	• Masculine

Królikowska et al. (2019)	Spanish−English	Spanish−English bilinguals from Granada (Spain), El Paso (Texas), State College (Pennsylvania) and San Juan (Puerto Rico).	director-matcher task	default strategy
		• Young adult Spanish− English bilinguals from Granada (Spain), El Paso (Texas)	• Unconstrained director-matcher task	• Translation equivalent strategy
Cisneros et al. (2023)	Spanish−English	• Spanish−English bilinguals in New Mexico (USA)	• Forced-switch toy task	• Masculine default strategy
			• Two-alternative forced-choice judgment task • Natural toy task • Forced-switch toy task	• Translation equivalent strategy
Cruz (2023)	Spanish−English	• Spanish−English bilinguals in Southern Arizona (USA)	• Force-choice elicitation task	• Masculine default strategy
Cruz (2021)	Spanish−English	• Spanish−English bilinguals in Southern Arizona (USA)	• The Corpus del Español en el Sur de Arizona (CESA) corpus	• Masculine default strategy • Biological gender is a reliable predictor for gender assignment
Bellamy et al. (2018)	Purepecha–Spanish	• Early sequential Purepecha–Spanish bilinguals in Michoacán (Mexico)	• Director-matcher task	• Masculine default strategy
			• Online forced-choice acceptability judgment task	• Feminine default strategy (due to phonological cues)
Rekun and Meir (2024)	Russian−Hebrew	• Russian−Hebrew bilinguals in Israel	• Online acceptability judgment tasks (auditory)	• Shape-based strategy
Fuller and Lehnert (2000)	German−English	• Late sequential German−English bilinguals	• Two German−English corpora	• Translation equivalent strategy
Bellamy and Wichers Schreur (2022)	Tsova-Tush−Georgian	• Tsova-Tush−Georgian bilinguals in Zemo Alvani (Georgia)	• Tsova-Tush corpus • Director-matcher task	• Translation equivalent strategy

Greidanus Romaneli et al. (2021)	Dutch–Portuguese	• Dutch–Portuguese bilinguals in Brazil	• Naturalistic production data (recorded conversation) • Director-matcher task • Acceptability judgment task	• Masculine default strategy

Appendix 3.B Number of masculine and feminine determiners in mixed NPs at different positions in the sentences in the repetition task in Chapter 3.

	Determiners	Beginning	Middle	End
Masculine determiner	el	3	3	2
	un	2	3	3
Feminine determiner	la	2	3	3
	una	3	3	2

Appendix 3.C Number of masculine and feminine determiners in mixed NPs at different positions in the sentences in the judgment task in Chapter 3.

	Determiners	Beginning	Middle	End
Masculine determiner	el	3	3	2
	un	2	3	3
Feminine determiner	la	3	3	2
	una	2	3	3

Appendix 3.D Examples (8), (9), (10), and (11), extracted from participant recordings, illustrate how participants revised sentences during their repetitions in the repetition task.

8. The matched mixed NP (i.e., feminine determiner + feminine translation equivalent) was revised to a mismatched mixed NP (i.e., masculine determiner + feminine translation equivalent).

Original sentence:

También quisiera hacer unos comentarios sobre la comparación con ***la 苹果***.
Also I'd like to make some comments about the comparison with ***DET.FEM apple***.
"I would also like to make a few comments about the comparison with ***the apple***."

Repetition:

También quisiera hacer unos comentarios sobre la comparación con ***el 苹果***.
[DET.MASC apple]

9. The mismatched mixed NP (i.e., Feminine determiner + Masculine translation equivalent) was revised to a matched mixed NP (i.e., Masculine determiner + Masculine translation equivalent).

Original sentence:

La 小提琴 es uno de los instrumentos más difíciles de aprender.
DET.FEM violin is one of the instruments most difficult to learn.
"**The violin** is one of the most difficult instruments to learn."

Repetition:

El 小提琴 es uno de los instrumentos más difíciles de aprender.
[DET.MASC violin]

10. The matched mixed NP (i.e., masculine determiner + masculine translation equivalent) was revised to a mismatched mixed NP (i.e., feminine determiner + masculine translation equivalent).

Original sentence:

Una pregunta retórica que a menudo recibe una respuesta no menos teórica es: ¿qué apareció antes, ***el 鸡蛋*** o la gallina?
DET.FEM question rhetoric that to slight receives an answer not less theoretical is: which appears first, ***DET.MASC huevo*** or the hen?
"A rhetorical question that often receives a no less theoretical answer is: which came first, **the egg** or the chicken?"

Repetition:

Una pregunta retórica que a menudo recibe una respuesta no menos teórica es: ¿qué apareció antes, ***la 鸡蛋*** o la gallina?
[DET.FEM huevo]

11. The mismatched mixed NP (i.e., masculine determiner + feminine translation equivalent) was revised to a matched mixed NP (i.e., feminine determiner + feminine translation equivalent).

Original sentence:

También podrá alquilar ***un 自行车*** para explorar esta hermosa ciudad.
Also can rent ***DET.MASC bicicleta*** to explore this beautiful city.
"You can also rent **a bike** to explore this beautiful city."

Repetition:

También podrá alquilar ***una 自行车*** para explorar esta hermosa ciudad.
[DET.FEM bicicleta]

Chapter 4

Classifier Congruency Effects in Chinese Noun Phrase Production: Behavioral and Electrophysiological Evidence from Spanish−Chinese Bilinguals

---

A version of this chapter is under review: Wu, R., Parafita Couto, M. C., & Schiller, N. O. (under review). Classifier congruency effects in Chinese noun phrase production: behavioral and electrophysiological evidence from Spanish−Chinese bilinguals.

102 *The Role of Lexico-Syntactic Features in Noun Phrase Production and Comprehension*

Abstract

Mandarin Chinese employs a classifier system requiring classifiers obligatorily positioned between demonstratives or numerals and nouns in noun phrases (NPs; e.g., *一张桌子*, *[one + specific-classifier-zhang1 + table]*, "one table"). While classifier congruency effects have been documented in monolingual Chinese speakers, it remains unclear whether such effects extend to early Spanish−Chinese bilinguals. This study examined classifier congruency effects in NP production by 29 early Spanish−Chinese bilinguals and 30 Mandarin Chinese speakers using behavioral and EEG measures. Participants completed picture-word interference (PWI) tasks with four conditions, manipulating classifier congruency (congruent vs. incongruent) and semantic relatedness (related vs. unrelated). Both groups showed significantly longer naming latencies for classifier-incongruent and semantically related conditions. ERP results revealed N400-like effects for classifier incongruency in both groups and semantic interference, reflected by N400-like effects in bilinguals and P600-like effects in Mandarin Chinese speakers. These findings indicate competitive selection of classifiers during NP production across both groups, highlighting the robustness of classifier congruency effects and the flexibility of morphosynactic processing in bilinguals.

Keywords: language production; classifier congruency effect; semantic interference effect; bilingualism; picture-word interference (PWI); N400; P600

4.1 Introduction

Speech production in monolinguals involves a complex, multi-staged process, but how does this process differ for bilinguals? Theories of bilingual lexical access suggest a general three-stage process, regardless of whether the lexical selection mechanism is language-specific or language-non-specific (e.g., Costa et al., 1999a; De Bot, 1992; De Bot & Schreuder, 1993; Dijkstra & Van Heuven, 1998; Green, 1986; Roelofs, 1998). First, lexical entries from both the active (language-in-use) and non-active (language-not-in-use) languages are activated in parallel. Second, a lexicon-external mechanism or production rule determines which language will be used. Third, lexical entries from the target language receive the highest activation and are ultimately selected for production, while those in the non-target language are suppressed.

A widely used method to investigate these processes is the picture-word interference (PWI) paradigm (e.g., Rosinski et al., 1975). In the PWI paradigm, participants name a picture while ignoring a distractor word presented before, simultaneously with, or after the picture onset. The LRM speech production model (Levelt et al., 1999) provides a framework for understanding the process of naming pictures in noun phrase (NP) production. According to this model, both the target picture and the distractor activate their conceptual and lexical representations first. Next, their syntactic features, such as grammatical gender, classifiers, or number, are retrieved. All related nodes are then activated, with the target-related node receiving the highest activation. Finally, the most activated node is selected, encoded, and articulated as part of the NP.

While monolinguals consistently follow this process, NP production in bilinguals involves additional layers of complexity. During bilingual production, syntactic representations and conceptually related nodes of both the target and distractor from target and non-target languages are activated. However, nodes in the target language receive higher activation, and the target-related node in the target language receives the highest activation. This is because an external production rule determines which language is intended to be produced. Thus, selection is restricted to nodes within the target language, with the highest activated node ultimately selected for production (e.g., Costa et al., 1999a; Roelofs, 1998).

Recent research has increasingly examined the processing of classifiers in bare noun and NP productions (e.g., Wang et al., 2019; Huang & Schiller, 2021; Wang & Schiller, submitted; Wang et al., submitted) and established the mechanisms underlying these processes in monolinguals. However, little is known about how bilinguals produce classifiers. Thus, this raises a critical question: does classifier production in bilingual NP processing involve competition and selection mechanisms similar to those observed in monolinguals?

4.1.1 The effects of processing lexico-syntactic features in language production

4.1.1.1 The gender congruency effect

Experimental research on speech production has mostly investigated lexico-syntactic features, such as grammatical gender and classifiers, using the PWI paradigm (Glaser, 1992; Schriefers, 1993). In this paradigm, it has been found that the relationship between distractor words and target pictures significantly affects picture-naming latencies. For example, Schriefers (1993) initially used the PWI paradigm to investigate the selection of grammatical gender in NP production in Dutch (a language that features common and neuter gender). Faster naming latencies were obtained when the grammatical gender of the target (*boek*$_{NEU}$, "book") was congruent with the distractor word (*dak*$_{NEU}$, "roof") compared to the incongruent one (*tafel*$_{COM}$, "table"). Moreover, these faster naming latencies in gender-congruent conditions than gender-incongruent conditions were continuously observed in Germanic languages (e.g., German in Bürki et al., 2016; Schiller & Caramazza, 2003 and Schriefers & Teruel, 2000; Dutch in Schiller, 2013 and Schiller & Caramazza, 2006) and Romance languages (e.g., Portuguese in Sá Leite et al., 2021; French in Alario & Caramazza, 2002; Spanish in Costa et al., 1999b and Wu & Schiller, 2023), as well as other gender-marking languages (see Wang & Schiller, 2019 and Sá Leite et al., 2022 for an overview). Overall, this faster and more accurate processing in gender-congruent conditions is known as the *gender congruency effect*.

4.1.1.2 The classifier congruency effect

More recently, the PWI paradigm has been employed to investigate classifier processing in Mandarin Chinese (hereafter, "Chinese" refers to Mandarin Chinese). For example, Wang et al. (2006) examined the selection of classifiers in Chinese bare noun and NP production using the PWI paradigm with Chinese monolingual speakers. Shorter naming latencies were observed when the classifier of target pictures was congruent with distractor words in NP production, but not in bare noun production. A similar effect in naming latencies was also observed in Chinese NP productions by Huang and Schiller (2021), where the classifier congruency (congruent vs. incongruent) and semantic relatedness (related vs. unrelated) between target pictures and distractor words were manipulated. Monolingual Chinese speakers named pictures while simultaneously ignoring presented distractor words, with their naming latencies and electrophysiological (EEG) activity being recorded simultaneously. Behavioral results showed that classifier-congruent distractors elicited faster naming latencies, while semantically related distractors increased naming latencies. EEG results showed a significant N400 effect modulated by classifier incongruency and semantically unrelatedness. This finding aligns with Wang et al. (2019) who observed a comparable N400 effect in Chinese bare noun production using similar experimental designs. They interpreted this effect as evidence of classifier activation during lexical selection. Given the similar inherent properties and processing mechanisms, researchers have identified the observed effects in naming latencies and electrophysiological responses during Chinese NP production, as well as the electrophysiological responses during Chinese bare noun production, as manifestations of the *classifier congruency effect*.

These behavioral and EEG findings from bare nouns and NP production in Chinese monolingual speakers confirm the robustness of the classifier congruency effect in Chinese speech production. However, since no studies have examined how bilinguals (e.g., Spanish−Chinese bilinguals) process classifiers in Chinese NP production, it remains unclear whether similar effects would be observed in Spanish−Chinese bilinguals. Moreover, given that Chinese classifiers and Spanish grammatical gender function as lexico-syntactic features with similar characteristics in NPs and comparable processing mechanisms, it is particularly interesting to investigate how Spanish−Chinese bilinguals process, select, and produce classifiers in their speech.

4.1.2 The semantic interference effect

A well-established effect using the PWI paradigm is the *semantic interference effect*. Greater interference occurs when distractor words are semantically related to target pictures (e.g., PIG, lion), compared to unrelated distractors (e.g., PIG, cup) or nonsense trigrams (Rosinski, 1977). This results in longer naming latencies in semantically related than unrelated conditions. According to the lexical selection by competition theory (Roelofs, 1992, 1993; Levelt et al., 1999), when the target picture (e.g., *pig*) and the distractor (e.g., *lion*) are semantically related, the corresponding lemmas activate each other through the spreading of activation to representations in the lexical network. This semantically related distractor (unlike unrelated distractors) receives activation from the semantic category node activated by the target. This results in more competition during the lexical selection process and delays the selection of the target lexical node, producing longer naming latencies. In the unrelated condition, the target and the distractor are not activating each other, resulting in relatively less competition and faster naming latencies.

4.1.3 Electrophysiological correlates of classifier congruency and semantic interference effects

Electroencephalography (EEG) and event-related potentials (ERPs) are particularly suitable methodologies for monitoring the complex processing of speakers' NP production. In previous studies, two ERP components seem to have been especially in focus, i.e., the N400 and the P600. The N400 is an ERP component characterized by a negative voltage amplitude peak around 400 ms post-stimulus onset, typically within the 250 ms and 600 ms time window over central and posterior electrode sites (Kutas & Hillyard, 1980; Chwilla et al., 1995). It is typically associated with lexical-semantic integration and lexical co-activation processes (e.g., Lau et al., 2008; Kutas & Federmeier, 2011; Chen et al., 2017; Leckey & Federmeier, 2019). Previous studies on the gender congruency effect and the classifier congruency effect have frequently reported modulations of the N400. Specifically, more negative N400 amplitudes were mostly reported as elicited by gender- or classifier-incongruent trials compared to gender- or classifier-congruent trials (see an overview of gender congruency effects in Wang & Schiller, 2019 and Sá Leite et al., 2022; classifier congruency effects in Wang et al., 2019 and Huang & Schiller, 2021). Additionally, the N400 was also reported for picture-naming tasks in semantically unrelated conditions relative to related conditions, reflecting semantic integration processes (e.g., Greenham et al., 2000; Blackford et al., 2012; Wang et al., 2019; Huang & Schiller, 2021). It was concluded that the N400 effect is elicited by competition at the lemma level, triggered by strong activation at the conceptual level (Wang et al., 2019). Thus, in this study, we used the N400 to detect the co-activation of the classifier congruency effect and lexico-semantic integration of the semantic interference effect.

The P600 is an ERP component defined by a positive-going deflection primarily localized in the centroparietal regions, typically occurring between 500 and 800 ms post-stimulus onset and peaking around 600 ms (Osterhout & Holcomb, 1992). Typically, the P600 is sensitive to syntactic violations in syntactic processing (Friederici et al., 1993; Hagoort et al., 1993). However, the P600 is not only necessarily associated with syntactic violations, rather, it has also been reported to be elicited by orthographic violations (Münte et al., 1998), garden path sentences (Friederici et al., 1996), grammatical violations (Hagoort et al., 1993), following N400 effects of semantic violations (Osterhout & Nicol, 1999), generally improbable events (Coulson et al., 1998), and semantic anomalies involving implausible or unexpected events (e.g., **The cat that fled from the mice ran through the room*; Kolk et al., 2003). Furthermore, the P600 was challenged in the view of an exclusively syntactic violation response, for instance, Kim and Osterhout (2005) proposed a "semantic P600" effect, typically distributed in the centro–parietal regions with a maximal difference occurring within 600 – 800 ms (also see Kuperberg et al., 2003; Van Herten et al., 2005; Bornkessel-Schlesewsky & Schlesewsky, 2008, for a review). This perspective redefines the P600 as an electrophysiological response to "integration difficulties", reflecting the cognitive effort required to semantically, rather than syntactically, integrate an entity into the discourse context. Delogu et al. (2019) provided support for this view by observing a P600-like effect when words, though semantically related to the context, were difficult to integrate into the syntactically correct discourses (i.e., implausible events). Additionally, Puhacheuskaya (2021) also pointed out that the regions and time windows of the "semantic P600" effect are largely consistent with the centro–parietal P600 effect, which is primarily documented by most studies as the effect associated with syntactic reanalysis and repair. Given the view of the "semantic P600" effect, we also consider the P600 to detect the semantic interference effect.

4.1.4 The feature of classifiers and NPs in Chinese

Almost all languages incorporate some form of nominal classification in their grammar, with grammatical gender systems (e.g., in Spanish) and numeral classification systems (e.g., in Chinese) being the most common (Seifart, 2010). In Spanish, nouns inherently possess an identifiable grammatical gender, classified as either masculine or feminine (Roca, 1989;

Harris, 1991). Determiners and adjectives within Spanish NPs, while not inherently gender-marked themselves, must obligatorily agree in gender with nouns they modify (Harris, 1991). Thus, the grammatical gender choice within NPs primarily depends on the properties of nouns. In contrast, Chinese does not feature a grammatical gender classification system but instead employs a classifier system that is in some way comparable to gender division (Wang et al., 2019). Classifiers, as free morphemes, classify the category of nouns they accompany, with their selection determined by particular properties of nouns (Li & Thompson, 1981). In Chinese NPs, determiners, numerals, or quantifiers cannot directly modify nouns. Instead, a numeral classifier is obligatorily inserted as a mediating element (Li, 2013). Thus, a typical Chinese NP consists of demonstrative/numeral /quantifier + classifier + noun (e.g., *一只老虎*, /yi1[5]zhi1lao3hu3/ *[one classifier-zhi1 tiger]*, "one tiger") (Li & Thompson, 1981). In general, classifiers function similarly to grammatical gender in nominal systems, as they represent inherent properties of nouns and cannot be omitted in NPs.

In Mandarin Chinese, there are several dozen classifiers with approximately 150 commonly used (Erbaugh, 1986). According to Erbaugh (2006), these classifiers can be generally classified into five categories based on their lexical type and function: (1) measure classifiers, used for measuring weight or length (e.g., *公里* /gong1li3/ "kilometer", *两* /liang3/ "ounce"); (2) kind classifiers that describe categories or types of entities (e.g., *种* /zhong3/ "type", *类* /lei4/ "kind, category"); (3) collective classifiers, indicating arrangements of entities (e.g., *排* /pai2/ "row of" or *捆* /kun3/ "bundle"); (4) event classifiers that are used for events (e.g.,*班* /ban1/ "run of a bus or train", *场* /chang3/ "performance of a show"); (5) sortal classifiers, used for particular categories based on prope-

[5] In Mandarin Chinese, tones are essential for phonemic distinction. The language has four lexical tones: Tone 1 is high-level, Tone 2 is high-rise, Tone 3 is low-dip, and Tone 4 is high-falls (Chao, 1948). In this study, numerical markers indicate the corresponding tone in Pinyin transcriptions.

rties of entities (e.g., 棵 /ke1/ "classifier for plants", 块 /kuai4/ "chunk or square"). Mandarin Chinese has approximately 75 sortal classifiers, each typically used with 5 to 20 nouns (Erbaugh, 2002). Typically, classifier selection is determined by properties of nouns, such as animacy, shape, function, and size (Myers & Tsay, 2000). It should be noted that the relationship between nouns and classifiers is, in many cases, opaque, although choices of classifiers in NPs depend to a great extent on semantic properties (Shao, 1993; Tzeng et al., 1991). Additionally, there is a special so-called "general" or "default" classifier *ge* (个 /ge4/), which is the most frequently used classifier (Li & Thompson, 1981), with approximately 40% of Chinese nouns only taking this classifier, such as unique objects (e.g., "the sun") and abstractions (e.g., "dream", "idea") (Erbaugh, 2006). Moreover, compared to other classifier categories, Mandarin Chinese speakers often prefer sortal classifiers when referring to unfamiliar nouns or those with new information (Erbaugh, 2006). Therefore, both the general classifier and sortal classifiers were considered as the primary target classifiers in this study.

4.1.5 The current study

This study explores the processing of classifiers in Chinese NPs from behavioral and neural perspectives, focusing on Mandarin Chinese speakers who were born and grew up in China (hereafter "Mandarin Chinese speakers" refer to Chinese speakers born and raised in monolingual Mandarin-speaking environments in China) and early Spanish−Chinese bilingual speakers in Spain. In this study, we manipulate classifier congruency (classifier-congruent vs. classifier-incongruent) and semantic relatedness (semantically related vs. semantically unrelated) between target objects and distractor words. The study serves the main objective of examining whether classifier features are automatically activated and competitively selected during Chinese NP production in bilingual speakers. Specifically, we aim to determine whether this activation is evident in specific ERP components (e.g., N400) and whether competitive selections influence naming latencies, thereby resulting in a classifier congruency effect in both behavioral and neural measures. This leads to our primary research question: is there a classifier congruency effect in Chinese NP production by early Spanish−Chinese bilinguals?

4.1.5.1 Hypotheses

Behavioral hypotheses. We predict effects on classifier congruency and semantic interference on behavioral measures of naming latencies in both Mandarin Chinese speakers and early Spanish−Chinese bilingual speakers. Based on the lexical selection by competition theory (Roelofs, 1992, 1993; Levelt et al., 1999), lexical selections involve competitive processes among activated lemmas. Accordingly, for classifier-congruent targets and distractors, we predict a *classifier congruency effect*, reflected in less competition and shorter naming latencies in classifier-congruent compared to incongruent conditions.

For targets that are semantically related to distractors (i.e., belonging to the same semantic category), we expect a *semantic interference effect*, due to more competition and thus longer naming latencies for semantically related compared to unrelated conditions. However, according to the LRM (Levelt et al., 1999), lexico-semantic and lexico-syntactic information are independently activated and selected at the lemma level, with each following a distinct competitive selection process. Additionally, given that the mapping between classifiers and semantic features of nouns is often opaque, classifier selection is primarily driven by individual nouns rather than broad semantic categories (Shao, 1993; Tzeng et al., 1991). Consequently, no interaction is expected between the classifier congruency effect and the semantic interference effect.

EEG hypotheses. We predict a more negative N400 amplitude for classifier-incongruent conditions compared to congruent conditions between 250 – 600 ms after picture onset in Mandarin Chinese speakers. Given the similarities in processing mechanisms of both monolingual and bilingual language production, we also expect a similar N400 effect in early Spanish−Chinese bilinguals under classifier-incongruent conditions.

For the semantic interference effect, we similarly expect, based on Wang et al. (2019) and Huang and Schiller (2021), a more negative N400 amplitude elicited by semantically unrelated conditions compared to related conditions in two groups. However, considering prior research identifying the "semantic P600" effect as a response to "semantic integration difficulties", it is plausible that semantically unrelated conditions may also elicit a "semantic P600" effect, reflecting increased integration challenges when target pictures are semantically unrelated to distractors.

4.2 Methods

4.2.1 Participants

This study involved two groups of participants: Mandarin Chinese speakers and early Spanish−Chinese bilinguals. The Mandarin Chinese group consisted of 30 healthy, right-handed native Chinese speakers (21 females) with a mean age of 24.36 years (SD = 3.02), recruited from Pompeu Fabra University in Barcelona, Spain. All speakers in this group were born and raised in China (see details below). Participants in this group completed the Language Experience and Proficiency Questionnaire (LEAP-Q; Marian et al., 2007) prior to the experiments to assess their linguistic profiles, including language experience and usage. The Spanish−Chinese bilingual group included 29 healthy, right-handed early bilinguals (24 females) with a mean age of 21.01 years (SD = 3.01), also recruited at Pompeu Fabra University. These participants were required to complete a revised BCSP questionnaire developed based on the Bilingual Code-Switching Profile (BCSP; Olson, 2022), a Chinese Elicited Imitation (EI) test (Yan et al., 2020) and a short version of the Diplomas de Español como Lengua Extranjera (DELE) Spanish proficiency test (Instituto Cervantes, 2010) to evaluate their language proficiency, experience, and usage. All participants reported having no language disorders, psychological conditions, or visual impairments. The informed consent was filled out before the experiments, and compensation was provided upon completion of the experiments.

4.2.1.1 LEAP-Q: Mandarin Chinese group

Within the Mandarin Chinese group, all speakers were born and raised in monolingual Mandarin Chinese-speaking environments in China, ensuring native-level proficiency and natural exposure to Mandarin Chinese from birth. Of these, twenty-two participants were pursuing master's degrees, while the remaining were enrolled in bachelor's (n = 3) or doctoral programs (n = 5) in Barcelona, Spain. Although our initial efforts were to recruit monolingual Chinese speakers without any knowledge of Spanish or other languages (e.g., English), fourteen participants with Spanish proficiency below the B1 level were ultimately included, due to Spanish universities' requirements for master's students in Spain to attain basic Spanish proficiency. Additionally, given the global prevalence of English and its integration into higher education in China and abroad, all participants inevitably had an average English proficiency of approximately the B2 level. On average, participants reported using Chinese for approximately 70.18% (*SD* = 0.18) of their daily language use. Detailed linguistic profiles are provided in Appendix 4.A1.

4.2.1.2 Revised BCSP: early Spanish−Chinese bilingual group

All early Spanish−Chinese bilingual participants were either born in Spain (n = 25) or China (n = 4), and all were raised and currently reside in Barcelona, Spain. They grew up in households where Chinese served as the primary language spoken by their parents, acquiring it as a heritage language. Spanish, however, is the dominant language in their broader social and educational environments and is extensively used in daily life. Most participants learned Chinese in informal contexts, primarily at home or through heritage language schools (e.g., weekend or community Chinese language schools), where formal exposure to Chinese was typically limited to a few hours per week. The majority of participants (n = 22) were pursuing bachelor's degrees, while the remaining participants were either continuing master's studies (n = 4) or had recently graduated from high school (n = 3). On average, participants reported using Chinese for 32.55% (*SD* = 0.17) and Spanish for 44.87% (*SD* = 0.16) of their daily communication. All participants reported Chinese and Spanish as their native languages. Additionally, sixteen participants also identified Catalan as their native language, while the remaining participants either learned Catalan as a second language (L2) (n = 11) or reported no proficiency in Catalan (n = 2). Further details of the participants' linguistic profiles are provided in Appendix 4.A2.

4.2.2 Materials and design

The Mandarin Chinese group completed the picture naming task only, while the early Spanish−Chinese bilingual group was requested to first complete the picture naming task, then the DELE and EI tests to assess their proficiency in Spanish and Chinese.

4.2.2.1 Tasks and Stimuli

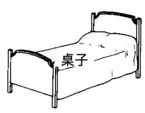
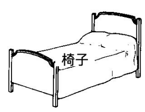
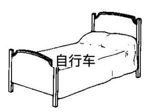
Picture naming task

Twenty-one black and white line drawings were obtained from Severens' picture database (Severens et al., 2005). The picture stimuli were selected based on two criteria: pictures should have easily recognizable features and represent concrete concepts with exact Chinese names. Each picture was associated with four distractor words, manipulated according to their classifier congruency (classifier-congruent vs. classifier-incongruent) and semantic relatedness type (semantically related vs. semantically unrelated) with target pictures. This yielded a total of 84 tokens, each consisting of one target picture and one distractor (see Appendix 4.B). The visual complexity of distractors ($F(3,80) = 0.466$, $p = 0.707$) and their stroke counts ($F(3,80) = 0.695$, $p = 0.558$) were balanced across four conditions. Additionally, the frequency of distractors, $F(3,80) = 0.228$, $p =$

0.877, based on the BCC corpus (Xun et al., 2016), was controlled to ensure uniformity across conditions. Distractors and target pictures were designed to avoid any phonological and orthographic similarities.

The experiment followed a fully factorial within-subjects design with two primary factors: classifier congruency (C) and semantic relatedness (S), resulting in a 2 by 2 design. The factor classifier congruency was manipulated based on whether the classifiers of targets and distractors were congruent (C+) or incongruent (C−). Similarly, the factor semantic relatedness distinguished between semantically related (S+) and semantically unrelated (S−) conditions, depending on whether or not targets and distractors belonged to the same semantic category. This design resulted in four experimental conditions for each target picture: C+S+, C+S−, C−S+, and C−S− (see Table 4.2.1).

Table 4.2.1. A sample set of target pictures paired with distractors for the picture naming task.

	Condition			
Target picture **BED** [床/chuang2/]	C+S+	C+S−	C−S+	C−S−
Target classifier 张 /zhang1/				
Distractors	table /zhuo1zi0/	photo /zhao4pian4/	chair /yi3zi0/	bicycle /zi4xing2che1/
Classifier of distractors	张 /zhang1/	张 /zhang1/	把 /ba3/	辆 /liang4/

Chinese Elicited Imitation (EI) test

The EI test of Chinese was programmed in E-prime 2.0 (Schneider et al., 2002), with identical instructions and stimuli used in Yan et al. (2020). This task aimed to assess Chinese as a second/foreign language (L2/FL) learners' general language proficiency, vocabulary, and grammatical knowledge (Yan et al., 2020). It consists of three sets of stimulus sentences, with each set including 24 sentences (k = 24*3 = 72). These three sets are targeted as follows: the first targets the beginning level, and the second and third sets target the intermediate level and advanced level, respectively. Each sentence consists of key vocabulary and/or grammatical knowledge that can be used for diagnostic assessment (Yan et al., 2020). For example, in (1), (2), and (3), the underlined words in the sentences are key vocabulary, and words in bold are target grammar. The scoring criteria are as follows: no points will be marked for minimal responses, such as complete silence, only function words, or one-word repetition; one point will be awarded for inadequate answers that demonstrate some understanding of the stimulus, such as half-sentence repetition with major grammatical errors, or few content words repetition but changing the main idea of the stimulus; two points will be marked for half repetition that includes more than half of sentences with more than one phrase containing grammatical mistakes; three points for repetition with minor deviation, repetitions with minor changes, or errors only; four points for exact repetition or appropriate paraphrased answers. The total score is max. 288 points.

1. 你是老师**还是**学生? (beginning level)
   Are you a teacher **or** a student?
2. 他去**过**上海很多次. (intermediate level)
   He went (**past tense**) to Shanghai a lot of times.
3. 那条裤子要**多少**钱? (advanced level)
   How much do those pants cost?

Spanish DELE proficiency test (short version)

The short version of the Spanish DELE proficiency test was developed based on the reading test of the traditional *Diplomas de Español como Lengua Extranjera* (*DELE*, Diplomas of Spanish as a Foreign language) proficiency test. The traditional test aimed to assess learners' general language proficiency in Spanish as a foreign language (FL) (Instituto Cervantes, 2010). The short version is designed as a multiple-choice and cloze test with identical test targets. It comprises 30 multiple-choice questions and a long cloze test (20 questions), with each question carrying one point, resulting in a total score of 50. Three distinct levels were distinguished based on the scores: beginning level (0–29), intermediate level (30–39), and advanced level (40–50).

4.2.3 Procedure

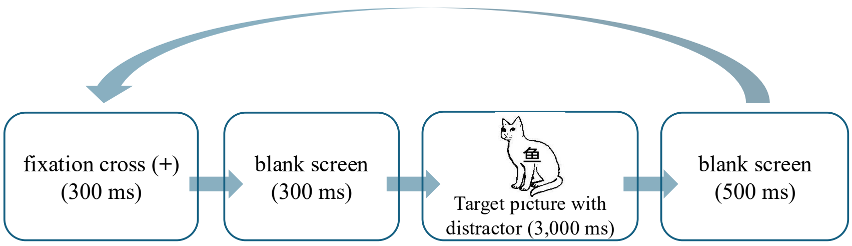
4.2.3.1 Picture naming task

All participants first completed the picture naming task. This task was programmed in E-prime 2.0 (Schneider et al., 2002) with a by-subject order design. This design was achieved using the Windows program Mix (Van Casteren & Davis, 2006) to generate a pseudo-random trial order, counterbalancing the effect of order. Thus, trials from the same condition or featuring the same target picture were not allowed to be presented consecutively, and no more than two consecutive trials could involve the same target classifier. As a result, the order of trials was randomized across participants.

The Mandarin Chinese and early Spanish−Chinese groups followed the same experimental procedure. Consistent with the traditional PWI paradigm, the experiment involved three key phases: familiarization, practice, and main experimental session. During the familiarization session, participants learned the exact names of 21 target pictures, each displayed on the screen for 3,000 ms along with its exact name underneath. Next, in the practice session, participants named target pictures using a bare noun while a meaningless string ("XX") appeared at the center of the screen. Each picture was displayed for 3,000 ms, and any unexpected responses were immediately corrected after all 21 pictures had been presented.

In the experimental session, each target picture randomly appeared four times, once with each of the four distractors, totaling 84 trials. In each trial, one or two identical pictures were presented simultaneously for 3,000 ms on the screen. Participants were instructed to name the target pictures quickly and accurately with Chinese noun phrases (e.g., *一只猫*, /yi1zhi1 mao1/ [one + classifier-zhi1 + cat], "one cat") while ignoring distractor words. Each trial followed a standard sequence (see Figure 4.2.1), i.e., a fixation cross at the center of the screen for 300 ms, followed by a blank screen for 300 ms. Subsequently, the target picture paired with a distractor was presented for 3,000 ms, followed by a blank screen for 500 ms. Vocal responses and EEG data were recorded during this session.

Figure 4.2.1. Illustration of the trial sequence in the experimental session.



4.2.3.2 EEG recordings

EEG data were collected with Ag/AgCl active electrodes via the BrainVision Recorder software by Brain Products GmbH. We used an actiCAP electrode cap (Easycap GmbH) with a BrainAmp standard amplifier (Brain Products GmbH) following a standard 32-electrode 10/20 montage (see Appendix 4.C). Of these 32 electrodes, FT9 was placed below the left eye to monitor the horizontal eye movement (HEOG).

Similarly, FT10 was placed at the outer canthus of the right eye to monitor vertical electrooculogram (VEOG). Additionally, TP9 and TP10 were placed on the left mastoid (M1) and right mastoid (M2), which were later reused as re-referencing channels. Thus, data were measured at twenty-eight channel locations at a sampling rate of 500 Hz. In addition, two external electrodes were plugged into the cap to use as a ground electrode (GND) and an online reference electrode (FCz/REF; see Appendix 4.C). The impedance was kept below 10 kΩ for all electrodes.

4.2.3.3 Chinese Elicited Imitation (EI) test

The second task was the Chinese EI test, which was only tested with the early Spanish−Chinese bilingual speakers to assess their Chinese proficiency. This task was programmed in E-prime with a by-participant design, with all sentences randomized across participants. Following instructions in Yan et al. (2020), procedures for each trial were as follows: first, each stimulus sentence was played to each participant only once; next, a blank screen with 2 seconds of silence was presented to avoid rote memory; third, a 0.5-second ringtone prompted the start of repetition, followed by participants' repetition of the sentences in 20 seconds. The voice response was recorded automatically.

4.2.3.4 Spanish DELE proficiency test

The third task was the Spanish DELE proficiency test, which was only taken by the early Spanish−Chinese bilingual speakers. This was done to evaluate their Spanish language proficiency. This task was completed in printed form, and participants were requested to answer all questions in 10 minutes. All answer sheets were collected for further calculations.

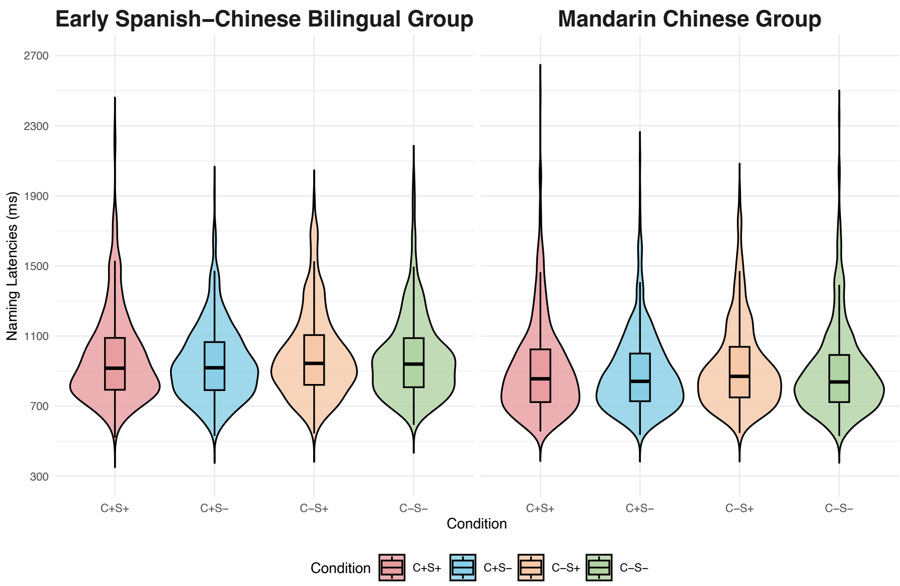
4.3 Results

4.3.1 Behavioral results

All recordings in two groups were first separately preprocessed using Praat (Boersma & Weenink, 2019), and naming latencies were then calculated and extracted for subsequent analysis (see Figure 4.3.1). Trials with incorrect, missed, or revised responses, and those marked by hesitation, were excluded from further analysis. Next, we analyzed the behavioral data for two groups separately using a generalized linear mixed model (GLMM) with a *gamma* distribution. This analysis was conducted via the *glmer* function from the *lme4* package (Bates et al., 2015) in Rstudio version 4.2.2 (*R Core Team*). To avoid the risk of inflating the Type I error rate, a top-down model selection approach was applied in the analysis of the behavioral data, starting with the theoretical maximum model (Barr, 2013). In order to fit the model to our data, the theoretical maximum model incorporated *classifier congruency* and *semantic relatedness* as fixed factors, with *item* and *participant* included as random factors.

For both groups, the full model structure remained consistent, incorporating *classifier congruency (congruent vs. incongruent)* and *semantic relatedness (related vs. unrelated)* as dummy-coded (1 vs. 0) fixed factors, random intercepts for participants and target items, by-participant random slopes for classifier congruency and semantic relatedness, as well as their correlation within the participant factor.

Figure 4.3.1. Mean naming latencies (ms) for each condition in the Mandarin Chinese group and the early Spanish−Chinese bilingual group.



4.3.1.1 Mandarin Chinese group

In the Mandarin Chinese group, a total of 2,520 trials were generated from the 21 picture stimuli. During preprocessing, 8.01% of the trials were excluded, and 0.79% were removed due to outliers, which were identified as naming latencies exceeding three standard deviations (SDs) from the mean response time of participants. This yielded 2,298 trials in total for subsequent statistical analysis. Considering our experiment design and model fit checks by plotting the model residuals against predicted values, we confirmed the full model was the best-fit model. Participants in the Mandarin Chinese group were significantly faster in classifier-congruent conditions with $\beta = -13.99$, $SE = 6.43$, $t = -2.18$, and $p = 0.03$ compared to the incongruent conditions. Moreover, participants responded significantly slower to semantically related conditions compared to semantically unrelated conditions, with $\beta = 22.71$, $SE = 6.09$, $t = 3.73$, and $p < .001$ (see Table 4.3.1).

Table 4.3.1. GLMM of best fit for naming latencies in the Mandarin Chinese group (n = 30), with classifier congruency and semantic relatedness as two predictors, including estimates, confidence intervals, and *p*-values.

Formula: Naming latency ~ Classifier congruency (congruent vs. incongruent) + Semantic relatedness (related vs. unrelated) + (Classifier congruency + Semantic relatedness | participant) + (1 | item)

Predictors	**RTs**			
	Estimate	*95% CI*	*t-value*	*Pr(>|z|)*
(Intercept)	933.484	910.903 – 956.065	81.068	<0.001***
Classifier [Congruent]	−13.988	−26.596 – −1.380	−2.176	0.030*
Semantic [Related]	22.713	10.779 – 34.646	3.732	<0.001***
Random Effects				
σ^2	0.04			
$\tau_{00\ \text{Participant}}$	4845.50			
$\tau_{00\ \text{Item}}$	744.10			
$\tau_{11\ \text{participant. classifier-congruent}}$	305.89			
$\tau_{11\ \text{participant. semantic-related}}$	20.81			
$\rho_{01\ \text{participant. classifier-congruent}}$	− 0.19			
$\rho_{01\ \text{participant. semantic-related}}$	− 0.16			
ICC	1.00			
$N_{\text{participant}}$	30			
N_{Item}	21			
Observations	2298			
Marginal R^2/Conditional R^2	0.032/1.000			

4.3.1.2 Early Spanish−Chinese bilingual group

In the bilingual group, the 21 picture stimuli resulted in a total of 2,436 trials. Of these, 9.11% were removed during the preprocessing stage, and 0.49% were excluded due to outliers. Thus, a total of 2,202 trials were included for subsequent analysis. The best-fit model was generated as follows: first, the full model was constructed; second, the correlation of classifier congruency and semantic relatedness for the participant factor was removed in the case of singular fit. As a result, the best model was fitted, including classifier congruency and semantic relatedness as fixed effects, random intercepts for participants and target items, and by-participant random slopes for classifier congruency and semantic relatedness. The best-fit model demonstrated significantly shorter naming latencies in classifier congruent trials, with $\beta = -24.42$, $SE = 8.37$, $t = -2.92$, and $p = 0.004$ compared to incongruent trials. Additionally, early Spanish−Chinese bilinguals showed significantly longer naming latencies in semantically related than in semantically unrelated trials, with $\beta = 22.19$, $SE = 9.07$, $t = 2.45$, and $p = 0.014$ (see Table 4.3.2). Additionally, the average naming latencies for Mandarin Chinese and early Spanish−Chinese bilingual groups are presented in Table 4.3.3.

Table 4.3.2. GLMM of best fit for naming latencies in the early Spanish−Chinese bilingual group (n = 29), with classifier congruency and semantic relatedness as two predictors, including estimates, confidence intervals, and *p*-values.

Formula: Naming latency ~ Classifier congruency (congruent vs. incongruent) + Semantic relatedness (related vs. unrelated) + (Classifier congruency| participant) + (Semantic relatedness | participant) +(1 | item)

Predictors	**RTs**			
	Estimate	*95% CI*	*t-value*	*Pr(>|z|)*
(Intercept)	994.597	960.662 – 1028.533	57.476	<0.001***
Classifier [Congruent]	−24.418	−40.836 – −8.000	−2.917	0.004**
Semantic [Related]	22.185	4.407 – 39.963	2.447	0.014*
Random Effects				
σ^2	0.05			
τ_{00} Participant. semantic-related	2901.42			
τ_{00} Participant. classifier-congruent	313.74			
τ_{00} Item	1319.24			
τ_{11} participant. semantic-related	886.43			
τ_{11} participant. classifier-congruency	825.88			
ρ_{01} participant. semantic-related	−0.09			
ρ_{01} participant. classifier-congruent	−0.78			
ICC	1.00			
N participant	29			

N_{Item}	21
Observations	2202
Marginal R^2/Conditional R^2	0.057/1.000

Table 4.3.3. Mean picture naming latencies by conditions in Mandarin Chinese and early Spanish−Chinese bilingual groups.

Conditions	Naming latencies (ms)			
	Mandarin Chinese group		Spanish−Chinese group	
	Mean	SD	Mean	SD
Classifier-congruent/Semantically Related (C+S+)	911	264	978	262
Classifier-congruent/Semantically Unrelated (C+S−)	888	223	945	212
Classifier-incongruent/Semantically Related (C−S+)	925	246	992	247
Classifier-incongruent/Semantically Unrelated (C−S−)	899	252	980	235

4.3.2 Chinese Elicited Imitation (EI) test results

The Chinese EL scores were calculated following the original scoring criteria. Almost all participants in the early Spanish−Chinese bilingual group accurately repeated or paraphrased sentences without altering their meanings (M_{score} = 283.41, SD_{score} = 3.01). Overall, all participants were classified within the advanced proficiency level.

4.3.3 Spanish DELE proficiency test (short version) results

The Spanish DELE proficiency scores were calculated by summing the correct answers from participants' answer sheets. Nearly all particip-ants achieved scores within the advanced range (M_{score} = 45.73, SD_{score} = 2.95), indicating an advanced proficiency level in Spanish. Note that scores of either the Chinese EL test or the Spanish DELE test were not considered an absolute measure but rather an indicator of language proficiency.

4.3.4 EEG data results

4.3.4.1 EEG data analysis

EEG data from two groups were pre-processed separately in Brain-Vision Analyzer 2.1 (Brain Products GmbH) following standardized procedures. First, signals were visually inspected for quality. Next, recordings were re-referenced to the average of the left and right mastoid electrodes (M1 and M2). The data were then filtered using a high-pass filter of 0.1 Hz and a low-pass filter of 30 Hz. Ocular independent component analysis (ICA) with VEOG and HEOG as parameters was used to correct blink activity. Artifact rejection was subsequently performed to minimize noise. Finally, the signals were segmented to include only correct trials, generating epochs around picture onsets to analyze the voltage amplitude of ERP components of interest. Each epoch was defined as the interval spanning 200 ms before and 1,200 ms after trial onset, determined based on the average naming latency duration. Baseline correction was applied to the 200 ms prior to picture onset until picture onset, after which the epochs were extracted for statistical analysis. Before the statistical analysis, permutation tests were computed using the permutes package (Voeten, 2019) for each group to tentatively detect the locus of the effect of *classifier congruency* and *semantic relatedness* on voltage amplitudes. Specifically, we calculate F-values across all electrodes and the entire time window between 200 ms and 1,200 ms with respect to stimulus onset. Next, we performed a single-trial linear mixed model (LMM) method with *lmer()* function on each group separately for statistical analysis.

4.3.4.2 Mandarin Chinese group (EEG results)

Results of classifier congruency. The permutation test results for the Mandarin Chinese group revealed potential modulatory effects of classifier congruency (congruent vs. incongruent) in electrodes Pz, P8, P7, P4, P3, Oz, O2, O1, Fz, FC2, FC1, F8, F7, F4, F3, Cz, CP5, CP2, and CP1 in 400 ms – 500 ms time windows post-stimulus onset (see the left panel in Figure 4.3.2). Further, we did a visual inspection of these electrodes and observed a negative-going wave elicited by classifier-incongruent conditions between 400 ms and 500 ms, consistent with the topographic distribution of N400 (Kutas & Hillyard, 1980; Chwilla et al., 1995). Descriptively speaking, we observe the smaller amplitude for the classifier-incongruent condition with M = 4.61 (SD = 10.85) compared to the classifier-congruent condition with M = 5.40 (SD = 10.73) in the 400 – 500 ms time window. To do further amplitude analyses, we grouped the electrodes based on their location, which are left fronto–central, left centro–parietal, right fronto–central, right centro–parietal, and central regions. Furthermore, *location* of grouped electrodes was included as a covariate for amplitude statistical analyses. Additionally, although visual inspection suggests the presence of an N2 component around 200 ms, its underlying cause in this study remains unclear and falls outside the scope of the current study. Therefore, we do not further interpret this component.

The best-fitting model for voltage amplitudes included *classifier congruency (congruent vs. incongruent)* and *semantic relatedness (related vs. unrelated)* as dummy-coded (0 vs. 1) main effects, with *location* of grouped electrodes as a covariate. *Participant* and *item* were included as random effects, with a by-participant random slope for the effects of *classifier congruency* and *semantic relatedness* (see Table 4.3.4 in Appendix 4.D). Voltage amplitudes were significantly more negative for classifier-incongruent compared to congruent conditions with $\beta = -0.74$, $SE = 0.241$, $t = -3.064$, $p = 0.0046$[6].

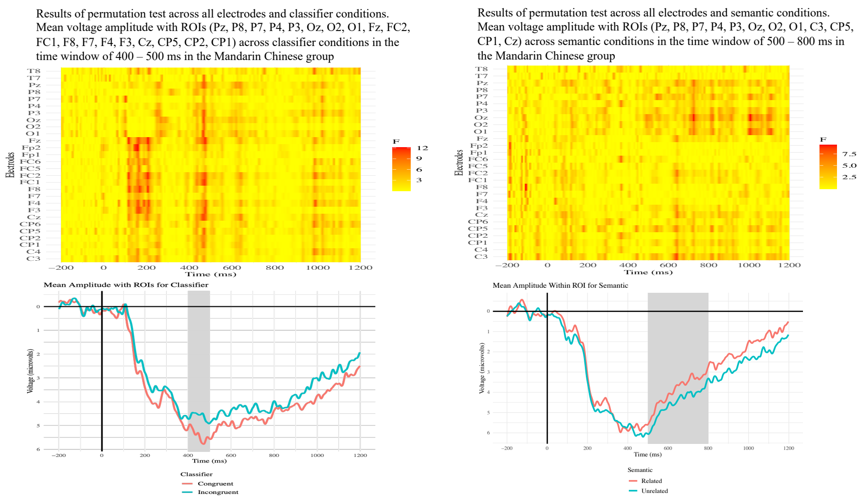
[6] Note: The p-values in the main text were computed using Satterthwaite's method for better accuracy in mixed-effects models. These values may differ slightly from those

Results of semantic relatedness. The permutation test showed potential modulatory effects of semantic relatedness (related vs. unrelated) with electrodes Pz, P8, P7, P4, P3, Oz, O2, O1, C3, CP5, CP1, and Cz within the 500 – 800 ms time window (see the right panel in Figure 4.3.2). Visual inspection illustrated a positive-going wave elicited by semantically unrelated conditions between 500 ms and 800 ms, which is aligned with the topographic distribution of the typical P600 (Osterhout & Holcomb, 1992) and the "semantic P600" (see Kuperberg et al., 2003; Van Herten et al., 2005; Bornkessel-Schlesewsky & Schlesewsky, 2008, for a review). Specifically, the amplitude for semantically unrelated conditions (M = 4.51, SD = 11.08) is larger than for semantically related conditions with M = 3.89 (SD = 11.39) in the 500 – 800 ms time window. Electrodes within ROIs were grouped based on their location, that is, left centro–parietal, right centro–parietal, and central regions. Further, *location* of grouped electrodes was a covariate for amplitude statistical analyses. We found no indication of an N400 effect prior to the 600 ms based on the permutation test and visual inspections.

The best-fit model for our data was with *classifier congruency* and *semantic relatedness* as fixed effects, and *participant* and *item* as random effects. *Location* was included as a covariate, with a by-participant random slope for the effects of *classifier congruency* and *semantic relatedness* (see Table 4.3.5 in Appendix 4.D). Semantically unrelated conditions elicited a more positive amplitude than semantically related conditions with $\beta = 0.68$, $SE = 0.24$, $t = 2.81$, $p = 0.0086$.

presented in Table 4.3.4, 4.3.5, 4.3.6 and 4.3.7 (Appendix 4.D and 4.E), which were derived using Wald tests with *tab_model* function.

Figure 4.3.2. Results of the permutation test for classifier congruency (left panel) and semantic relatedness (right panel) across all electrodes from −200 ms to 1,200 ms, showing F-values and corresponding mean voltage amplitudes within ROIs and time windows in the Mandarin Chinese group.



4.3.4.3 Early Spanish−Chinese bilingual group (EEG results)

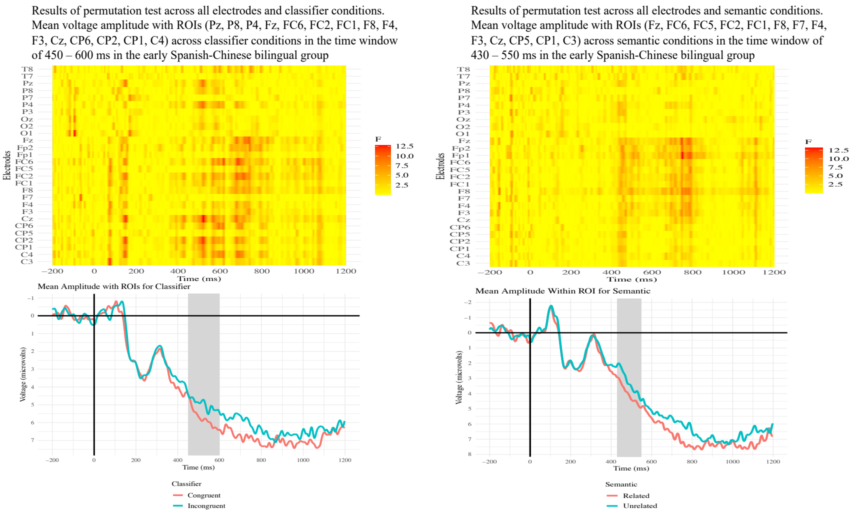
Results of classifier congruency. By following the same procedure as in the Mandarin Chinese group, we found a potential modulation effect of classifier congruency with electrodes Pz, P8, P4, Fz, FC6, FC2, FC1, F8, F4, F3, Cz, CP6, CP2, CP1, and C4, showing a negative-going wave in classifier-incongruent condition in the 450 – 600 ms time window (see left panel in Figure 4.3.3). This is consistent with the potential N400 effect found in the Mandarin Chinese group. Further, these electrodes were grouped into five ROIs: left fronto–central, left centro–parietal, right fronto–central, right centro–parietal, and central regions. In general, classifier-incongruent conditions (M = 4.99, SD = 12.2) showed a smaller amplitude than classifier-congruent conditions with M = 5.76 (SD = 12.0).

The best-fitting model incorporated *classifier congruency* and *semantic relatedness* as fixed effects, with *participant* and *item* as random effects.

Location was included as a covariate, and a by-participant random slope was specified for *classifier congruency* and *semantic relatedness* (see Table 4.3.6 in Appendix 4.E). The result demonstrated that classifier-incongruent conditions elicited a more negative amplitude than classifier-congruent conditions with $\beta = -0.77$, $SE = 0.33$, $t = -2.38$, $p = 0.0242$.

Results of semantic relatedness. Following the same steps, a potential modulation effect of semantic relatedness with a negative-going wave in the semantically unrelated condition was detected with electrodes Fz, FC6, FC5, FC2, FC1, F8, F7, F4, F3, Cz, CP5, CP1, C3 in the 430 – 550 ms time window (see the right panel in Figure 4.3.3). This effect aligned with the topographic distribution of the N400 effect (Kutas & Hillyard, 1980; Chwilla et al., 1995). The descriptive result showed that the amplitude in semantically unrelated conditions ($M = 3.26$, $SD = 11.6$) is smaller than in semantically related conditions ($M = 4.06$, $SD = 12.2$). For further statistical analysis, these electrodes were grouped into four ROIs, including left fronto–central, right fronto–central, left centro–parietal, and central regions. By fitting the *classifier congruency* and *semantic relatedness* as fixed factors, *participant* and *item* as random factors, *location* as a covariate, as well as by-participant random slopes for *classifier congruency* and *semantic relatedness*, the best model was generated. The result demonstrated that semantically unrelated conditions elicited a more negative amplitude than semantically related conditions with $\beta = -0.89$, $SE = 0.39$, $t = -2.30$, $p = 0.0287$ (see Table 4.3.7 in Appendix 4.E).

Figure 4.3.3. Results of the permutation test for classifier congruency (left panel) and semantic relatedness (right panel) across all electrodes from −200 ms to 1,200 ms, showing F-values and corresponding mean voltage amplitudes within ROIs for time windows of interest in the early Spanish–Chinese bilingual group.



4.4 Discussion

In this study, we explored the processing of classifier congruency and semantic relatedness in Mandarin Chinese NPs among early Spanish–Chinese bilinguals and Mandarin Chinese speakers using the PWI paradigm. We manipulated classifier congruency and semantic relatedness between targets and distractors by categorizing them as either classifier-congruent or classifier-incongruent and as either semantically related or unrelated. This allowed us to compare naming latencies between classifier-congruent and classifier-incongruent NPs to assess the behavioral representation of the classifier congruency effect. Meanwhile, we analyzed amplitude differences between these two conditions to detect potential ERP components (e.g., N400 or P600) to determine whether the classifier congruency effect was present electrophysiologically. We made a similar comparison between semantically related and semantically unrelated conditions to identify the semantic interference effect behaviorally and electrophysiologically.

Classifier congruency effect

In line with our expectations, we observed a significant classifier congruency effect behaviorally in both groups. Specifically, in both groups, behavioral results showed that participants' naming latencies were shorter when naming target pictures that were classifier-congruent with distractor words, compared to classifier-incongruent conditions. In other words, the classifier congruency effect was identified behaviorally in naming classifier-congruent targets and distractors. This aligns with findings in NP production observed in previous studies by Wang et al. (2006), Huang and Schiller (2021), Wang and Schiller (submitted) and Wang et al. (submitted), which reported longer naming latencies in classifier-incongruent conditions compared with congruent conditions in monolingual Chinese speakers' NP production.

Notably, the classifier congruency effect was not observed in the behavioral results during bare nouns production in Wang et al. (2019) and Wang et al. (2006). The main reason for the absence of such an effect in these studies is presumably that bare noun naming does not require the production of classifiers in participants' responses. According to the LRM speech production model (Levelt et al., 1999), when naming a picture, the conceptual and lexical representation of to-be-produced words need to be encoded first. Next, syntactic features, such as classifier features in Chinese, are automatically activated and eventually selected. Finally, the to-be-produced word is competitively selected and produced. In bare noun naming tasks, the classifier feature is not needed for production, and as a result, it is not selected. Therefore, no classifier congruency effect is observed in bare noun naming. In our study, the classifier is required for producing a Chinese NP, where it is automatically activated and eventually selected. As a result, the *classifier congruency effect* was observed in both groups.

Meanwhile, we observed that the classifier congruency effect was accompanied by a more negative amplitude elicited by classifier-incongruent conditions compared to congruent conditions in both groups. In the Mandarin Chinese group, negative amplitudes were observed within the 400 – 500 ms time window post-stimulus onset, with the largest effects observed at the fronto–central, central, and centro–parietal electrodes. Similarly, in the early Spanish–Chinese bilingual group, negative waves were maximal at the fronto–central, central, and centro–parietal regions within the 450 – 600 ms time window post-stimulus onset. This time window and regions closely align with the typical N400 effect (Kutas & Hillyard, 1980; Chwilla et al., 1995). Crucially, in both groups, amplitudes were significantly more negative in classifier-incongruent conditions compared to classifier-congruent conditions. This aligns with the electrophysiological result in NP production reported by Huang and Schiller (2021) and the bare noun production results reported by Wang et al. (2019). Both of these studies observed significantly more negative ERP waves elicited by classifier-incongruent compared to congruent conditions within the time windows of 370 – 430 ms or 385 – 585 ms in fronto–central and centro–parietal, respectively, reflecting an N400-like effect. Additionally, the more negative wave in classifier-incongruent conditions is consistent with previous EEG studies on classifier congruency (Chou et al., 2014; Zhang et al., 2012; Zhou et al., 2010). Given the consistency in time windows and electrode regions with previous studies, we also identified an N400-like effect for the classifier congruency effect in both the Mandarin Chinese and early Spanish–Chinese bilingual groups.

Semantic interference effect

On the other hand, behavioral results from semantic relatedness manipulation (i.e., semantically related vs. semantically unrelated) in both groups are in accordance with the well-established *semantic interference effect*. This is evidenced by longer naming latencies were observed in both groups when targets were named in the presence of distractors from the same semantic category (e.g., Rosinski, 1977; La Heij, 1988; Glaser, 1992;

Schriefers et al., 1990). Similar findings were reported by Wang et al. (2019), Huang and Schiller (2021), Wang and Schiller (submitted) and Wang et al. (submitted) with monolingual Chinese speakers, as well as Dutch late learners of Mandarin Chinese (L2 learners) in Wang and Schiller (submitted), where monolingual Chinese speakers and L2 learners exhibited prolonged naming latencies in semantically related targets and distractors in NP and bare noun production. The rationale is that when naming targets that are semantically related to the simultaneously presented distractors, the lemmas of both the target and distractor, along with their associated lexical nodes, are activated simultaneously. The lemmas of semantically related distractors are activated from the same semantic category as the target, leading to an increase in competition in lexical selection and making selection more difficult, thereby prolonging the lexical selection process (Roelofs, 1992, 1993; Levelt et al., 1999). This view was also used to explain the observation of semantic interference in other studies (Belke et al., 2005; Bloem & La Heij, 2003; Roelofs, 2003). The early Spanish–Chinese bilinguals, raised in heritage Chinese households and Spanish society, demonstrated advanced proficiency in both Spanish and Chinese (see Sections 4.3.2 & 4.3.3), which facilitated the observation of the semantic interference effect.

Along with the significant semantic interference effect observed in the behavioral data, we also identified two ERP components associated with this effect in the Mandarin Chinese group and the early Spanish–Chinese bilingual group, respectively, elicited by semantically unrelated conditions against related conditions. Specifically, in the Mandarin Chinese group, semantically unrelated conditions elicited a positive wave within the 500 – 800 ms time window after stimulus onset at centro–parietal regions. This seems to align with the "semantic P600" or centro–parietal P600 effect, which typically has a centro–parietal scalp distribution within a 500 – 800 ms time window (see Kuperberg et al., 2003; Van Herten et al., 2005; Bornkessel-Schlesewsky & Schlesewsky, 2008). In the early Spanish–Chinese bilingual group, we found a negative wave that peaked around 400 ms within the 430 – 550 ms time window elicited by semantically unrelated conditions in the fronto–central and centro–parietal regions. These results largely overlap with the typical N400 effect (Kutas & Hillyard, 1980; Chwilla et al., 1995) and are consistent with the results by Huang and Schiller (2021) and Wang et al. (2019), where they reported a more negative ERP wave around 400 ms evoked by the presence of target pictures with semantically unrelated distractors, reflecting an N400-like effect.

There appears to be a discrepancy in the ERP components observed in the semantic interference effect between the Mandarin Chinese group and the early Spanish–Chinese bilingual group. Specifically, the Mandarin Chinese group showed a semantic P600-like effect, while the early Spanish–Chinese group exhibited an N400-like effect. To address this discrepancy, it is important to consider the characteristics of the N400 and P600 components. As discussed in section 4.1.3, the N400 effect is typically semantically driven, reflecting lexico-semantic integration and lexical co-activation processes (Lau et al., 2008; Kutas & Federmeier, 2011; Chen et al., 2017; Leckey & Federmeier, 2019). In contrast, the P600 effect is typically syntactically driven, reflecting the syntactic integration process (Friederici et al., 1993; Hagoort et al., 1993), while the "semantic P600" effect has been proposed and identified as an electrophysiological response to "integration difficulties" in semantic processing of integrating an unexpected entity into the discourse context (see Kuperberg et al., 2003; Van Herten et al., 2005; Bornkessel-Schlesewsky & Schlesewsky, 2008, for a review), which can be interpreted as semantically driven within syntactic contexts. For the early Spanish–Chinese bilingual group, especially the fact that they have high proficiency in both languages, it is plausible that the semantic integration process is quicker and automatic, which would trigger the N400-like effect as a marker of semantic processing difficulty in semantically unrelated conditions. For the Mandarin Chinese group, given that the semantic P600 effect was proposed and linked to difficulties in semantic integration, the more positive wave observed could reflect a semantic P600-like effect.

However, the P600-like effect elicited by semantically unrelated conditions in the current study seems unexpected, particularly when compared to the N400 effect commonly reported in PWI-based ERP studies in Mandarin Chinese and other languages (e.g., Dell'Acqua et al., 2010; Wicha et al., 2003; Zhu et al., 2015), as few studies have explored or documented this effect in semantic interference. Yet, when considering the production of Chinese NPs, such effects are not entirely surprising. In classifier-NPs, the noun has a comparable lexico-syntactic feature marked by the classifier, and its processing is thought to involve both semantic and syntactic information (Wang et al., 2019). As demonstrated above, classifiers encode inherent properties of nouns, and their selection must be determined by and align with the properties of nouns. This implies that the processing of nouns within NPs inherently requires activation and retrieval of lexico-syntactic information, even when producing bare nouns. Evidence for this comes from the P600 effect observed in PWI-based Chinese bare noun naming, which was elicited by the activation of inherent lexico-syntactic features of nouns, suggesting syntactically driven processing of bare nouns (Wang et al., 2024). Given the role of lexico-syntactic features in Chinese NPs, producing a noun within an NP likely automatically requires the activation and retrieval of lexico-syntactic information. Additionally, since Mandarin Chinese speakers typically acquire classifier-noun combinations by the age of four to five (Erbaugh, 1986), and these combinations must be rotely memorized due to the opaque mapping between classifiers and nouns, the connection between the lexico-syntactic feature of classifiers and the noun remains relatively fixed and reliable. This stable association likely contributes to the P600-like effect observed in Mandarin Chinese speakers. Therefore, the observed P600-like effect may presumably reflect the processing of lexico-syntactic information within NPs.

Additionally, the absence of an N400 response to the semantic interference effect in our data could also be attributed to the inclusion of a relatively high proportion (i.e., 19%) of animate nouns (e.g., McRae et al., 2005; also see Wang et al., 2024 for exploring the role of animacy in bare noun production). Evidence from functional neuroimaging studies suggests that distinct brain regions are involved in processing living (animate nouns) and non-living (inanimate nouns) categories (Chao et al., 1999). In the processing of classifier-noun combinations, the animacy information is not immediately used for semantic integration between nouns and classifiers but rather plays a role in the later processing stage, which would reflect in the P600 effect (Zhang et al., 2012). In contrast, studies that have reported an N400 to semantic interference included a lower proportion of animate nouns, such as 12% in Huang and Schiller (2021) and 15% in Wang et al. (2019). It should be noted that Blackford et al. (2012) dissociated the connection between the behavioral semantic interference effect and the electrophysiological N400-like effect. While the N400-like effect may also arise from semantic priming (Kreher et al., 2006; Blackford et al., 2012), it remains unclear whether it directly corresponds to the semantic interference effect observed in behavioral results. Additionally, Costa et al. (2009) found no significant correlation between naming latencies and the ordinal position of pictures from the same categories within the 400 ms range of the ERP component. Thus, further research is needed to clarify the electrophysiological mechanisms underlying semantic interference.

4.5 Conclusion

To conclude, the current study examined the processing of Chinese classifiers in noun phrase production by early Spanish–Chinese bilinguals and Mandarin Chinese speakers through overt picture naming experiments. Using a picture-word interference paradigm, we investigated naming latencies and electrophysiological correlates of Chinese NP production in four conditions, manipulating classifier congruency (congruent vs. incongruent) and semantic relatedness (related vs. unrelated). Behavioral results showed that both Mandarin Chinese speakers and early Spanish–Chinese bilinguals took a significantly longer time to name targets with classifier-incongruent and semantically related distractors relative to classifier-congruent and semantically unrelated conditions. Electrophysiological data revealed significant N400-like effects in response to classifier-incongruent conditions for both groups, with the bilingual group showing N400-like negativity and the Mandarin Chinese group exhibiting significant P600-like positivity in the semantically unrelated relative to semantically related conditions. Moreover, bilinguals exhibited longer naming latencies and a delayed N400-like effect in classifier-incongruent conditions compared to the Mandarin Chinese group. Overall, this study provides behavioral and electrophysiological evidence that early Spanish–Chinese bilinguals exhibit robust classifier congruency effects in Chinese noun phrase production, mirroring the processing patterns of monolingual speakers in previous studies. These findings suggest that early and sustained bilingual exposure enables the acquisition of morphosyntactic features absent in one of bilinguals' first languages (L1s), supporting models of competitive selection in bilingual language production. Future research should explore the generalizability of these effects across different bilingual populations and morphosyntactic domains, further illuminating the mechanisms underlying bilingual language processing.

Acknowledgements

We sincerely thank all participants for their time and commitment to this study. We are also grateful to the research team at the Center for Brain and Cognition, Universitat Pompeu Fabra (UPF), for their valuable assistance throughout the project. Our deepest appreciation goes to Prof. Núria Sebastián Gallés, head of the Speech Acquisition and Perception (SAP) group at UPF, for her kind support and hospitality. We further acknowledge the technical guidance and logistical support provided by Francesc Xavier Mayoral Terron and Silvia Blanch Rabat from the Department of Technology at UPF, whose contributions were instrumental during the experimental setup and data collection phases.

Declaration of conflicting interest

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

This work was partially funded by the China Scholarship Council (CSC) NO. 202107720015. Niels O. Schiller is supported by CityUHK start-up grant no.9380177 from City University of Hong Kong.

Data Availability Statement

The data supporting the findings of this study are openly available in Open Science Framework at https://osf.io/qwe7p/?view_only=a232bf276df54febbb162b46d940790d (view-only link).

Appendix

Appendix 4.A

1. Overview of linguistic profile for the Mandarin Chinese group (N = 30) according to the LEAP-Q (Marian et al., 2007).

Mandarin Chinese speakers	Mean (SD)/Distribution
Number of female/male participants	21/9
Mean age in years (SD; range)	24.36 (3.05; 22-30)
Number of participants pursuing a Bachelor degree (BA)	3
Number of participants pursuing a Master degree (MA)	22
Number of participants pursuing a Doctoral degree (PhD)	5
Number of participants with Spanish proficiency (A1)	4
Number of participants with Spanish proficiency (A2)	7
Number of participants with Spanish proficiency (B1)	3
Daily use of Chinese (frequency)	70.17% (SD = 0.17)
Daily use of English (frequency)	25.6% (SD = 0.16)
Daily use of Spanish (frequency)	4.23% (SD = 0.13)

2. Overview of linguistic profile for the early Spanish–Chinese bilingual group (N = 29) according to the revised BCSP (Olson, 2022).

Spanish-Chinese bilinguals	Mean (SD)/Distribution
Number of female/male participants	24/5
Mean age in years (SD; range)	21 (3.05; 18-29)
Number of participants born in Spain	25
Number of participants born in China	4
Number of participants with a high school diploma	3
Number of participants pursuing a Bachelor's degree (BA)	22
Number of participants pursuing a Master's degree (MA)	4
Daily use of Spanish (frequency)	44.86% (SD = 0.16)
Daily use of Chinese (frequency)	32.56% (SD = 0.17)
Daily use of Catalan (frequency)	12.17% (SD = 0.11)
Daily use of English (frequency)	10.17% (SD = 0.09)
Daily use of French (frequency)	0.21% (SD = 0.002)
Daily use of German (frequency)	0.03% (SD = 0.0003)
Number of participants with Catalan proficiency (C1/B2)	16/11
Number of participants with French proficiency (below B1)	14
Number of participants with German proficiency (A2)	1
Number of participants with Slovak proficiency (A2)	1
Number of participants with Italian proficiency (A1)	1

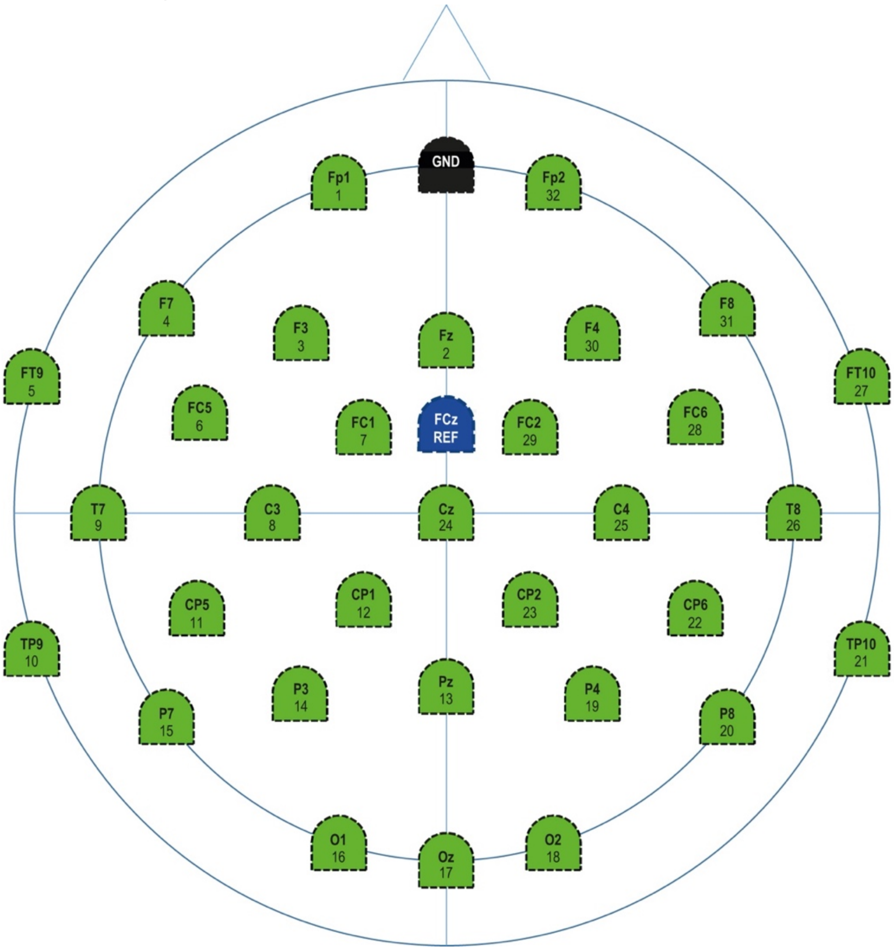
Appendix 4.B Stimuli used in the picture-naming task in Chapter 4.

Target picture	**Classifier**	**Distractor type**			
		Classifier-congruent		Classifier-incongruent	
		Semantically related	Semantically unrelated	Semantically related	Semantically unrelated
腿 (tui3) leg	条 (tiao2)	胳膊 (ge1bo0) arm	毛巾 (mao2jin1) towel	脚 (jiao3) foot	石头 (shi2tou0) stone
马 (ma3) horse	匹 (pi3)	狼 (lang2) wolf	布 (bu4) cloth	牛 (niu2) cow	刀 (dao1) knife
老虎 (lao3hu3) tiger	只 (zhi1)	猴子 (hou2zi0) monkey	船 (chuan2) boat	狮子 (shi1zi0) lion	骨头 (gu3tou2) bone
床 (chuang2) bed	张 (zhang1)	桌子 (zhuo1zi0) table	照片 (zhao4pian4) photo	椅子 (yi3zi0) chair	自行车 (zi4xing2che1) bicycle
耳朵 (er3duo1) ear	只 (zhi1)	眼睛 (yan3jing1) eye	鸽子 (ge1zi0) dove	鼻子 (bi2zi0) nose	棉花 (mian2hua1) cotton
叉子 (cha1zi0) fork	把 (ba3)	勺子 (shao2zi0) spoon	伞 (san3) umbrella	筷子 (kuai4zi0) chopsticks	牛奶 (niu2nai3) milk
围巾 (wei2jin1) scarf	条 (tiao2)	领带 (ling3dai4) tie	绳子 (sheng2zi0) rope	手套 (shou3tao4) glove	飞机 (fei1ji1) airplane
苹果 (ping2guo3) apple	个 (ge4)	桔子 (ju2zi0) mandarin	杯子 (bei1zi0) cup	香蕉 (xiang1jiao1) banana	竹子 (zhu2zi0) bamboo
花生 (hua1sheng1) peanut	粒 (li4)	玉米 (yu4mi3) corn	沙子 (sha1zi0) sand	小麦 (xiao3mai4) wheat	手指 (shou3zhi3) finger
梨 (li2) pear	个 (ge4)	桃 (tao2) peach	盘子 (pan2zi0) plate	葡萄 (pu2tao2) grape	报纸 (bao4zhi3) newspaper
裤子 (ku4zi0) pants	条 (tiao2)	裙子 (qun2zi0) skirt	尾巴 (wei3ba1) tail	衬衫 (chen4shan1) shirt	太阳 (tai4yang2) sun
嘴 (zui3) mouth	张 (zhang1)	脸 (lian3) face	表格 (biao3ge2) form	舌头 (she2tou0) tongue	熊猫 (xiong2mao1) panda
西红柿 (xi1hong2shi4) tomato	个 (ge4)	土豆 (tu3dou4) potato	书架 (shu1jia4) bookshelf	黄瓜 (huang2gua1) cucumber	云 (yun2) cloud
蛋糕	块	饼干	手表	面包	项链

(dan4gao1) cake	(kuai4)	(bing3gan1) cookie	(shou3biao3) watch	(mian4bao1) bread	(xiang4lian4) necklace
塔 (ta3) tower	座 (zuo4)	寺庙 (si4miao4) temple	雪山 (xue3shan1) snow mountain	楼 (lou2) building	西瓜 (xi1gua1) watermelon
包子 (bao1zi0) bun	个 (ge4)	饺子 (jiao3zi0) dumpling	网球 (wang3qiu2) tennis	面条 (mian4tiao2) noodle	剪刀 (jian3dao1) scissors
猫 (mao1) cat	只 (zhi1)	狗 (gou3) dog	袜子 (wa4zi0) sock	鱼 (yu2) fish	月亮 (yue4liang4) moon
鸭子 (ya1zi0) duck	只 (zhi1)	鸡 (ji1) chicken	手 (shou3) hand	蛇 (she2) snake	树 (shu4) tree
桥 (qiao2) bridge	座 (zuo4)	长城 (chang2cheng2) The great wall	岛 (dao3) island	地道 (di4dao4) tunnel	帽子 (mao4zi0) hat
猪 (zhu1) pig	只 (zhi1)	羊 (yang2) sheep	皮鞋 (pi2xie2) leather shoe	大象 (da4xiang4) elephant	书 (shu1) book
兔子 (tu4zi0) rabbit	只 (zhi1)	老鼠 (lao3shu3) mouse	行李箱 (xing2li3xiang1) suitcase	熊 (xiong2) bear	米 (mi3) rice

Appendix 4.C

Illustration of 32-electrode locations following a standard 32-electrode 10/20 montage.



Appendix 4.D

Table 4.3.4. LMM for Classifier congruency (N400) in the Mandarin Chinese group (n = 30).

Formula: Voltage amplitude ~ Classifier congruency (congruent vs. incongruent) + Semantic relatedness (related vs. unrelated) + Location (left frontal central vs. left central parietal vs. right frontal central vs. right central parietal vs. central) + (Classifier congruency + Semantic relatedness | participant) +(1 | item)

Predictors	**Voltage amplitude**			
	Estimate	*95% CI*	*t-value*	*Pr(>|z|)*
(Intercept)	5.621	4.385 – 6.858	8.913	<0.001***
Classifier [Incongruent]	−0.739	−1.211 – −0.266	−3.064	0.002**
Semantic [Unrelated]	0.161	−0.331– 0.654	0.643	0.521
Location [left central parietal]	0.930	0.890 – 0.970	45.392	<0.001***
Location [left frontal central]	−2.136	−2.182 – −2.091	−91.567	<0.001***
Location [right central parietal]	0.356	0.314 – 0.398	16.480	<0.001***
Location [right frontal central]	−2.029	−2.075 – −1.984	−86.981	<0.001***
Random Effects				
σ^2	103.07			
$\tau_{00\ \text{Participant}}$	10.70			
$\tau_{00\ \text{Item}}$	0.86			
$\tau_{11\ \text{participant. classifier-incongruent}}$	1.74			
$\tau_{11\ \text{participant. semantic-unrelated}}$	1.89			
$\rho_{01\ \text{participant. classifier-incongruent}}$	0.00			
$\rho_{01\ \text{participant. semantic-unrelated}}$	− 0.24			
ICC	0.10			
$N_{\text{participant}}$	30			
N_{Item}	21			
Observations	2,098,474			
Marginal R^2/Conditional R^2	0.014/0.116			

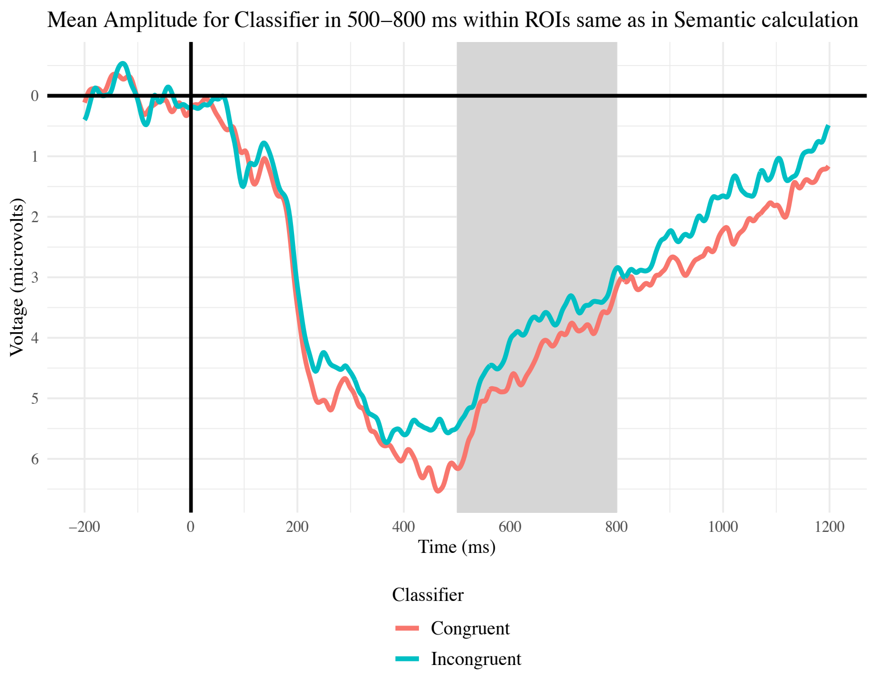
Table 4.3.5. LMM for semantic relatedness (P600) in the Mandarin Chinese group (n = 30).

Formula: Voltage amplitude ~ Classifier congruency (congruent vs. incongruent) + Semantic relatedness (related vs. unrelated) + Location (left central parietal vs. right central parietal vs. central) + (Classifier congruency + Semantic relatedness | participant) +(1 | item)

Predictors	**Voltage amplitude**			
	Estimate	*95% CI*	*t-value*	*Pr(>|z|)*
(Intercept)	4.217	2.945 – 5.490	6.497	<0.001***
Classifier [Incongruent]	−0.474	−0.932 – −0.015	− 2.024	0.043*
Semantic [Unrelated]	0.675	0.204 – 1.146	2.810	0.005**
Location [left central parietal]	0.779	0.753 – 0.804	59.663	<0.001***
Location [right central parietal]	−1.842	−1.871 – −1.812	– 122.211	<0.001***
Random Effects				
σ^2	114.39			
$\tau_{00\ \text{Participant}}$	11.46			
$\tau_{00\ \text{Item}}$	0.83			
$\tau_{11\ \text{participant. classifier-incongruent}}$	1.64			
$\tau_{11\ \text{participant. semantic-unrelated}}$	1.73			
$\rho_{01\ \text{participant. classifier-incongruent}}$	− 0.15			
$\rho_{01\ \text{participant. semantic-unrelated}}$	− 0.48			
ICC	0.09			
$N_{\text{participant}}$	30			
N_{Item}	21			
Observations	4,030,152			
Marginal R^2/Conditional R^2	0.010/0.098			

Note: The classifier congruency is significant (p = 0.043) in the semantic model, however, it is not the classifier congruency effect here, as the voltage amplitude elicited by classifier incongruent trials has a negative-going direction in these ROIs and time windows of 500 – 800 ms. Please see Figure 4.3.4 for classifier congruency in these ROIs and time windows of 500 – 800 ms.

Figure 4.3.4. Mean amplitude for Classifier in 500 – 800 ms within the same ROIs as in Semantic calculations.



Appendix 4.E

Table 4.3.6. LMM for semantic relatedness (N400) in the early Spanish–Chinese bilingual group (n = 29).

Formula: Voltage amplitude ~ Classifier congruency (congruent vs. incongruent) + Semantic relatedness (related vs. unrelated) + Location (left frontal central vs. left central parietal vs. right frontal central vs. central) + (Classifier congruency + Semantic relatedness | participant) +(1 | item)

Predictors	**Voltage amplitude**			
	Estimate	*95% CI*	*t-value*	*Pr(>|z|)*
(Intercept)	3.877	2.194 – 5.560	4.515	<0.001***
Classifier [Incongruent]	−0.500	−1.189 – 0.190	−1.420	0.156
Semantic [Unrelated]	−0.888	−1.643 – −0.132	−2.303	0.021*
Location [left central parietal]	2.801	2.745 – 2.857	98.526	<0.001***
Location [left frontal central]	−0.759	−0.812 – −0.706	−28.134	<0.001***
Location [right frontal central]	0.261	0.209 – 0.314	9.695	<0.001***
Random Effects				
σ^2	122.22			
$\tau_{00\ \text{Participant}}$	19.34			
$\tau_{00\ \text{Item}}$	1.47			
$\tau_{11\ \text{participant. classifier-incongruent}}$	3.58			
$\tau_{11\ \text{participant. semantic-unrelated}}$	4.30			
$\rho_{01\ \text{participant. classifier-incongruent}}$	− 0.03			
$\rho_{01\ \text{participant. semantic-unrelated}}$	− 0.70			
ICC	0.13			
$N_{\text{participant}}$	29			
N_{Item}	21			
Observations	1,638,312			
Marginal R^2/Conditional R^2	0.014/0.141			

Table 4.3.7. LMM for Classifier congruency (N400) in the early Spanish–Chinese bilingual group (n = 29).

Formula: Voltage amplitude ~ Classifier congruency (congruent vs. incongruent) + Semantic relatedness (related vs. unrelated) + Location (left frontal central vs. left central parietal vs. right frontal central vs. right central parietal vs. central) + (Classifier congruency + Semantic relatedness | participant) +(1 | item)

Predictors	**Voltage amplitude**			
	Estimate	*95% CI*	*t-value*	*Pr(>|z|)*
(Intercept)	6.283	4.701 – 7.866	7.781	<0.001***
Classifier [Incongruent]	−0.772	−1.408 – −0.136	−2.379	0.017*
Semantic [Related]	−0.656	−1.336 – 0.023	−1.893	0.058
Location [left central parietal]	2.130	2.066 – 2.195	64.725	<0.001***
Location [left frontal central]	− 2.501	−2.552 – −2.450	−96.108	<0.001***
Location [right central parietal]	1.290	1.249 – 1.331	61.970	<0.001***
Location [right frontal central]	−1.405	−1.447 – −1.362	−64.528	<0.001***
Random Effects				
σ^2	128.42			
$\tau_{00\ \text{Participant}}$	17.14			
$\tau_{00\ \text{Item}}$	1.27			
$\tau_{11\ \text{participant. classifier-incongruent}}$	3.05			
$\tau_{11\ \text{participant. semantic-unrelated}}$	3.48			
$\rho_{01\ \text{participant. classifier-incongruent}}$	− 0.03			
$\rho_{01\ \text{participant. semantic-unrelated}}$	− 0.65			
ICC	0.11			
$N_{\text{participant}}$	29			
N_{Item}	21			
Observations	2,370,960			
Marginal R^2/Conditional R^2	0.017/0.128			

Note: The p-values presented in the *tab_model* output were derived from Wald tests, which may slightly differ from those reported in the main text, as the latter were computed using Satterthwaite's method for better accuracy in mixed-effects models.

Chapter 5

Classifier Assignment Strategies in Mixed Chinese–Spanish Noun Phrases: Insights from a Multilingual Community in Barcelona

---

A version of this chapter is in preparation for submission: Wu, R., Parafita Couto, M. C., & Schiller, N. O. (in preparation). Classifier assignment strategies in mixed Chinese–Spanish noun phrases: insights from a multilingual community in Barcelona.

150 *The Role of Lexico-Syntactic Features in Noun Phrase Production and Comprehension*

Abstract

This study investigates how early Spanish–Chinese bilinguals assign Mandarin Chinese classifiers to Spanish nouns in mixed Chinese–Spanish noun phrases (NPs). Mandarin Chinese compulsorily uses classifiers in NPs, with most of them being selected largely based on the semantic properties of associated nouns. In contrast, Spanish does not have a classifier system. While Chinese employs several dozen classifiers, the general classifier *ge* is used as a default, raising the question of whether bilinguals will similarly default to *ge* when assigning classifiers to Spanish nouns, or whether they will employ other strategies, such as selecting classifiers based on Chinese translation equivalents of the noun. To address these questions, we adopted a multi-method approach, including a director-matcher task, a sentence repetition task, and a judgment task, with thirty early Spanish–Chinese bilinguals, heritage speakers of Chinese residing in Barcelona, Spain, in unilingual Chinese NPs and mixed NPs. Results showed that, in unilingual Chinese NPs, bilinguals reliably used the expected-specific classifiers in both production and comprehension. However, in mixed NPs, while both the general classifier *ge* and expected-specific classifiers were used in production, comprehension was overwhelmingly guided by expected-specific classifiers. These findings flexibly utilize both default and translation equivalent strategies in classifier assignment. This parallels previously observed gender assignment strategies in the same bilingual community, highlighting the role of community characteristics and social networks in shaping code-switching. This is the first study to explore classifier assignment in mixed Chinese–Spanish NPs, offering novel insight into bilingual flexibility and adaptability in navigating multiple linguistic systems. While these findings provide valuable initial evidence, additional research will help determine their applicability across broader populations and contexts.

Keywords: code-switching; mixed noun phrases; Chinese classifiers; Chinese; Spanish; bilingualism

5.1 Introduction

A unique feature of multilingual communication is that speakers fluently alternate or integrate elements from their languages into one single conversation, a practice known as code-switching (CS) (Poplack, 1980). In CS research, mixed noun phrases (NPs) have emerged as a key area of interest, where one element (e.g., determiners or articles) belongs to one language, while other elements (e.g., nouns or adjectives) come from the other. This interest likely stems from the frequent code-switches observed between determiners and the adjective-noun cluster within mixed NPs in naturalistic data (e.g., Parafita Couto & Gullberg, 2019). For instance, Spanish–English bilinguals might produce *los dry walls* ([the$_{\text{PL.MASC}}$ dry walls] “the dry walls”), combining a Spanish determiner (*los*) with an English adjective-noun cluster (*dry walls*) (Parafita Couto & Gullberg, 2019). Research on these mixed determiner-NPs has predominantly focused on grammatical gender assignment in language combinations that are either both gender-marked or consist of one gender-marked and one non-gender-marked language (see Bellamy & Parafita Couto, 2022 for an overview). This has well established the patterns of gender assignment and the corresponding gender assignment strategies used in mixed NPs across various language combinations and bi/multilingual communities (see Wu et al., 2025 for gender assignment in mixed Spanish–Chinese NPs). It is known that exploring the mechanisms of gender assignment in mixed NPs enhances our understanding of how bilingual speakers navigate linguistic structure differences and integrate elements from their languages. However, grammatical gender is only one type of nominal classification system. To develop a more comprehensive understanding of how bilinguals process such systems across languages, it is crucial to expand such research to include language combinations that do not encode grammatical gender but have analogous systems, such as classifiers. One such language combination that deserves attention is mixed Chinese–Spanish classifier-NPs.

Nominal classification systems are a prevalent feature across the world's languages, with grammatical gender (e.g., Spanish) and numeral classifiers (e.g., Mandarin Chinese) representing the most widespread types (Seifart, 2010). In Spanish, every noun is inherently assigned a grammatical gender, either masculine or feminine, with masculine functioning as the default (Harris, 1991; Roca, 1989). Determiners and adjectives within Spanish NPs do not possess inherent gender but instead agree in gender with associated nouns (Harris, 1991). Thus, grammatical gender in Spanish is determined by the properties of nouns. Similarly, Mandarin Chinese features a lexico-syntactic numeral classifier system in which specific morphemes, known as "classifiers", are obligatorily positioned between numerals or demonstratives and nouns within NPs, for example, 一辆汽车, /yi1liang4qi4che2/ [one + classifier-liang4 + car], "one car" (Li & Thompson, 1981). Similar to grammatical gender in Spanish, classifier selection in Chinese largely depends on the particular semantic properties of nouns (Li & Thompson, 1981). Given that both grammatical gender and classifier systems organize nouns into categories based on intrinsic properties, mixed Chinese–Spanish classifier noun phrases present a good context for investigating how bilinguals manage structural differences and integrate elements from both languages during code-switching (see Parafita Couto et al., in press). Accordingly, this study examines classifier assignment patterns and strategies in mixed Chinese–Spanish NPs within an underexplored bilingual community. To our knowledge, limited code-switching research has addressed this phenomenon among early Spanish–Chinese bilinguals in Barcelona. Specifically, we analyze which Chinese classifiers are selected for Spanish nouns and identify the strategies these bilinguals employ in classifier assignment within mixed NPs.

5.2 Research background

5.2.1 Classifiers in Mandarin Chinese and mixed Chinese –Spanish NPs

5.2.1.1 Classifiers in Mandarin Chinese

In Mandarin Chinese (hereafter, "Chinese" refers to Mandarin Chinese), there are a multitude of classifiers employed in a variety of contexts. The 1988 edition of the *Hanyu Liangci Cidian* (*A Dictionary of Chinese Classifiers*, 1988) lists a total of 902 classifiers[7]. These include both nominal classifiers (cf. *名量词 "mingliang ci"* in Chinese), paired with nouns in NPs, and verb classifiers (cf. *动量词 "dongliang ci"* in Chinese), paired with verbs in verb phrases (Zhang, 2007). Considering that this study focuses on NPs, only nominal classifiers (hereafter, "classifier" refers to nominal classifiers) are described in detail.

Erbaugh (1986) notes that approximately 150 classifiers are commonly used in Chinese. The categorization of these classifiers varies across studies, with different scholars proposing overlapping classification schemes. For instance, Chao (1968) delineates six major categories: (1) individual classifiers to classify individual objects based on their shape or other properties (e.g., *棵* /ke1/ "classifier for plants"); (2) group classifiers to group or represent collections of individual objects (e.g., *捆* /kun3/ "bundle"); (3) partition classifiers, which represent portions of objects (e.g., *一截绳子* /yi1jie2sheng2zi0/ "a section of rope"); (4) temporary classifiers, measuring entities by external extent (e.g., *一身雪* /yi1shen1 xue3/ "a bodyful of snow"); (5) standard measure classifiers, i.e., measure weight or length (e.g., *公里* /gong1li3/ "kilometer"); (6) container classi-

---

[7] Earlier research did not clearly differentiate between the terms "classifier" and "measure word", instead, they were often used interchangeably or subsumed under one another. For example, Chao (1968) referred to all Chinese classifiers as "measure words". In contrast, Erbaugh (2006) adopted the term "classifier" but mentioned that the Chinese word for classifier is literally "measure word" (*量词, "liangci" in Chinese*). However, Li and Thompson (1981) used both terms, treating "measure words" as a subset of "classifiers" and stating that any measure word can be a classifier. In this paper, we will not focus on distinguishing these terms but instead consistently use "classifier".

fiers, i.e., container nouns used as measures (e.g., *一瓶水* /yi1ping2shui3/ "a bottle of water"). These categories reflect the diverse functions and semantic criteria underlying classifier selection in Chinese.

Based on similar criteria, Erbaugh (1986) categorized five groups: (1) sortal classifiers, which are based on the properties of entities (e.g., *棵* /ke1/); (2) collective classifiers, i.e., indicate arrangements of entities (e.g., *捆* /kun3/); (3) event classifiers, measuring events (e.g., *场* /chang3/ "performance of a show"); (4) kind classifiers, which refer to the kind or type of entities, e,g., *种* /zhong3/ "type"; (5) measure classifiers, measuring weight or length (e.g., *公里* /gong1li3/). Among these categories, sortal classifiers as described by Erbaugh (1986) and individual classifiers as outlined by Chao (1968) are the most frequently used and tend to show considerable overlap in their classification criteria. These classifiers are typically associated with nouns that are countable and are the preferred choice of monolingual Chinese speakers when referring to unfamiliar nouns or introducing new information (Erbaugh, 2006). Therefore, this study focuses on the use of sortal (individual) classifiers within the Chinese classifier system.

In Chinese, there are approximately 75 sortal classifiers, with each typically associated with 5 to 20 nouns (Erbaugh, 2002). Typically, in Chinese NPs, these classifiers must occur after a numeral (e.g., *yi1* "one", *ban4* "half"), a demonstrative (e.g., *zhe4* "this", *na*4 "that"), and/or quantifiers (e.g., *zheng3* "whole", *ji3* "a few"), but precede the noun, i.e., "numeral/demonstrative/quantifiers + classifier + noun" (Li & Thompson, 1981). The selection of classifiers is largely determined by the semantic properties of nouns, including animacy, shape, function, and size (Myers & Tsay, 2000). For instance, the classifier *tiao* can be shared by both *蛇* /she2/ "snake" and *绳子* /sheng2zi0/ "rope" due to the fact that they both have a long shape. While these semantic features play significant roles in classifier choices within NPs, the relationship between nouns and their corresponding classifiers is not always transparent (Shao, 1993; Tzeng et al., 1991; also see Wang & Schiller, in press, and Qian, in press, for an overview). In many cases, the "proper"[8] classifier for a given noun must be learned through rote memorization. Additionally, it is not always the case that each noun corresponds to a single, fixed classifier; rather, speakers frequently use multiple classifiers with the same noun (Erbaugh, 2006). For example, classifiers include *只* (*zhi1*, "for animals or small objects"), *条* (*tiao2*, "extended, long shape"), *头* (*tou2*, "head"), and *角* (*jiao3*, "horn") were used for the referent "goat" in the Pear Stories film by one single speaker, which was also acceptable (Erbaugh, 2006). Lakoff (1987) argued that classifiers are a form of conceptual category, which reflects the imaginative aspects of the mind. Denny (1976) also suggested that the selection of classifiers for nouns is more likely to emphasize the inherent properties of the objects themselves. Nevertheless, Chinese speakers are likely to share a common sense of which classifier categories are suitable for categorizing objects overall (Allan, 1977).

In addition to specific sortal classifiers, there is an exception to classifiers, a so-called *general* or *default* classifier (Myers & Tsay, 2000). In Chinese, the most frequently used classifier is *ge* (*个* /ge4/), which is gradually becoming a general (Li & Thompson, 1981) and default classifier (Myers & Tsay, 2000). On the one hand, the general classifier *ge* is typically used for nouns that do not fit into any particular or specialized classifier category (Erbaugh, 1986), with approximately 40% of nouns only take this classifier, including unique objects (e.g., "earth", "city"), abstractions (e.g., "dream", "idea"), and many common objects (e.g., "ball") (Erbaugh, 2006). On the other hand, it has a function as a default to substitute for sortal classifiers or to be used for nouns that do not take a sortal classifier (Aikhenvald, 2000). Thus, it possesses considerable grammatical flexibility and can be used in a wide range of contexts, which allows it to replace a more specialized classifier in many cases (Li & Thompson, 1981; Zhang, 2013). For example, 1a and 1b, the proper classifiers for 1a and 1b are *ke*

---

[8] In this article, we consistently use the term "proper" as in Li and Thompson (1981) to refer to the classifier that is most associated with the noun they accompany.

and *tiao*, respectively. However, many monolingual Mandarin Chinese speakers have replaced *ke* and *tiao* with the general classifier *ge*. For these speakers, both the general classifier *ge* and the proper classifiers *ke* and *tiao* are equally acceptable.

1. a. san *ge/ke* xigua
   three $CL_{general}$/$CL_{specific}$ watermelon
   "three watermelons"

   b. san *ge/tiao* jianyi
   three $CL_{general}$/$CL_{specific}$ suggestion
   "three suggestions"

(Zhang, 2013)

However, the alternation of *ge* and other classifiers does not always happen on the basis of the acceptability of grammaticality. In fact, many nouns still require a specific classifier that should be memorized (Li & Thompson, 1981). For example, classifiers in NPs that include kind classifiers (2a) or non-abstract mass nouns (2b) cannot be replaced by *ge*.

2. a. san *zhong* shu
   three $CL_{kind\text{-}specific}$ book
   "three kinds of books"

   b. san *di* you
   three $CL_{specific\text{-}mass}$ oil
   "three drops of oil"

(Zhang, 2013)

In general, the classifier *ge* is the *default* classifier that can replace any other classifiers under the right pragmatic conditions (Myers & Tsay, 2000). The use of the general classifier *ge* as a default is commonly observed in the speech of monolingual Chinese adults (e.g., Erbaugh, 1986; Myers et al., 1999), children (e.g., Erbaugh, 1986; Hu, 1993), and non-native Chinese learners (Polio, 1994). For instance, Myers et al. (1999) found that monolingual Chinese adult speakers used the classifier *ge* approximately 70% of the time in a classifier elicitation study. A similar pattern was found in the storytelling study in Erbaugh (1986), in which the classifier *ge* was used almost 95% of the time. Based on these findings, it is interesting to explore how Spanish–Chinese bilingual speakers use the general classifier *ge* and specific proper classifiers in both Chinese unilingual contexts and Chinese–Spanish bilingual contexts. In other words, it becomes imperative to employ multiple tasks to investigate whether the classifier *ge* is the default when early Spanish–Chinese bilinguals process and produce classifier-NPs in unilingual Chinese and Chinese–Spanish bilingual contexts.

5.2.1.2 Classifiers in mixed Chinese–Spanish NPs

Typically, mixed Chinese–Spanish NPs consist of at least a Chinese numeral, followed by a Chinese classifier, and a Spanish noun (i.e., Chinese numeral + Chinese classifier + Spanish noun, see examples 3a and 3b below with Spanish nouns in italics).

3. a. 一 朵 *flor*
   one CL.duo$_{\text{specific}}$ flower
   "one flower"

   b. 一 辆 *coche*
   one CL.liang$_{\text{specific}}$ car
   "one car"

Given that Chinese has several dozen classifiers that can or cannot be assigned based on the semantic properties of nouns, as well as the exception of the general classifier *ge*, this raises questions: how do Spanish–Chinese bilinguals assign Chinese classifiers to Spanish nouns in mixed Chinese–Spanish NPs? Do they default to the general classifier *ge* for Spanish nouns? Research on grammatical gender, a comparable classification system to the classifier system in Chinese, in many mixed-language contexts has documented several strategies, such as defaulting other-language nouns to masculine gender or assigning gender to other-language nouns based on the gender of Spanish translation equivalents (see section 5.2.2). Considering the parallel noun-based nature of both grammatical gender and classifier systems, this study investigates whether similar strategies are employed in classifier assignment within mixed Chinese–Spanish NPs. To provide some context for this exploration, the following section first reviews gender assignment strategies in mixed NPs and then draws on these findings to hypothesize analogous strategies for classifiers.

5.2.2 Classifier assignment strategies

5.2.2.1 Default strategy

The default strategy was introduced by Poplack et al. (1982) and elaborated in subsequent studies (for an overview, see Bellamy & Parafita Couto, 2022) to explain gender assignment patterns in mixed NPs. In studies of gender assignment, this strategy refers to preferences for defaulting to a gender for most other-language inserted nouns, regardless of the gender of their translation equivalents, semantic or morpho-phonological properties (see Bellamy & Parafita Couto, 2022). For example, Spanish–English bilinguals in Miami default Spanish masculine gender to English nouns and produce *el cookie* "the$_{\text{MASC}}$ cookie", rather than *la cookie* "the$_{\text{FEM}}$ cookie", although the Spanish equivalent *galleta* is feminine (see Valdés Kroff, 2016). Similarly, Chinese includes the general classifier *ge*, which often functions as a default choice, particularly when speakers refer to nouns that do not fit neatly into specific categories or when concepts are unfamiliar (see Myers & Tsay, 2000 and Erbaugh, 2006). This pattern allows us to hypothesize a comparable default strategy for classifier assignment in mixed Chinese–Spanish NPs. Under this approach, Spanish–Chinese bilinguals might regularly use *ge* for Spanish nouns, regardless of whether a more specific proper classifier exists for their Chinese equivalents or based on semantic properties. For example, Spanish–Chinese bilinguals would produce 一个 *árbol*, [one + general classifier-ge4 + tree], "one tree", rather than 一棵 *árbol*, [one + specific-proper classifier-ke1 + tree], "one tree", although the classifier 棵 /ke1/ is specific for nouns related to plants.

5.2.2.2 Translation equivalent strategy

The translation equivalent strategy, also referred to as analogical criterion (Liceras et al., 2008) or analogical gender (Montes-Alcalá & Lapidus Shin, 2011), is commonly associated with assigning gender to inserted nouns in mixed NPs based on the gender of their translation equivalents in the recipient language (for an overview, see Bellamy & Parafita Couto, 2022). For example, Spanish–English bilinguals produced *la party* "the_{FEM} party", aligning with the feminine gender of its Spanish translation equivalent *fiesta,* although *party* has no grammatical gender in English (see López, 2020). By analogy, we propose a translation equivalent strategy for classifier assignment in mixed Chinese–Spanish NPs. According to this approach, Spanish–Chinese bilinguals would select the specific proper classifier that corresponds to the Chinese translation equivalent of the inserted Spanish noun. For instance, bilinguals would produce *一把 tenedor,* [one + specific-proper classifier-ba3 + fork], "one fork", based on its Chinese equivalent *叉子* /cha1zi0/ that matches with the specific classifier *把* /ba3/ (i.e., commonly used for objects with handles).

5.2.3 Spanish–Chinese bilingual community in Spain

To explore how classifiers are assigned in bilingual contexts, this study focuses on the Spanish–Chinese bilingual community in Barcelona, a linguistically rich and dynamic population shaped by sustained contact between typologically distinct languages and offering valuable insights into cross-linguistic influence and code-switching. This community has grown significantly over recent decades: since the 1980s, a major wave of Chinese migration has taken place (Beltrán Antolín & López, 2013), making the Chinese population the second-largest non-EU immigrant group in Spain since 2016 (Robles-Llana, 2018). Between 2003 and 2020, the number of Chinese residents rose from 13,416 to over 56,000, with the population of children under 16 increasing from 2,400 to more than 12,000 (He, 2024). This growth of the second generation has led to the emergence of a relatively young but increasingly stable Spanish–Chinese bilingual community. In this community, most first-generation immigrants share a common regional origin, occupational background (e.g., family-run businesses), limited educational background, and strong ties to Chinese cultural traditions (Beltrán Antolín, 2008). These conditions have fostered a cohesive and homogeneous community environment that supports Chinese heritage language transmission.

Moreover, second generations often serve as linguistic and cultural mediators within and beyond the community, a role necessitated by their parents' limited integration into broader Spanish society (Robles-Llana, 2018). This dynamic has resulted in second generations developing proficiency in Chinese alongside dominant exposure to Spanish (and often Catalan) through schooling and broader societal interaction. In addition, growing up in a multilingual environment exposed these heritage speakers to regular interactions with Chinese and Spanish monolinguals, Spanish–Chinese bilinguals, Catalan–Spanish bilinguals, and other bi/multilinguals. This environment fosters rich cross-linguistic contact and encourages flexible, adaptive language practices, such as frequent intra-clausal code-switching and the dynamic management of typologically distinct language systems. Despite being a relatively young community, Spanish–Chinese bilinguals in Barcelona demonstrate vibrant use of both Spanish and Chinese, making this context especially valuable for examining how bilinguals navigate and manage their languages, as well as their code-switching patterns.

5.3 The current study

5.3.1 Research questions

In this study, we investigate code-switching patterns involving classifiers in mixed Chinese–Spanish NPs with an understudied bilingual population. Given that Chinese features several dozen classifiers and a default classifier *ge*, we thus focus on examining how early Spanish–Chinese bilinguals, heritage speakers in Barcelona, Spain, use Chinese classifiers in mixed Chinese–Spanish NPs. Based on this, we further identify which classifier assignment strategies they prefer in mixed NPs. Thus, three main research questions are addressed:

1. Do early Spanish–Chinese bilinguals default to the classifier *ge* when producing Spanish nouns within mixed Chinese–Spanish NPs?
2. To what extent do the classifier patterns in mixed Chinese–Spanish NPs align with those observed in unilingual Chinese NPs?
3. What classifier assignment strategies do early Spanish–Chinese bilinguals employ in the production and comprehension of mixed Chinese–Spanish NPs?

5.3.2 Hypotheses

Building on the findings from Chapter 3, which showed that early Spanish–Chinese bilinguals flexibly adopted different gender assignment strategies depending on task type and the sociolinguistic characteristics of their community, we hypothesize that similar patterns will emerge for classifier assignment in mixed Chinese–Spanish NPs.

1. We predict that bilingual speakers will not uniformly default to the general classifier *ge* when producing Spanish nouns in mixed NPs. Instead, they are expected to alternate between general *ge* and expected-specific classifiers, reflecting their flexibility in processing classifiers and the influence of their existing Chinese classifier knowledge.
2. We predict that classifier assignment patterns in mixed Chinese–Spanish noun phrases will diverge from those observed in unilingual Chinese contexts, showing adaptations to cross-linguistic influences and bilingual language mode.
3. We expect that bilinguals' classifier choices will vary across production and comprehension tasks, with a greater tendency to use specific classifiers in comprehension, where metalinguistic awareness is more activated, and a relatively balanced use of *ge* and specific classifiers in production, consistent with patterns of task-dependent language

processing observed in Chapter 3. Specifically, we predict bilinguals will employ both default and translation equivalent strategies in production, while they will prefer translation equivalent strategies in comprehension.

5.4 Methods

5.4.1 Participants

Thirty early Spanish–Chinese bilinguals (M_{age} = 20.5 years, SD_{age} = 1.66 years, 20 females) from Barcelona (Spain) participated in this study (see Table 5.4.1). Of these, 26.67% (n = 8) reported having lived in Spain only, while the remaining 73.33% (n = 22) had lived in both Spain and China. Notably, one participant from the group was born in France and moved to Spain, and one participant was born in Italy, lived in China for four years, and then moved to Spain. Overall, most participants were born either in Spain or China and currently live in Spain. All participants were raised in Chinese heritage families, with all family members of Chinese nationality. They learned Chinese at home or at language schools, and acquired Spanish with an average age of 3.47 years (*SD* = 3.42) in educational contexts in Spain. All participants filled in a background questionnaire adapted from the Bilingual Code-Switching Profile (BCSP) (Olson, 2022) to gain insight into their language profile. Moreover, they reported their daily language of Chinese ($Frequency_{mean}$ = 40.7%, $Frequency_{SD}$ = 0.17), Spanish ($Frequency_{mean}$ = 41.9%, $Frequency_{SD}$ = 0.156), Catalan ($Frequency_{mean}$ = 9.12%, $Frequency_{SD}$ = 0.059), and English ($Frequency_{mean}$ = 6.18%, $Frequency_{SD}$ = 0.053). All participants voluntarily participated in the study and completed two production and one comprehension tasks in CS mode and unilingual Chinese mode sequentially. Participants received monetary compensation after the completion of all tasks. All related materials, including stimulus lists for three tasks and the results for the questionnaire and three tasks, can be found on the OSF.

Table 5.4.1. Participant characteristics in Experiments 1, 2, and 3.

	Experiments 1, 2, and 3
Number of Female/Male participants	20/10
Mean age in years (SD; range)	20.5 (1.66;18-24)
Number of participants born in Spain/Italy/France	18/1/1
Number of participants born in China	10
Age of Chinese acquisition	After birth
Age of Spanish acquisition	3.47 years (*SD* = 3.42)
Daily use of Chinese (frequency)	40.7% (*SD* = 0.170)
Daily use of Spanish (frequency)	41.9% (*SD* = 0.156)
Daily use of Catalan (frequency)	9.12% (*SD* = 0.059)
Daily use of English (frequency)	6.18% (*SD* = 0.053)
Participants with living experience in China (i.e., more than one year)	22

5.4.2 Production tasks

5.4.2.1 Experiment 1: Director-Matcher task (Toy task)

Materials

Thirty-two colored pictures were obtained from the Multipic database (Duñabeitia et al., 2018) based on five criteria: 1) objects on pictures should relate to count nouns with non-abstract concepts; 2) each object should relate to only one Chinese name and one corresponding Spanish translation equivalent; 3) objects should associate predominantly with only one proper specific classifier based on *Xiandai Hanyu Liangci Yongfa Cidian* (*A Dictionary of Modern Chinese Classifier Usage,* Guo, 2002); 4) no more than two objects associated with the same classifier; 5) the general classifier *ge* is grammatically acceptable for all objects. In addition to consulting the dictionary to ensure the accurate use of classifiers for each object, the most proper classifiers were also determined based on consensus among ten native Chinese PhD students in linguistics, who filled out the classifier they considered most proper for each object. Consequently, 32 target pictures were selected, each with a concrete concept and a corresponding proper classifier. Hereafter, we refer to the proper classifiers, identified through native speaker consensus and dictionary consultation, as *expected-specific classifiers*. Given that the relationship between nouns and their corresponding classifiers is not always transparent and that speakers often use multiple classifiers for the same noun (see Section 5.2.1.1), if participants produced a specific classifier that differed from the expected one for a given noun, we label those as *unexpected-specific classifiers*.

Procedure

Participants completed the toy task in pairs. First, participants were presented with the 32 target pictures and learned the exact Chinese name and Spanish equivalent of each picture for 3 seconds. Then, they sat opposite each other at a table with a hardboard between them. They were randomly assigned a role as either "director" or "matcher" to complete the task. Last, they were presented with two separate grids of blank space, on which the same 32 target pictures were displayed at different positions (see Figure 5.4.1).

The toy task was completed sequentially in CS mode and unilingual Chinese mode. In each mode, participants followed the same procedures. First, printed instructions were provided in CS mode for the CS task and in unilingual Chinese for the unilingual task. Second, the "director" described each picture in sequence, including information such as name, color, shape, and position, to the "matcher". The "matcher" then moved the corresponding picture to a specified position based on the description. Third, the "matcher" verified details of each picture with the "director" once the descriptions were complete. In the CS mode, "director" and "matcher" were always asked to complete the task in Chinese while naming pictures in Spanish. In the unilingual Chinese mode, descriptions of pictures were only in Chinese. All participants completed the toy task in CS mode prior to doing so in unilingual Chinese mode. Target NPs are as follows (see examples 4 and 5), with the Spanish noun in italics.

4. 一	面/个	*espejo*	[mixed NP]
one	CL.mian$_{specific}$/ CL.ge$_{default}$	mirror	

“a mirror” (cf. 一面镜子)

5. 一	架/个	钢琴	[unilingual Chinese NPs]
one	CL.jia$_{specific}$/CL.ge$_{default}$	piano	

“a piano”

Figure 5.4.1. Illustration of the toy task setup.

5.4.2.2 Experiment 2: Repetition task

Materials

In this task, Chinese names and their Spanish equivalents for the 32 colored pictures used in the director-matcher task served as target nouns in unilingual Chinese NPs and mixed Chinese–Spanish NPs, respectively.

Thirty-two unilingual Chinese sentences were initially selected and generated based on the BCC corpus (Xun et al., 2016), each comprising a unilingual Chinese NP composed of a Chinese noun referring to the name of each picture.

Unilingual mode: In unilingual mode, the original 32 unilingual Chinese sentences were modified to reflect classifier comparisons regarding expected-specific classifiers and general classifier *ge* within unilingual NPs. Thus, 64 unilingual Chinese sentences with different classifiers were obtained.

CS mode: In CS mode, target Chinese nouns within unilingual Chinese NPs in the 32 unilingual sentences were replaced by their Spanish equivalents, resulting in 32 CS sentences featuring 32 mixed Chinese–Spanish NPs (i.e., Chinese numeral + expected-specific classifiers + Spanish nouns). These sentences were then modified by replacing expected-specific classifiers with the general classifier *ge*, resulting in 32 comparison sentences. Thus, a total of 64 CS sentences were constructed, with the contrast of sentence pairs being the classifiers (i.e., expected-specific classifiers vs. general classifier *ge*). To eliminate the potential impact of the NP position, the placement of NPs in each sentence was counterbalanced in each mode. Consequently, NPs at the beginning (11) (see Example 6), in the middle (11) (Example 7), and at the end (10) (Example 8) of 32 CS and 32 unilingual Chinese comparison sentences were identified.

6. CS sentences with mixed Chinese–Spanish NPs:

一支/个 lápiz 和一些纸就能让他画出世界万物，而且画得惟妙惟肖，栩栩如生.

Unilingual Chinese sentence:

一支/个铅笔和一些纸就能让他画出世界万物，而且画得惟妙惟肖，栩栩如生.

"***A pencil*** and some paper allowed him to draw everything in the world, and to do so in a beautiful, lifelike way."

7. CS sentences with mixed Chinese–Spanish NPs:

   我看到天空中飘着***一朵/个 nube***，乍一看，还以为是个小山包.

   Unilingual Chinese sentence:

   我看到天空中飘着***一朵/个云***，乍一看，还以为是个小山包.
   "I saw ***a cloud*** in the sky that, at first glance, looked like a small hill."

8. CS sentences with mixed Chinese–Spanish NPs:

   直升机开始在空中摇晃，然后坠入树林中，撞上了***一棵/个 arbol.***

   Unilingual Chinese sentence:

   直升机开始在空中摇晃，然后坠入树林中，撞上了***一棵/个树***.
   "The helicopter began to shake in the air, then crashed into the woods and hit ***a tree***."

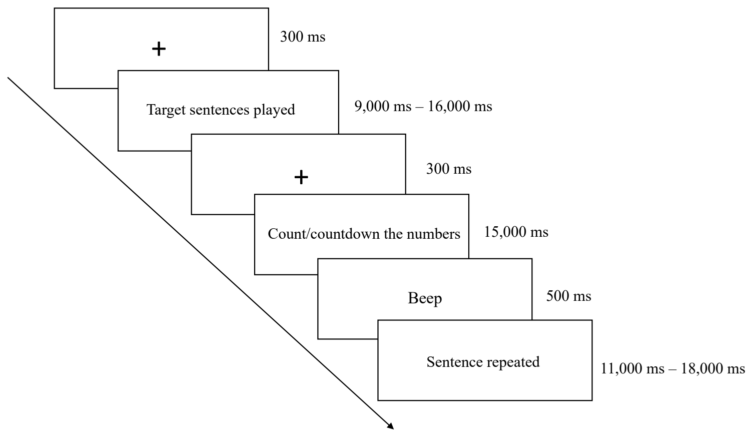
Procedure

The repetition task was programmed in E-prime 2.0 software (Schneider et al., 2002) with a by-subject order design, whereby the order of both the blocks and sentences was randomized across participants. Each participant completed this task in CS mode and unilingual Chinese mode sequentially and individually. In all instances, the CS mode is always completed prior to those in the unilingual Chinese mode. For each participant, the procedure in both CS mode and unilingual Chinese mode was as follows (see also Figure 5.4.2).

CS mode: First, each participant sat in front of a screen and was provided with instructions in CS mode, instructing them to listen to and repeat 64 CS sentences. Second, each participant saw a fixation cross on the screen for 300 ms, and then heard a CS sentence. Each sentence was played only once, and participants were instructed to repeat it as much as possible in a manner and tempo that was natural to them. Third, another fixation cross was presented for 500 ms, and participants were then instructed to count from ten to twenty or down from twenty to ten in their preferred language for 15 seconds. This intervening task was designed to prevent rote memorization and instead elicit responses that reflect participants' underlying grammatical knowledge, particularly when repetition exceeded their capacity of short-term memory (Gullberg et al., 2009). Fourth, a beep sound was played for 500 ms as a prompt, after which participants repeated each sentence within a specified time range (11–18 seconds) based on the length of each sentence. Participants' vocal responses were automatically recorded.

Unilingual mode: In the unilingual mode, participants read unilingual Chinese instructions and repeated 64 unilingual Chinese sentences sequentially. They followed the same experimental procedure as in the CS mode.

Figure 5.4.2. Example trial procedure in the repetition task.



5.4.3 Comprehension task

5.4.3.1 Experiment 3: Acceptability judgment using a two-alternative forced-choice task (2AFC judgment task)

Materials

***Critical sentences*:** 64 CS sentences containing mixed Chinese–Spanish NPs and 64 unilingual Chinese sentences comprising unilingual Chinese NPs were generated by following the same procedure as in the repetition task. Each sentence pair was compared regarding the use of classifiers within NPs, i.e., the general classifier *ge* vs. expected-specific classifiers. The location of the NPs in each sentence was counterbalanced in each mode, with the NPs occurring at the beginning (10), middle (11), or end (11) of the sentence.

***Filler sentences*:** a total of 72 CS filler sentences were constructed, comprising 36 in Spanish with either inserted Chinese nouns or adjectives and 36 in Chinese with either inserted Spanish nouns or adjectives. Additionally, 36 filler unilingual sentences were created, including 18 in unilingual Spanish and 18 in unilingual Chinese. Sentences were presented in pairs, with the contrast between each sentence pair being the placement of adjectives, either pre-nominal or post-nominal, in the NPs. The outcomes of these filler sentences fall outside the scope of the present study and will be reported in other studies.

Procedure

This experiment was conducted online using Qualtrics (Qualtrics, LLC, Provo, UT, USA). In this task, participants chose either the expected-specific classifier or the general classifier *ge* to accompany the Spanish noun in a CS sentence or to accompany the Chinese noun in a unilingual Chinese sentence. Critical and filler sentences, presented in the same language mode, were randomized across participants. In each mode, participants read sentence pairs sequentially and made selections before proceeding to the next. Notably, participants always completed this task in CS mode before proceeding to the unilingual Chinese mode.

5.5 Results and analysis

All recorded NPs were initially transcribed, and those containing at least a Chinese classifier with a Spanish noun (in CS mode) or with a Chinese noun (in unilingual Chinese mode) were extracted and proceeded for further analysis. In unilingual mode, production and comprehension data were analyzed descriptively, as participants consistently exhibited a clear-cut preference for expected-specific classifiers across the three tasks. In CS mode, to assess whether classifier choices significantly differed when Spanish nouns were paired with different Chinese classifiers in CS mode, a one-way repeated-measures ANOVA was conducted for production tasks. Given the nature of the comprehension data, i.e., the violation of normality and the presence of a within-subject factor with two levels, a non-parametric Wilcoxon signed-rank test was used. Results showed that classifier choices are significantly different when Spanish nouns pair with distinct Chinese classifiers in all three tasks: director-matcher task ($F(1.28, 37.14) = 13.18$, $p < .001$, $\eta^2\_G = .31$), repetition task ($F(1.22, 35.26) = 42.29$, $p < .001$, $\eta^2\_G = .57$), and judgment task ($V = 392$, $p < .001$, $r = .70$).

5.5.1 Production tasks

5.5.1.1 Director-Matcher task (Toy task)

Unilingual mode

In unilingual mode, 960 trials were initially collected, of which 959 contained identifiable classifiers within NPs for further analysis (i.e., one trial was removed due to error). Three distinct patterns were observed in unilingual Chinese NPs, with NPs presented in bold and italics for clarity in examples (9), (10), and (11).

9. 第一行　第一列　　是一 ***辆***　　 ***自行车***. [expected-specific classifier]
First raw first column is ***one CL.liang4 bicycle***.
NUM $CL_{expected}$ N
"The first column of the first column is a bicycle."
(speaker 10M)

10. 第三排　第三列　　是一　***朵***　　***假发***. [unexpected-specific classifier]
Third raw third column is ***one CL.duo3 wig***.
NUM $CL_{unexpected}$ N
"The third column of the third row is a wig."
(speaker 30M)

11. 第六列　　　是 一　***个***　　　***硬币***. [general classifier]
Sixth column is ***one CL.ge4 coin***.
NUM $CL_{general}$ N
"The sixth column is a coin."
(speaker 3D)

Examples (9), (10), and (11) highlight three combinations of classifiers and Chinese nouns in NPs: NPs with expected-specific classifiers, NPs with unexpected-specific classifiers, and NPs with the general classifier *ge* (see an overview in Table 5.5.1).

Table 5.5.1. Distribution of classifiers in unilingual Chinese NPs in the toy task.

	Total
Unilingual NPs with expected classifiers	711 (74.14%)
Unilingual NPs with general classifier *ge*	144 (15.02%)
Unilingual NPs with unexpected classifiers	104 (10.84%)
Total	959

Table 5.5.1 illustrates a strong tendency to use expected-specific classifiers for Chinese nouns in unilingual Chinese NPs (74.14%), compared to the combinations of the general classifier *ge* (15.02%) and unexpected-specific classifiers (10.84%). Overall, participants overwhelmingly preferred expected-specific classifiers in unilingual NPs in the director-matcher task. Moreover, a highly consistent pattern of classifier assignment was observed across participants and Chinese nouns (see Appendix 5.A).

CS mode

In CS mode, 960 mixed NPs were obtained, of which 9 trials (0.94%) were removed from further analysis as unexpected responses (e.g., errors) occurred. In general, three patterns with different identifiable classifiers in mixed NPs were observed in examples (12), (13), and (14), and mixed NPs are presented in bold and italics:

12. 在第一行 的 第一列 是一 ***艘 巨大的 barco***. [expected-specific classifier]
    In first row of first column is ***one CL-sou1 huge ship***.
    NUM $CL_{expected}$ ADJ N
    "In the first column of the first row is a huge ship."
    (e.g., speaker 17D)

13. 第一行 第二列 我有 ***一 头 蓝色的 peluca***. [unexpected-specific classifier]
    First row second column I have ***one CL-tou2 blue wig***.
    NUM $CL_{unexpected}$ ADJ N
    "I have a blue wig in the second column of the first row. "
    (e.g., speaker 13D)

14. 第四个 图片 是 ***一 个 sombrero***. [general classifier]
    The fourth picture is ***one CL-ge4 hat***.
    NUM $CL_{general}$ N
    "The fourth picture is a hat."
    (e.g., speaker 22M)

Similar to the NPs in unilingual Chinese, participants produced three types of classifiers for Spanish nouns within mixed NPs: mixed NPs with expected-specific classifiers, mixed NPs with unexpected-specific classifiers, and NPs with the general classifier *ge* (see an overview in Table 5.5.2).

Table 5.5.2. Distribution of classifiers in mixed NPs in the toy task.

	Total
Mixed NPs with expected-specific classifiers	440 (46.27%)
Mixed NPs with the general classifier *ge*	367 (38.59%)
Mixed NPs with unexpected-specific classifiers	144 (15.14%)
Total	951

Table 5.5.2 illustrates that the occurrence of expected-specific classifiers (46.27%) and the general classifier *ge* (38.59%) with Spanish nouns in mixed NP is comparable ($p = 1$, $t = 0.93$), and that unexpected classifiers (15.14%) also sometimes occur with Spanish nouns. In sum, participants preferred both expected-specific classifiers and the general classifier *ge* for Spanish nouns within mixed NPs. Notably, the patterns of classifier assignment are also relatively consistent across participants and Spanish nouns in mixed NPs (see Appendix 5.B).

5.5.1.2 Repetition task

Unilingual mode

In unilingual mode, 1,920 trials were obtained, of which 221 (11.51%) were excluded due to unexpected responses (e.g., producing single nouns without classifiers) or no responses. Table 5.5.3 shows the distribution of general classifier *ge*, expected, and unexpected classifiers in participants' responses. Out of 1,217 responses with expected classifiers, 472 trials (38.8%) originally with the general classifier *ge* were changed to expected classifiers, while 745 trials (61.2%) remained with the original expected classifiers. For responses with general classifier *ge* (375 trials), 315 (84%) trials that originally had the general classifier *ge* were maintained in participants' responses, whereas 60 (16%) were changed from expected classifiers to general classifier *ge*. Additionally, there are responses with unexpected classifiers, which were changed from NPs with either the general classifier *ge* (66 trials, 71.7%) or expected classifiers (41 trials, 38.3%). Overall, participants showed preferences for using expected-specific classifiers in their repetitions. In addition, a similar trend of classifier assignment was observed across participants and items (see Appendix 5.C).

Table 5.5.3. Distribution of classifiers in unilingual NPs in the repetition task.

	General classifiers *ge*	Expected-specific classifiers	Total
NPs with general classifier *ge*	315 (84%)	60 (16%)	375
NPs with expected-specific classifiers	472 (38.8%)	745 (61.2%)	1,217
NPs with unexpected-specific classifiers	66 (61.7%)	41 (38.3%)	107
Total			1,699

Notes: 1. *NPs with general classifier (ge)*: NPs where the general classifier *ge* was used in participants' responses. The "**general classifier *ge***" column refers to NPs where the general classifier *ge* was maintained in participants' responses, while the "**expected-specific classifiers**" column indicates cases where participants changed expected-specific classifiers to general classifier *ge*. 2. *NPs with expected classifiers*: these are NPs that used a specific expected classifier. The "**general classifier *ge***" column reflects cases where participants replaced general classifier *ge* with expected-specific classifiers; the "**expected-specific classifiers**" column indicates cases where expected-classifiers remained unchanged. 3. *NPs with unexpected classifiers*: NPs that used specific **unexpected** classifiers. The "**general classifier *ge***" column refers to cases where participants changed the general classifier *ge* to unexpected-specific classifiers. The "**expected-specific classifiers**" column indicates cases where participants replaced expected-specific classifiers with unexpected-specific classifiers.

CS mode

In CS mode, 1,920 responses were collected, of which 312 (16.25%) were removed due to the same reason as in the unilingual mode. Table 5.5.4 illustrates that participants produced 815 mixed NPs with expected classifiers, of which 561 (68.8%) trials maintained the original use of expected classifiers, while 254 (31.2%) trials had the general classifier *ge* were replaced with expected classifiers. Regarding responses with the general classifier *ge* (663 trials), 473 (71.3%) mixed NPs retained the originally assigned general classifier *ge*, while 190 trials (28.7%) were changed from expected classifiers to the general classifier *ge.* Additionally, 130 responses involved the use of unexpected classifiers, with 76 (58.5%) changed from mixed NPs originally using the general classifier *ge,* and 54 trials (41.5%) modified from mixed NPs that initially contained expected classifiers. Overall, participants produced mixed NPs with expected classifiers at a rate comparable to those with the general classifier ge (p = 0.477, t = 1.45). Additionally, a relatively consistent pattern of classifier assignment was observed across participants and items (see Appendix 5.D).

Table 5.5.4. Distribution of classifiers in mixed NPs in the repetition task.

	General classifiers *ge*	Expected classifiers	Total
NPs with general classifier *ge*	473 (71.3%)	190 (28.7%)	663
NPs with expected classifiers	254 (31.2%)	561 (68.8%)	815
NPs with unexpected classifiers	76 (58.5%)	54 (41.5%)	130
Total			1,608

Note: The interpretation of Table 5.5.4 is the same as that in Table 5.5.3.

5.5.1.3 2AFC Judgment task

Unilingual mode

In the 2AFC judgment task, participants chose the more natural sentences from classifier-varied pairs within unilingual Chinese NPs, yielding 960 responses for the unilingual mode. Table 5.5.5 illustrates the distribution of preferences for classifiers within unilingual Chinese NPs. Table 5.5.5 shows an overwhelming preference for unilingual NPs with expected-specific classifiers (858 trials, 89.37%), compared to those with the general classifier *ge* (102 trials, 10.63%). In sum, participants exhibited a strong tendency to use expected-specific classifiers for unilingual NPs in the judgment task. Additionally, a prevalent trend of assigning classifiers across participants and items was observed (see Appendix 5.E).

Table 5.5.5. Distribution of classifiers in unilingual NPs in the 2AFC judgment task.

	Total
Unilingual NPs with general classifier *ge*	102 (10.63%)
Unilingunal NPs with expected classifiers	858 (89.37%)
Total	960

CS mode

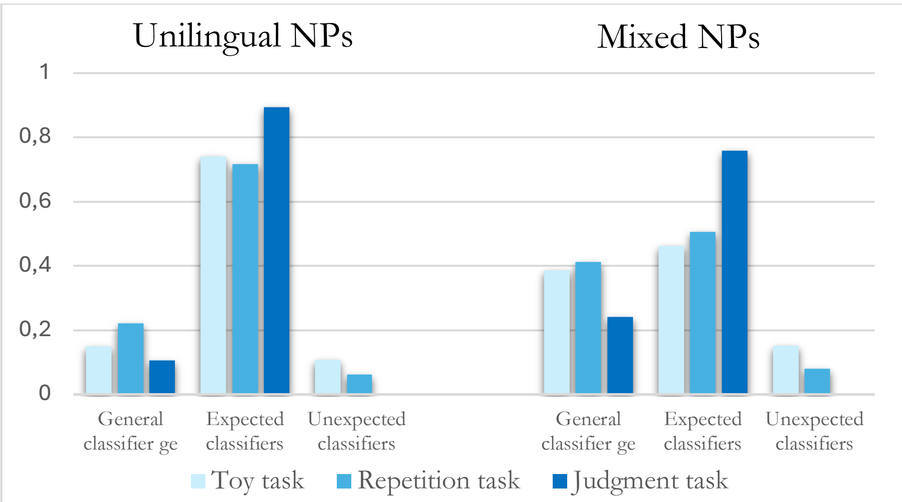
In the CS mode, 960 responses were collected, with the majority being mixed NPs using expected-specific classifiers (728 trials, 75.83%), while the remaining used the general classifier *ge* (232 trials, 24.17%, see an overview in Table 5.5.6). Overall, participants showed a strong preference for expected-specific classifiers over the general classifier *ge* in mixed NPs ($V = 392, p < .001, r = .70$), consistent with patterns observed in unilingual NPs. Additionally, a largely consistent pattern of classifier assignment was observed across participants and items, with only a few exceptions (see Appendix 5.F).

Table 5.5.6. Distribution of classifiers in mixed NPs in the 2AFC judgment task.

	Total
Mixed NPs with general classifier *ge*	232 (24.17%)
Mixed NPs with expected classifiers	728 (75.83%)
Total	960

Overall, participants showed a clear preference for expected-specific classifiers (i.e., 74.14% for the toy task, 71.63% for the repetition task, and 89.37% for the 2AFC task) with Chinese nouns in unilingual Chinese NPs across all three tasks, though the general classifier *ge* and unexpected classifiers were occasionally used. In contrast, for Spanish nouns in mixed NPs, participants showed comparable preferences for using the general classifier *ge* (i.e., 38.59% for the toy task, 41.23% for the repetition task) and expected-specific classifiers (46.27% for the toy task, 50.68% for the repetition task) in production tasks, but exhibited a strong tendency for expected-specific classifiers (75.85%) in the comprehension task (see Figure 5.5.1).

Figure 5.5.1. Distribution of classifiers in unilingual and mixed NPs across three tasks.



5.6 Discussion

This study investigated how early Spanish–Chinese bilinguals assign Chinese classifiers in mixed Chinese–Spanish NPs across production and comprehension, particularly examining whether they default to the general classifier *ge* when producing Spanish nouns and whether they use comparable patterns to assign classifiers in unilingual and CS mode. Building on this, we further examined which classifier assignment strategies bilinguals would employ when assigning classifiers to Spanish nouns in mixed NPs.

Variation in classifier assignment across language mode and tasks

The first key finding is that early Spanish–Chinese bilinguals exhibited a consistent preference for expected-specific classifiers in unilingual NPs across production and comprehension. In contrast, in mixed Chinese–Spanish NPs, their classifier choices showed task-dependent variations. In production, bilinguals used both the general classifier *ge* and expected-specific classifiers for Spanish nouns at relatively equal rates. Conversely, in comprehension, expected-specific classifiers were predominantly assigned to Spanish nouns. In sum, in response to the first two research questions, which refer to whether bilinguals default to the classifier *ge* for Spanish nouns in mixed NPs and whether the classifier patterns in mixed NPs resemble those in unilingual NPs, the findings suggest that bilinguals' classifier choices exhibited language mode-dependent and task-dependent variations. These findings support the first two hypotheses.

The varying patterns of classifier assignment across language contexts highlight differences in the underlying processing of classifiers in unilingual and CS modes. In the unilingual mode, bilinguals relied heavily on both the semantic properties of Chinese nouns and rote memory of their associated classifiers (Li & Thompson, 1981). When producing unilingual Chinese NPs, the selection of a semantically "proper" classifier is relatively straightforward, as the obvious semantic features of Chinese nouns (e.g., 树 /shu4/ "tree") readily activate the corresponding classifiers (棵 /ke1/, "specific classifier for plants"), facilitating the frequent use of expected classifiers. In CS mode, bilinguals drew on both the general classifier *ge* that defaults to Spanish nouns and expected-specific classifiers that are highly associated with Chinese equivalents of Spanish nouns during production, resulting in relatively equal use of the general classifier *ge* and expected classifiers in mixed NPs. When producing classifiers for Spanish nouns, which often lack the straightforward semantic cue typically available in Chinese, bilinguals may fail to recall the "proper" classifier or rote-remembered classifiers due to weaker or absent obvious semantic associations. In such cases, the general classifier *ge* becomes a convenient and accessible substitute. This interplay between defaulting and semantic mapping led to high frequencies of both classifiers in production. However, in comprehension, bilinguals' classifier choices were influenced strongly by the design of the task, which explicitly contrasted classifier-noun pairings, increasing their metalinguistic awareness and prompting them to evaluate classifiers based on their appropriateness for the Chinese equivalents (see details below). This, in turn, prompted the high occurrence of expected-specific classifiers for Spanish nouns in comprehension.

Classifier assignment strategies in production and comprehension

The second key finding, addressing the third research question concerning the classifier assignment strategies employed by bilinguals in mixed NPs, reveals distinct strategies across production and comprehension. In production, bilinguals adopted both default and translation equivalent strategies, as evidenced by the relatively equal frequency of Spanish nouns being paired with the general classifier *ge* and with expected-specific classifiers. This suggests both strategies are frequently utilized. However, in comprehension, bilinguals predominantly selected expected-specific classifiers highly associated with Chinese equivalents of Spanish nouns, reflecting a preference for the translation equivalent strategy. This exclusive preference is likely influenced by the task design, which presented mixed NPs in pairs of contrasting nouns with either the general classifier *ge* or expected-specific classifiers. Such a task design may increase participants' metalinguistic awareness of comparing two classifiers, amplifying their attention to the semantic fit between classifiers and Chinese equivalents of Spanish nouns. Consequently, the use of the translation equivalent strategy is likely promoted. Overall, early Spanish–Chinese bilinguals employed varied classifier assignment strategies in production and comprehension, which supports the third hypothesis.

Influence of community norms

Interestingly, the classifier assignment strategies in mixed Chinese–Spanish NPs parallel findings from the study on Spanish grammatical gender assignment in mixed Spanish–Chinese NPs using similar tasks (e.g., Wu et al., 2025). In both cases, early Spanish–Chinese bilinguals from the same community in Barcelona favored a combination of default and translation equivalent strategies in production, but preferred only translation equivalents in comprehension (see Wu et al., 2025). This cross-language consistency of strategies may suggest that the choice of assignment strategies in this community may be shaped not solely by shared underlying mechanisms for managing cross-linguistic structure, but also by broader sociolinguistic factors. A key factor appears to be the highly diverse linguistic interactions within their community, as documented in the study of grammatical gender assignment among Spanish–Chinese bilinguals within the same community (see Wu et al., 2025). Spanish–Chinese bilinguals in Barcelona grow up in a dynamic multilingual environment, engaging regularly with Chinese and Spanish monolinguals, Catalan–Spanish bilinguals, Chinese–Spanish–Catalan trilinguals, and individuals from other language backgrounds such as English. Wu et al. (2025) reported that such exposure to various language practices fosters a heightened level of adaptability and flexibility in bilinguals' adoption of gender assignment strategies. Rather than adhering to a uniform mapping system, bilinguals in this diverse linguistic context seem to adopt community-shaped strategies, such as defaulting or drawing on translation equivalents. Likewise, the current findings suggest that classifier assignment strategies among Spanish–Chinese bilinguals also reflect this adaptability and flexibility, suggesting that classifier selection, like gender assignment, is likely to be shaped by community diversity and sociolinguistic context. These findings add to existing research by demonstrating that these flexible approaches are not limited to gender assignment but also extend to classifier systems, pointing to a more general bilingual strategy for navigating multiple linguistic systems and norms.

Inter-speaker variation of classifier assignment

Notably, early Spanish–Chinese bilinguals exhibited inter-speaker variation in classifier assignment, especially in mixed Chinese–Spanish NPs (see Appendixes 5.A–F). For a given Spanish noun, individuals assigned different types of classifiers, i.e., expected-specific classifiers, the general classifier *ge*, or unexpected-specific classifiers. These variations likely reflect multiple interacting factors, including the semantic properties of nouns, speakers' memorized associations between specific classifiers and nouns based on prior exposure, and the relative frequency of classifier usage. Such variation is not unexpected, given the non-binary nature of the Chinese classifier system. Unlike binary grammatical systems (e.g., masculine and feminine in Spanish), the Chinese classifier system involves dozens of classifiers with overlapping functions and semantic usage rules, making it more prone to variability. Similar patterns of variability have been observed in non-binary gender systems, such as the five-gender Tsova-Tush system (Bellamy & Wichers Schreur, 2022), where Tsova-Tush–Georgian bilinguals also showed variable gender assignment patterns shaped by linguistic and extra-linguistic factors and usage frequency. Another likely source of variation in classifier use among participants is their language acquisition background. Most participants acquired Chinese in informal settings, primarily at home or heritage language schools (i.e., weekend or community Chinese language schools), where formal exposure is typically limited to a few hours per week. Consequently, many did not develop a classifier system as robust and stable as that of monolingual speakers raised in China (who typically receive explicit and systematic instruction in the language), leading to inconsistencies in classifier choice and a higher reliance on the general classifier *ge*. Nonetheless, more naturalistic data on Chinese–Spanish intra-clausal code-switching are needed to better understand how classifier variability unfolds in spontaneous bilingual speech. Such data would allow us to evaluate whether the patterns observed in experimental settings align with those found in everyday language use and shed light on how context and interaction shape bilingual choices.

This study offers the first exploration of classifier assignment in mixed noun phrases among early Spanish–Chinese bilinguals. The findings provide important initial insights into how bilinguals manage grammatical "conflict sites", such as the differing presence of classifiers in their two languages, and how they process classifiers within mixed NPs. However, these results should be regarded as preliminary. To enhance our understa-nding and assess the broader applicability of these patterns, future research should include a wider range of bilingual populations and employ varied task designs. Such work will be crucial for establishing the generalizability and underlying mechanisms of classifier assignment strategies in bilingual contexts.

5.7 Conclusion

This study examined how early Spanish–Chinese bilinguals assign classifiers to Spanish nouns and further identified the classifier assignment strategies they preferred in mixed Chinese–Spanish NPs. By employing a multi-task approach, we found that bilinguals preferred expected-specific classifiers in the unilingual mode, and they did not uniformly default to the general classifier *ge* when producing Spanish nouns in CS mode. Instead, they used both the general classifier *ge* and expected-specific classifiers in production, while they showed a strong preference for expected classifiers in comprehension. These patterns indicate that bilinguals draw on various strategies, the default and translation equivalent strategies, when navigating classifier assignment, with task type and language mode influencing their choices. As the first study to examine classifier assignment in mixed NPs among early Spanish–Chinese bilinguals using a multi-task design, this work sheds light on a previously underexplored aspect of cross-linguistic interaction and provides a comprehensive view of classifier assignment strategies. Notably, the findings reveal striking parallels between classifier assignment and grammatical gender assignment observed in the same community, underscoring the role of community characteristics, social network, and task type in shaping language use and highlighting the flexibility and adaptability of bilinguals. However, given that this is the first study on classifier assignment in mixed NPs, further studies with larger and more diverse participant samples and varied methodologies are needed to fully understand classifier use in bilingual speech and to determine the generalizability of these strategies across other bilingual populations.

Acknowledgements

We are sincerely grateful to all participants for their time and dedication to this research. We would also like to acknowledge the invaluable support of Prof. Núria Sebastián Gallés, head of the Speech Acquisition and Perception (SAP) group at Universitat Pompeu Fabra (UPF), for her generous guidance and warm welcome. Our special thanks go to Francesc Xavier Mayoral Terron and Silvia Blanch Rabat for their expert technical assistance and logistical support, which were essential for the successful setup and execution of the experiments.

Funding

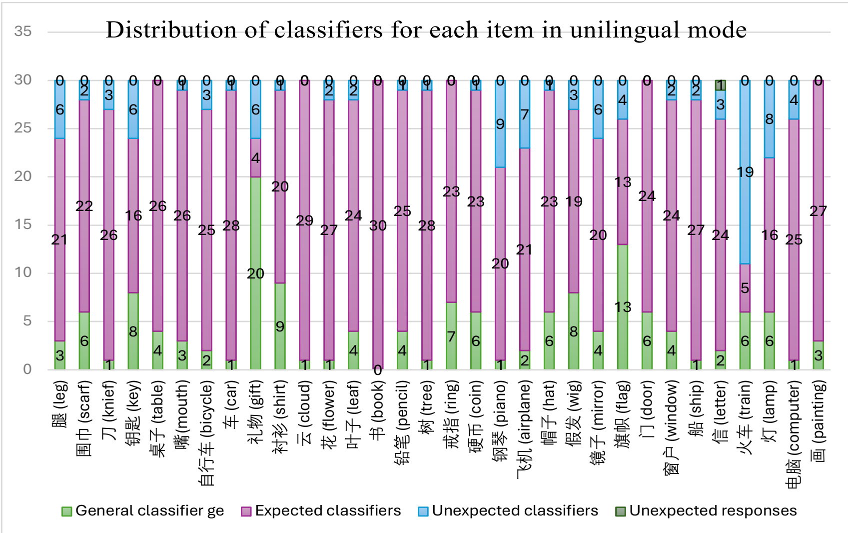
This work was partially funded by the China Scholarship Council (CSC) NO. 202107720015.

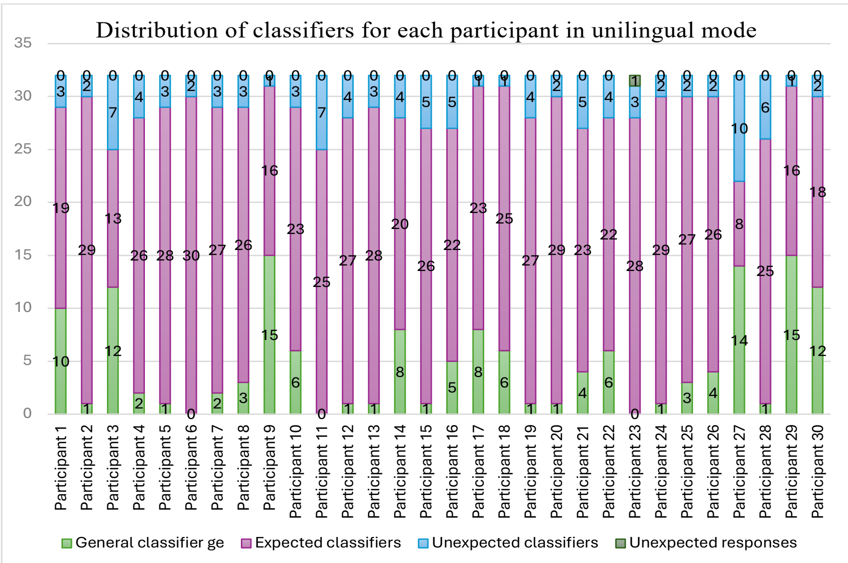
Data Availability Statement

All related materials, including experimental stimulus lists, BCSP questionnaire, statistical analyses, and results for three tasks are openly available in Open Science Framework at https://osf.io/ar3w4/?view_only=622bd917c8124c86ac2031cf4b8b5584 (view-only link).

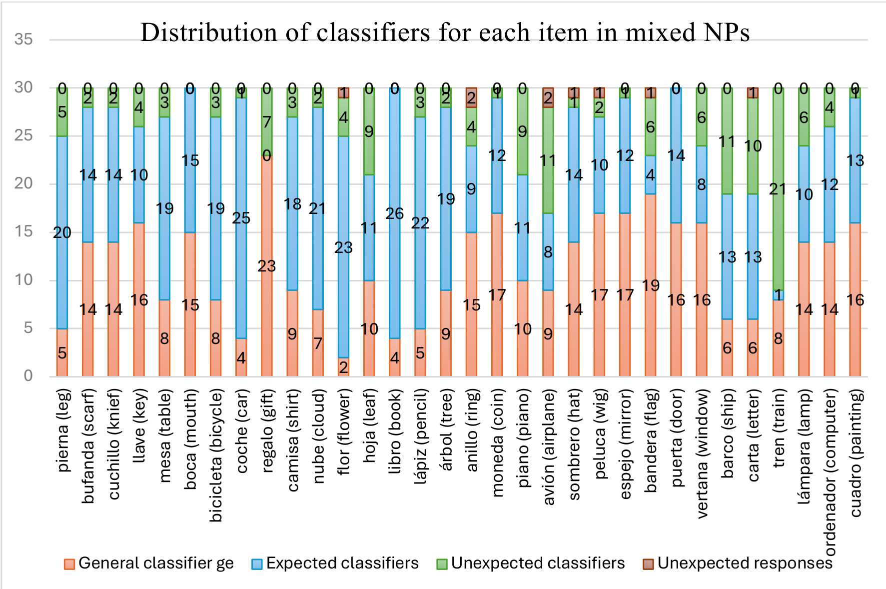
Appendix

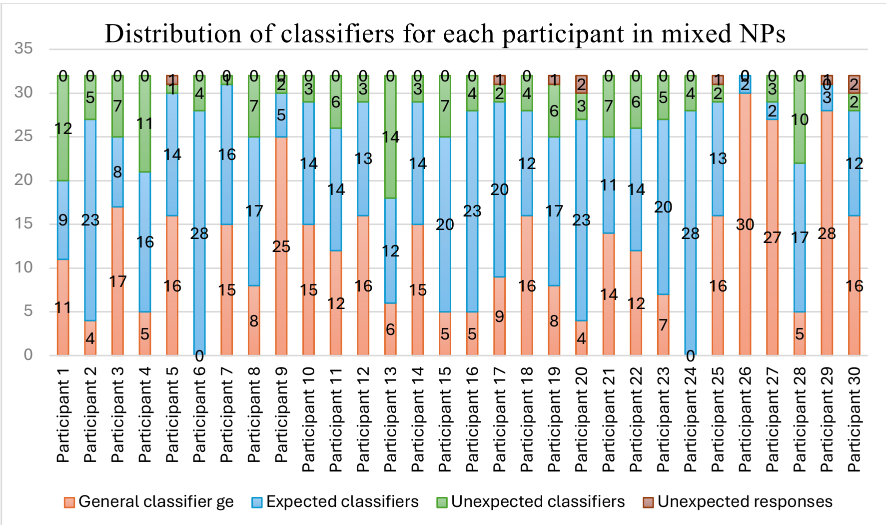
Appendix 5.A Distribution of classifiers across items and participants in unilingual NPs in the Director-Matcher task.



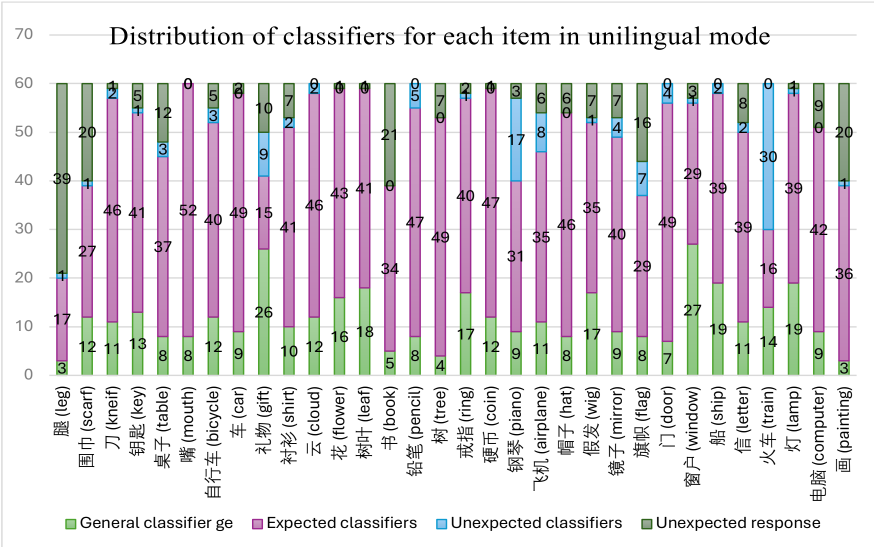


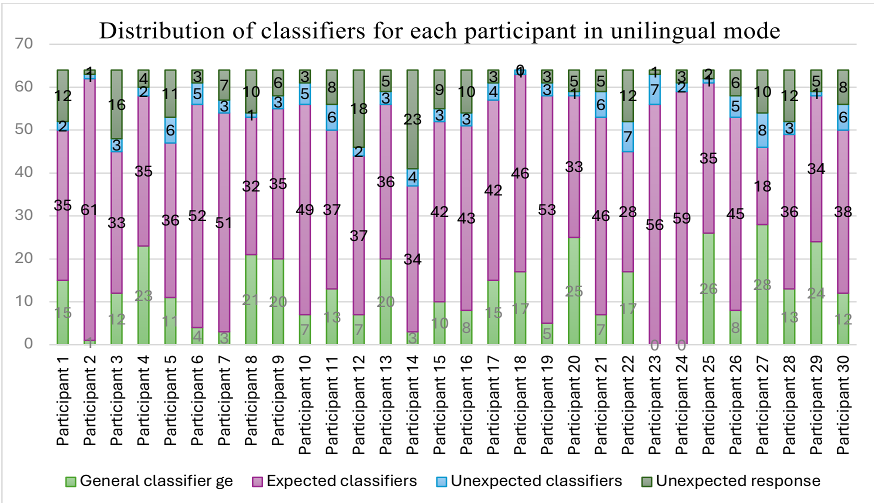
Appendix 5.B Distribution of classifiers across items and participants in mixed NPs in the Director-Matcher task.



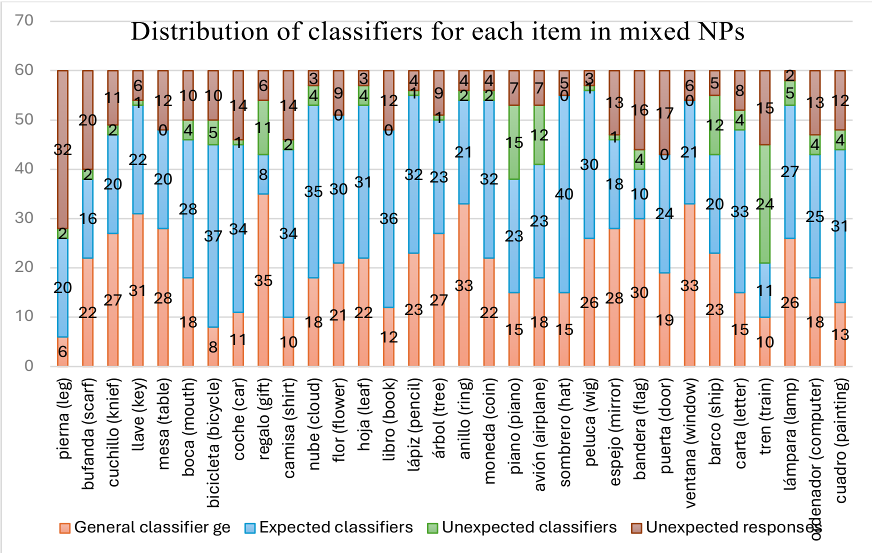


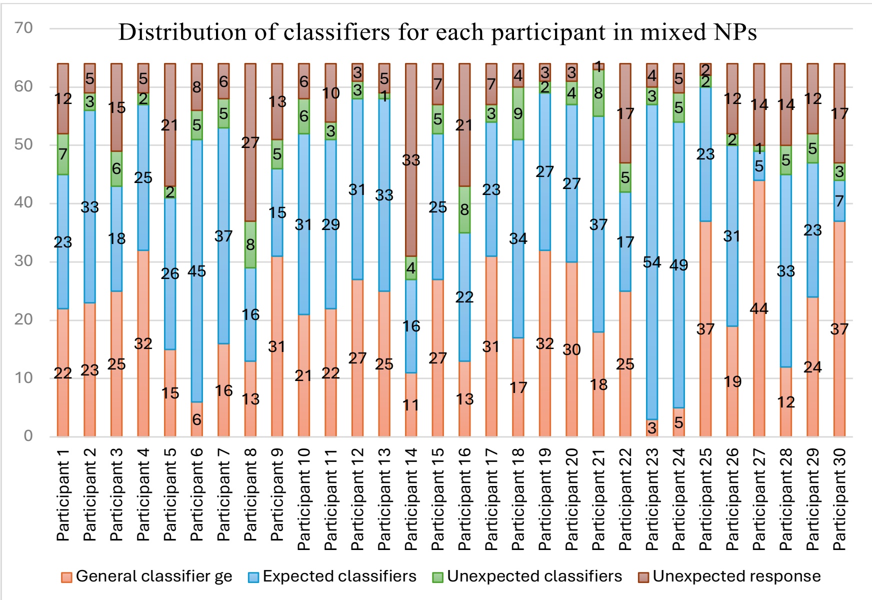
Appendix 5.C Distribution of classifiers across items and participants in unilingual NPs in the repetition task.



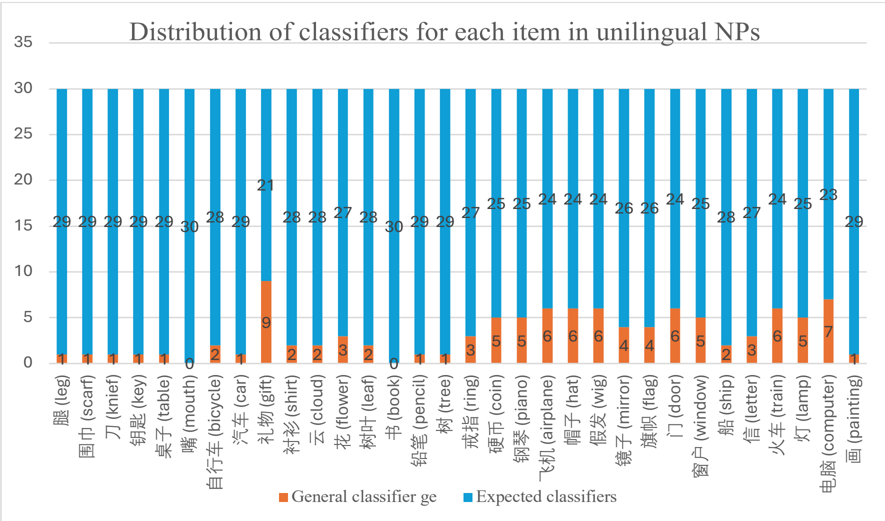


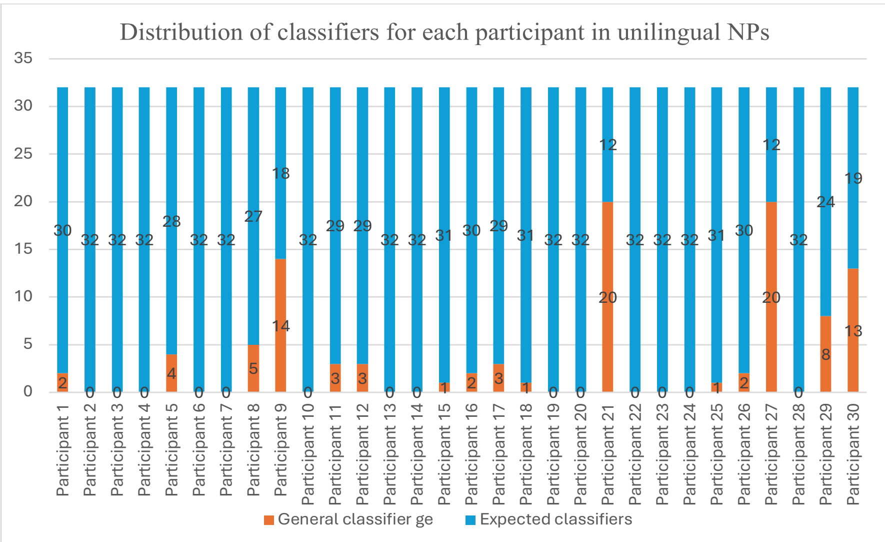
Appendix 5.D Distribution of classifiers across items and participants in mixed NPs in the repetition task.



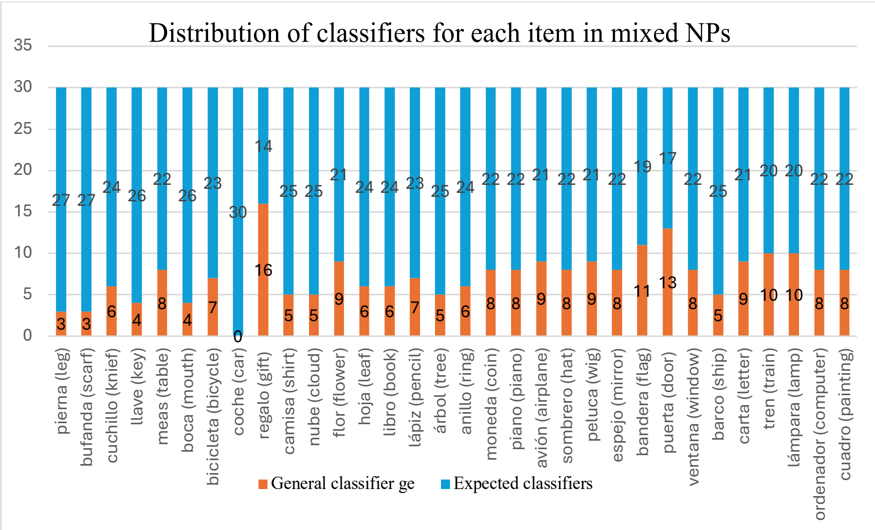


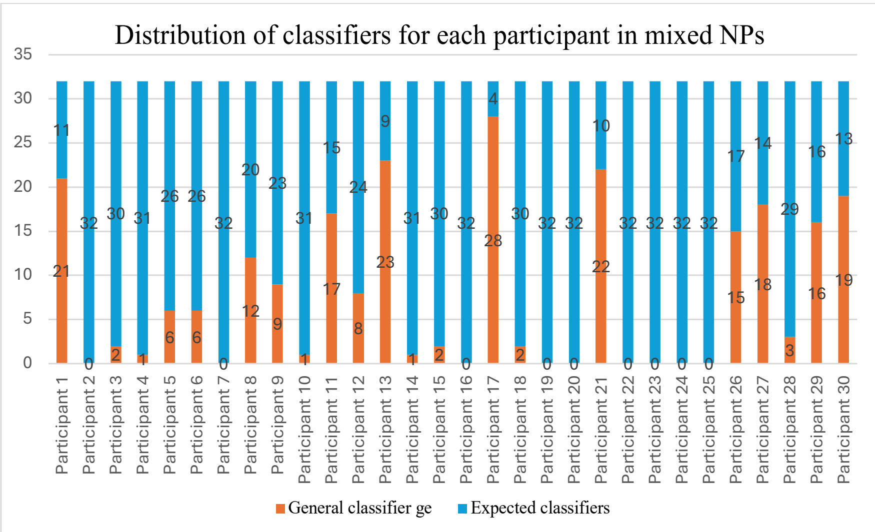
Appendix 5.E Distribution of classifiers across items and participants in unilingual NPs in the 2AFC judgment task.





Appendix 5.F Distribution of classifiers across items and participants in mixed NPs in the 2AFC judgment task.





Chapter 6

Adjective-noun Order in Spanish–Chinese Code-switching: Resolving Syntactic Conflict

---

A version of this chapter is in preparation for submission: Wu. R., Parafita Couto, M. C., & Schiller, N. O. (in preparation). Adjective-noun order in Spanish–Chinese code-switching: resolving syntactic conflict.

192 *The Role of Lexico-Syntactic Features in Noun Phrase Production and Comprehension*

Abstract

This study investigates adjective-noun placement in Spanish–Chinese code-switched constructions, addressing how bilinguals resolve syntactic differences between Spanish (postnominal adjectives, e.g., *manzana roja* "apple red") and Chinese (prenominal adjectives, e.g., *红色的苹果* "red apple"). Two models offer predictions for navigating this conflict: the Matrix Language Frame (MLF) model suggests that adjective-noun order follows the matrix language (ML), while Minimalist Program (MP)-based approaches arrive at the descriptive generalization that word order aligns with the language of adjectives. To test these predictions, thirty early Spanish–Chinese bilinguals completed a director-matcher task and a forced-choice acceptability judgment task in four contexts: unilingual Spanish and Chinese, and mixed language contexts with either Spanish or Chinese as the matrix language. Production data were analyzed descriptively, and comprehension data using Thurstone's law (case V) and one-way ANOVA. Results showed consistent expected default word order in unilingual contexts, which referred to prenominal adjectives in Chinese and postnominal adjectives in Spanish. In Spanish matrix contexts, while the matrix language, as predicted by the MLF, accounted for word order of adjective-noun patterns in production, the language of adjectives, as predicted by the MP, explained more patterns in both production and comprehension, suggesting better prediction by the MP overall. In Chinese matrix contexts, both the matrix language and the adjective languages explained adjective-noun patterns, aligning with both models. Additionally, noun insertions were preferred over adjective insertions, consistent with trends in other bilingual communities. This study is the first to empirically examine how grammatical constraints, derived from the MLF and MP, interact with Spanish–Chinese CS patterns, complementing existing research on adjective placement in bilingual speech. Moreover, our finding challenges the idea that a single theoretical model fully explains code-switching patterns; instead, it strongly suggests that both matrix language structure (MLF) and the features of individual lexical items (MP) interact and contribute to mixed constructions, leading to a more fine-grained understanding of syntactic integration.

Keywords: code-switching (CS), bilingualism, adjective-noun order, matrix-language frame (MLF), minimalist program (MP), Chinese, Spanish

6.1 Introduction

In multilingual communities, a phenomenon known as code-switching (CS) occurs when speakers integrate elements from two or more languages in a single discourse (Poplack, 1980; Deuchar, 2012). Research has established that CS is not random; instead, it follows specific patterns and rules (e.g., Bullock & Toribio, 2009). One area of particular interest in CS research is how bilinguals manage the "conflict sites" between two languages (Poplack & Meechan, 1998), where the grammatical rules of the two languages differ. These sites serve as key indicators of linguistic integration in bilingual speech, providing valuable insights into how bilinguals resolve syntactic differences. One such syntactic difference for Spanish–Chinese bilinguals arises in adjective-noun constructions. In Spanish, the typical word order is noun-adjective (e.g., *manzana roja*, [apple red], "red apple"), whereas Chinese features a reversed order with adjectives preceding nouns (e.g., *红色的苹果*, [red apple], "red apple"). Consequently, Spanish–Chinese bilinguals in code-switching may produce four possible mixed patterns: *苹果 roja* "$\text{apple}_{\text{N.CN}}$ $\text{red}_{\text{ADJ.SP}}$", *roja 苹果* "$\text{red}_{\text{ADJ.SP}}$ $\text{apple}_{\text{N.CN}}$", *manzana 红色的* "$\text{apple}_{\text{N.SP}}$ $\text{red}_{\text{ADJ.CN}}$", and *红色的 manzana* "$\text{red}_{\text{ADJ.CN}}$ $\text{apple}_{\text{N.SP}}$". Thus, the question arises whether bilinguals favor all these patterns or reject some combinations in their production and comprehension.

In Mandarin Chinese, there are two common adjective-noun structures: adjective *de*[9] noun (A *de* N) and adjective plus noun (A N) (Li & Thompson, 1981; Paul, 2010). In the "A *de* N" structure, the nominalizing particle *de* functions as an intermediary between the adjective and the head noun, and the adjective has the function of further clarifying features or

[9] In adjective *de* noun (A *de* N) constructions, *de* is the Pinyin transcription of Chinese character *"的"*.

references of the noun (e.g., *红色的花,* /hong2se4de0hua1[10]/ [red flower], "red flower") (Li & Thompson, 1981). Regarding the "A N" structure, the adjective and the noun are simply juxtaposed, and this structure is frequently used in a way that makes the adjective a name for a category of the noun (e.g., *黄豆* /huang2dou4/ [yellow bean], "soybean" (Li & Thompson, 1981; Paul, 2010). While certain adjectives can modify nouns with or without the particle *de*, most cases of adjectives require the nominalizing particle *de* when modifying a noun (Li & Thompson, 1981). In other words, the "A *de* N" structure occurs more frequently than the "A N" structure in Mandarin Chinese. Note that adjectives always precede nouns in both "A *de* N" and "A N" structures.

In contrast, Spanish predominantly adheres to a postnominal adjective order in the nominal constructions, in which the adjective is positioned after the noun. Nevertheless, a small number of exceptions exist where certain adjectives may appear before nouns with a change in meaning (Terker, 1985). In Spanish adjective-noun constructions, the postnominal adjective tends to convey a basic, attributive meaning to specify the noun (see 1a), whereas the prenominal adjective indicates a restricted and inherent characteristic of the noun (see 1b) (Terker, 1985). Notably, adjectives that relate to defining the kind of objects, such as color, nationality, shape, and religion, consistently follow nouns they modify and do not change position (Bull, 1950).

1. a. un hombre grande
   a man tall
   "a tall man"

   b. un gran hombre
   a great man
   "a great man"

(Stadthagen-González et al., 2019)

[10] In this paper, the number paired with the pinyin represents a specific tone. In Mandarin Chinese, there are four distinct phonemic tones. Tone 1 has a high-level pitch, Tone 2 has a high-rising pitch, Tone 3 has a low-dipping pitch, and Tone 4 has a high-falling pitch (Chao, 1948).

Given the differing word order in Chinese (2a) and Spanish (2b), whereby in Mandarin Chinese, adjectives precede nouns, and in Spanish, adjectives follow nouns, the adjective-noun constructions in Spanish–Chinese speech are therefore "conflict sites".

2. a. 绿色的 裙子 [Chinese: prenominal adjectives]
   green skirt
   "green skirt"

   b. falda verde [Spanish: postnominal adjectives]
   skirt green
   "green skirt"

Theoretical models have attempted to explain how the "conflict site" is managed in CS speech. On the one hand, Poplack (1980) proposed the Equivalence Constraint stating that code-switching only occurs where the syntactic rules of both languages align, implying that adjective-noun "conflict site" should block switching. Building on this, Torres Cacoullos and Vélez Avilés (2024) introduced the Variable Equivalence Hypothesis, suggesting that CS favors shared grammatical boundaries without privileging the syntax of one language. Like the Equivalence Constraint, it predicts that switching at the "conflict site" (e.g., between adjectives and nouns) will be disfavored, with switches more likely before, after, or at phrase boundaries (see Table 6.1.1, 3a–4b). However, since this study specifically focuses on code-switching between nouns and adjectives, rather than broader patterns across shared grammatical boundaries, the Variable Equivalence Hypothesis will not be examined further.

Table 6.1.1. Example CS sentences align with the Variable Equivalence Hypothesis.

CS contexts			
Chinese as matrix language		**Spanish as matrix language**	
Examples	Switch boundaries	Examples	Switch boundaries
3a.我妈妈有一只 **gato travieso**. My mom has a **cat naughty**.	before the phrase/at the phrase boundaries	3b. Mi mamá tiene un **淘气的猫**. My mom has.3S a_{Masc} **naughty cat**.	before the phrase/at the phrase boundaries
4a.我妈妈有一只**淘气的猫**, Muy lindo. My mom has a **naughty cat**, very cute.	after the phrase	4b. Mi mamá tiene un **gato travieso**, 很可爱. My mom has.3S a_{Masc} **cat naughty**, very cute.	after the phrase
Translation [English]			
My mom has a naughty cat, (very cute).			

For this specific conflict site, the placement of adjectives and nouns in code-switching, the Matrix Language Framework (MLF; Myers-Scotton, 1993, 2002) and Minimalist Program (MP)-based approaches (MacSwan, 1999) offer targeted predictions (see Section 6.2.1 for detailed information). According to the MLF (Myers-Scotton, 1993, 2002), code-switching involves a matrix language (ML) that provides the morphosyntactic frame and an embedded language (EL) that supplies lexical items. As such, in code-switched adjective-noun constructions, the MLF (Myers-Scotton, 1993, 2002) posits that the adjective position should be compatible with the word order of the matrix language. Another approach (Cantone & MacSwan, 2009), rooted in the MP (MacSwan, 1999), argues that code-switching should be governed by the same grammatical principles as monolingual syntax, with grammatical features determined by the properties of individual lexical items. Thus, following Cinque's (1994, 1999, 2005) proposal that adjectives are universally placed before nouns, Cantone and MacSwan (2009) reached the descriptive generalization that in adjective-noun switching, the adjective placement should be determined by the language of adjectives. A key aspect to mention is that, although these two models capture different aspects of adjective-noun patterns and differ in assumptions, their predictions in explaining these patterns sometimes converge or diverge. Specifically, certain sentences may be explained by both models (MLF+/MP+), by only one of them (e.g., MLF+/MP− or MLF−/MP+), or by neither (MLF−/MP−).

Several studies examining predictions of these two models in fully explaining grammatical patterns in adjective-noun code-switching have failed to find compelling evidence in support of either model (see Section 6.2.2 for detailed descriptions). For example, studies using naturalistic conversations and/or elicitation tasks, such as those in English–Welsh (Parafita Couto et al., 2015), Welsh–English, Spanish–English (Miami), and Papiamento–Dutch (Parafita Couto & Gullberg, 2019), Spanish–English (Balam & Parafita Couto, 2019) and Spanish–Dutch/Papiamento–Dutch (Van Osch et al., 2023), provided support for predictions of both models. While other studies, such as those in French–Dutch (Vanden Wyngaerd, 2017), Welsh–English (Parafita Couto et al., 2017; Vaughan-Evans et al., 2020), Spanish–English (Stadthagen-González et al., 2019) and Papiamento–Dutch (Pablos et al., 2019), showed that CS adjective-noun patterns could be explained by either the MLF or MP, or neither of them. These findings suggest that while each model captures different aspects of adjective-noun patterns in CS, neither fully accounts for the complexity observed. This underscores the need for further research, particularly with understudied language combinations, to more thoroughly evaluate these theoretical frameworks and refine our understanding of adjective-noun switching.

With this in mind, in this study, we investigate adjective-noun CS patterns in mixed nominal constructions with an underexplored language pair and bilingual community. Specifically, we examine Spanish–Chinese mixed constructions produced by early Spanish–Chinese bilinguals in Barcelona, Spain, to evaluate the MLF and MP predictions regarding adjective-noun order. Using a multitask approach, we aim to not only test these models but also uncover new insights into CS behaviors in this unique linguistic context.

6.2 Theoretical background

6.2.1 Theoretical models

6.2.1.1 The Matrix Language Frame (MLF) model

According to the Matrix Language Frame (MLF) model (Myers-Scotton, 1993, 2002), code-switching is structurally asymmetrical: one language, the Matrix Language (ML), provides the morphosyntactic frame, while the Embedded Language (EL) contributes content elements (see Table 6.2.1). A fundamental concept in the MLF is the distinction between content and system morphemes. Content morphemes, such as nouns, verbs, adjectives, and prepositions, carry semantic and pragmatic meaning and receive thematic roles (Myers-Scotton, 2002). On the other hand, system morphemes, including function words and inflections, relate to constituent structure and indicate relations between content morphemes. Typically, system morphemes come from the ML, while content morphemes can be sourced from both the ML and EL (Myers-Scotton, 2002). To determine the ML, Myers-Scotton (1993) proposed two principles based on distinctions of content and system morphemes:

The System Morpheme Principle: "In ML+EL constituents, all system morphemes which have grammatical relations external to their head constituent (i.e., which participate in the sentence's thematic role grid) will come from the ML."

The Morpheme Order Principle: "In ML+EL constituents consisting of singly occurring EL lexemes and any number of ML morphemes, surface morpheme order (reflecting surface syntactic relations) will be that of the ML."

(Myers-Scotton, 1993:83)

In code-switched clauses, both principles apply simultaneously, with the ML determining word order and providing inflections and function words (Myers-Scotton, 2002). The model predicts that finite verb morphology and clause word order will follow the ML. If the bound morphology of the finite verb is from language A, then the noun phrase word order, including adjective placement, should also match language A. For instance, with Chinese as the ML, adjectives should precede nouns, whether in Chinese or Spanish; with Spanish as the ML, adjectives should follow nouns, regardless of their language (see examples 5a–6b in Table 6.2.1).

Table 6.2.1. Example CS sentences align with the MLF predictions.

CS contexts			
Chinese as matrix language		**Spanish as matrix language**	
Examples	Predicted acceptability	Examples	Predicted acceptability
5a. 我妈妈有一只**淘气的** **gato**. My mom has a **naughty cat**.	**MLF+**/MP+	6a. Mi mamá tiene un **猫** **travieso**. My mom has.3S a$_{Masc}$ **cat naughty.**	**MLF+**/MP+
5b. 我妈妈有一只 **travieso** **猫**. My mom has a **naughty cat**.	**MLF+**/MP–	6b. Mi mamá tiene un **gato** **淘气的**. My mom has.3S a$_{Masc}$ **cat naughty.**	**MLF+**/MP–
Translation [English]			
My mom has a naughty cat.			

Note: MLF+/MP+: sentences accepted by both MLF and MP; MLF+/MP–: sentences accepted by only MLF.

6.2.1.2 The Minimalist Program (MP)

MacSwan (1999) applied the Minimalist Program (MP) (Chomsky, 1995) to explain CS grammars. Rather than proposing constraints specific to code-switching, MacSwan (2005) suggested that the phenomenon is constraint-free and can be explained by general grammatical principles, as "Nothing constrains code-switching apart from the requirements of the mixed grammars" (MacSwan, 1999:146). Regarding adjective-noun order, Cantone and MacSwan (2009) follow Cinque's (1994, 1999, 2005) proposal that the adjective placement follows a Universal Base, where all adjectives are universally placed before nouns. According to Cantone and MacSwan (2009), in languages with strong morphological features such as case or φ features (e.g., number, person, and gender), nouns may undergo overt movement across the adjectives to check strong agreement features. In this framework, such overt movements are triggered when strong case or φ features must be evaluated before pronunciation, resulting in visible word order changes within the determiner phrase (DP; Cantone & MacSwan, 2009). Accordingly, postnominal placements in Romance languages such as Spanish likely arise from overt noun movement preceding adjectives triggered by strong φ features. In Cantone and MacSwan's (2009) proposal, the adjective placement in code-switched NPs aligns with the syntactic constraints of its source language. By analysing Italian–German naturalistic data, they observed that adjectives from Italian are consistently positioned after nouns due to the noun movement, regardless of the ML, while German adjectives are placed prenominally. Thus, they reach the descriptive generalization that in the NP-internal code-switching, adjective language determines word order. Accordingly, adjectives should be placed before nouns when they are Chinese (see 7a and 7b with the highlighted code-switched constructions in Table 6.2.2) and after nouns when they are Spanish (see 8a and 8b), regardless of the ML in the clause.

Table 6.2.2. Example CS sentences align with the MP predictions.

CS contexts			
Chinese as matrix language		**Spanish as matrix language**	
Examples	Predicted acceptability	Examples	Predicted acceptability
7a. 我妈妈有一只**淘气的 gato**. My mom has a **naughty cat**.	MLF+/**MP+**	7b. Mi mamá tiene un **淘气的 gato**. My mom has.3S a_{Masc} **naughty cat**.	MLF–/**MP+**
8a. 我妈妈有一只**猫 travieso**. My mom has a **cat naughty**.	MLF–/**MP+**	8b. Mi mamá tiene un **猫 travieso**. My mom has.3S a_{Masc} **cat naughty**.	MLF+/**MP+**
Translation [English]			
My mom has a naughty cat.			

Note: MLF+/MP+: sentences accepted by both MLF and MP; MLF–/MP+: sentences accepted by only MP.

6.2.2 Previous studies in testing MLF and MP predictions in mixed adjective-noun patterns

Several attempts have been made to evaluate these two theoretical frameworks in accounting for adjective-noun code-switching across bilingual communities with different language pairs (see an overview in Appendix 6.A). However, findings of these studies posed challenges for both models. On the one hand, studies based on (semi-)naturalistic data generally reported preferences for noun insertion, a pattern that is commonly aligned with both models, although there remain cases that cannot be adequately explained by either model. For instance, Parafita Couto et al. (2015) used corpora and elicitation data to examine English–Welsh adjective-noun switches and found a preference for noun insertions, which align with both the matrix language (MLF) and the language of adjectives (MP). Similarly, Parafita Couto and Gullberg (2019) analyzed three corpora, Welsh–English, Spanish–English (Miami), and Papiamento–Dutch, and observed that adjective-noun patterns conformed to both models. Balam and Parafita Couto (2019) also found that most cases of Spanish–English adjective-noun constructions in the Northern Belize corpus aligned with both models. Additionally, Van Osch et al. (2023) examined adjective-noun switches in Spanish–Dutch and Papiamento–Dutch in an elicitation task and found a clear preference for noun insertion, which aligned with both models, although they pointed out that either the ML or the adjective language, or both, play roles in determining word order. However, Bierings et al. (2019), in an elicitation study of Kaqchikel–Spanish adjective-noun switches, found that these patterns could not be explained by either model, although they attributed these results to task effects and methodological issues.

On the other hand, studies with other methodologies revealed different CS patterns in adjective-noun switching. For example, French–Dutch adjective-noun code-switching in Vanden Wyngaerd (2017), which used a 3-point grammatical judgment task, found the MP better explained adjective-noun patterns, although the potential role of ML and noun insertions was also observed. However, Stadthagen-González et al. (2019) employed a 2-Alternative Forced Choice (2AFC) judgment task to examine the acceptability of adjective-noun order in Mexican Spanish–American English and found neither the MLF nor MP alone could fully explain adjective-noun patterns. Instead, their results revealed an additive effect, suggesting that both the ML and the language of adjectives influence word order acceptability. Further, Olson (2024) replicated the study of Stadthagen-González et al. (2019) in a 2AFC judgment task and found comparable findings. Additionally, Parafita Couto et al. (2017) used event-related brain potentials (ERPs) to investigate Welsh–English adjective-noun switches and observed the left anterior negativity (LAN), an ERP component sensitive to early grammatical processing, in conditions predicted by the MLF to induce a violation of adjective position. They argued that these findings may support the MLF over MP, though they refrained from making definitive conclusions. Similarly, Vaughan-Evans et al. (2020) examined Welsh–English adjective-noun switches with behavioral and ERPs measures based on Parafita Couto et al. (2017). Their behavioral results indicated that sentences adhering to the MLF assumptions were more likely to be acceptable. While electrophysiological data revealed a greater LAN elicited by MP violations, significant LAN and P600, an ERP component associated with global grammatical findings, were elicited by MLF violations, providing support for the MLF over the MP at a global sentence processing level. Conversely, Pablos et al. (2019) investigated Papiamento–Dutch adjective-noun switches using ERPs measures to detect LAN, and found no LAN modulation elicited by the adjectives in sentences that include MLF or MP violations. Their null results may suggest either that both noun-adjective orders are acceptable in modification contexts or that all tested code-switched patterns were generally dispreferred.

Overall, these findings suggest that bilinguals may not only integrate information from the clausal structure but also be sensitive to the morphosyntactic properties of individual lexical items when processing and producing mixed NPs. In other words, these studies did not provide compelling evidence for either model in fully explaining adjective-noun patterns. Additionally, preferences for noun insertion over adjective insertion were identified across these studies. Moreover, these findings primarily stem from well-established bilingual or multilingual communities (e.g., Spanish–English in Miami or Papiamento–Dutch in the Netherlands), while research on adjective-noun order in relatively young bilingual communities remains absent. Building on this, the present study explores how bilinguals in a relatively young yet stable bilingual community process grammatical constraints proposed by two theoretical models in constructing mixed adjective-noun nominal constructions, aiming to contribute to a more comprehensive understanding of bilingual syntactic integration.

6.2.3 Spanish–Chinese bilingual community in Spain

The most representative young yet stable bilingual community is the Spanish–Chinese bilingual community in Barcelona, Spain, which offers a unique and valuable context for exploring CS habits and patterns. As a relatively young community, it brings together rich cross-linguistic input, consistent heritage language maintenance, and diverse multilingual interactions, which ensure the vibrant use of Chinese and Spanish in this community and make it especially suitable for investigating how bilinguals manage and switch between typologically distinct language systems. The Chinese community, as the second-largest non-EU immigrant group in Spain (Robles-Llana, 2018), has its roots in the largest wave of Chinese migration to Spain in the 1980s (Beltrán Antolín & López, 2013), which was primarily driven by economic and political reasons. Since then, the Chinese population, 70% of whom migrated from Qingtian, Zhejiang province, has grown steadily and reached the second-largest non-EU immigrant group by 2016 (Robles-Llana, 2018; Beltrán Antolín, 2006). In Barcelona, the Chinese population increased from 13,416 in 2003 to 56,017 in 2020, with children under 16 (second generations) rising from 2,412 to 12,285 (He, 2024), fostering the emergence and growth of Spanish–Chinese bilingual communities. Most first-generation immigrants in this community came from the same region, worked in similar jobs (i.e., family-run restaurants), and preserved Chinese cultural values, creating a stable and homogenous environment for Chinese heritage language transmission to second generations.

Additionally, these second generations often act as linguistic and cultural mediators due to their parents' limited integration into broader Spanish society (Robles-Llana, 2018), and their upbringing in a multilingual environment has also led them to routinely interact with Chinese and Spanish monolinguals, Spanish–Catalan bilinguals, and other bi/multilinguals. This exposure has resulted in their high proficiency in Chinese, Spanish, and other languages (e.g., Catalan), fostering rich cross-linguistic interactions and adaptive language practices, such as CS, and the ability to manage typologically different language systems. Despite the community being relatively young (with a second-generation average age of 21 years), second-generation bilinguals share a relatively homogeneous linguistic and educational background, ensuring a valuable context for exploring how bilinguals navigate CS patterns.

6.3 Research questions

In this study, we investigate how early Spanish–Chinese bilinguals, heritage speakers who use Chinese at home and Spanish in the broader society (Barcelona, Spain), navigate the syntactic conflict in adjective-noun placement in mixed Spanish (postnominal adjectives)–Chinese (prenominal adjectives) constructions. Based on this, we further examine how grammatical constraints derived from the two theoretical models, the MLF and MP, influence the construction of mixed adjective-noun patterns. Additionally, given that previous studies have found a preference for noun insertions, we explore whether bilinguals similarly favor the insertion of nouns over adjectives. Building on these objectives, our research questions are as follows:

1. What code-switching patterns between nouns and adjectives will early Spanish–Chinese bilinguals favor in production and

comprehension? Specifically, is there a stronger preference for noun insertions compared to adjective insertions?

2. What insights do the MLF and the MP offer for understanding adjective placement in mixed adjective-noun constructions among early Spanish–Chinese bilinguals, and how can these insights help refine our understanding of CS patterns?

6.4 Materials and methods

6.4.1 Participants

Thirty early Spanish–Chinese bilinguals (age_{mean}= 20.5 years, age_{SD} = 1.66 years; 20 females) were recruited from Pompeu Fabra University in Barcelona, Spain (see Table 6.4.1). Participants completed a sequence of tasks, beginning with a production task and a comprehension task, followed by a questionnaire based on the Bilingual Code-Switching Profile (BCSP; Olson, 2022). The questionnaire assessed participants' language proficiency, experience, and language use. Based on the results of BCSP, participants reported being born either in Spain (n = 18) or China (n = 10), with two exceptions: one participant was born in France but moved to Spain immediately after birth, and another was born in Italy, spent four years in China shortly after birth, and has since lived in Spain. All participants are Chinese heritage speakers raised in Chinese-speaking families, acquiring Chinese after birth, while growing up in a Spanish-speaking society (i.e., learning/speaking Spanish outside home and at schools) with an average age of acquiring Spanish of 3.47 years old (*SD* = 3.42). Moreover, they reported their daily use of Spanish ($M_{frequency}$ = 41.9%, $SD_{frequency}$ = 0.156), Chinese ($M_{frequency}$ = 40.7%, $SD_{frequency}$ = 0.170), Catalan ($M_{frequency}$ = 9.12%, $SD_{frequency}$ = 0.059), and English ($M_{frequency}$ = 6.18%, $SD_{frequency}$ = 0.053). The study was approved by the Ethics Committee of the Faculty of Humanities at Leiden University, and all participants provided informed consent before taking part in the experiments. Upon completing all tasks, participants received monetary compensation for their participation.

Table 6.4.1. Participant characteristics for production and comprehension tasks.

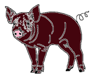
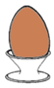
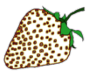
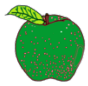
	Experiments 1 and 2
Number of Female/Male participants	20/10
Mean age in years (SD; range)	20.5 (1.66;18–24)
Number of participants born in Spain/Italy/France	18/1/1
Number of participants born in China	10
Age of Chinese acquisition	After birth
Age of Spanish acquisition	3.47 years (*SD* = 3.42)
Daily use of Chinese (frequency)	40.7% (*SD* = 0.170)
Daily use of Spanish (frequency)	41.9% (*SD* = 0.156)
Daily use of Catalan (frequency)	9.12% (*SD* = 0.059)
Daily use of English (frequency)	6.18% (*SD* = 0.053)
Participants without living experience in China (i.e., more than one year)	8

6.4.2 Production task: Director-Matcher task

6.4.2.1 Materials

Twenty-one target-colored line drawings were obtained from the Multipic database (Duñabeitia et al., 2018) based on the following procedures: first, we selected nine colored pictures that are related to familiar, concrete, and easily recognizable features as basic objects; second, each picture was modified by altering properties such as color, size, or condition, resulting in one or two new variants per original picture. This process generated a final set of 21 distinct target pictures with different properties (see Table 6.4.2). For detailed materials, including target pictures, sentences, task results, and questionnaires, please refer to the supplementary materials available at OSF.

Table 6.4.2. Objects used in the director-matcher task.

Target objects		
Grey cat 	Pink pig 	Yellow glove 
Orange cat 	Black pig 	Black glove 
White cat 	Brown pig 	Brown glove 
Orange balloon 	Whole egg 	Big potato 
Blue balloon 	Broken egg 	Small potato 
Green skirt 	Red strawberry 	Red apple 
Black skirt 	White strawberry 	Green apple 

6.4.2.2 Procedure

The experiment involved two sessions, i.e., a familiarization and an experimental session. During familiarization, participants were introduced to the Spanish and Chinese names of nine basic pictures. This step ensured that all participants were equally familiar with the visual stimuli and their corresponding labels before proceeding to the experimental tasks.

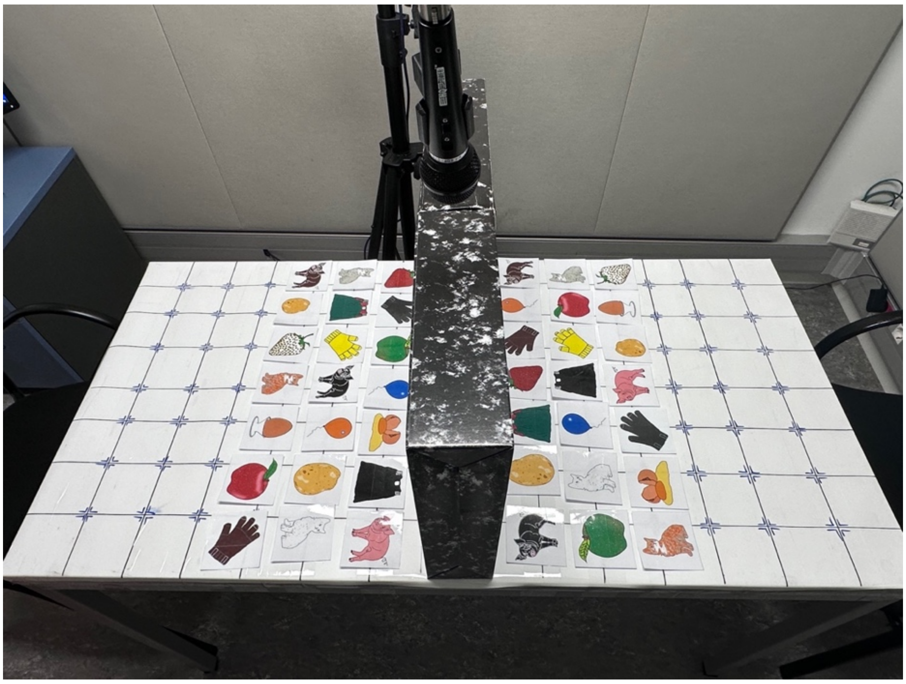
The experimental session consisted of four director-matcher tasks (Gullberg et al., 2009), administered across two language contexts: code-switching (i.e., Chinese matrix language and Spanish matrix language) and unilingual (i.e., Chinese and Spanish). In CS contexts, participants completed two tasks where mixed constructions occurred (i.e., adjectives from one language and nouns from another). They were instructed to name objects in a language different from the matrix language, which was either Chinese or Spanish. Specifically, when Chinese served as the matrix language, objects were named in Spanish, and vice versa. In unilingual contexts, participants performed two tasks exclusively in Spanish or Chinese. The order of tasks within each context was randomized, and participants completed four toy tasks sequentially, with CS contexts always preceding unilingual contexts.

In each language context, two participants were randomly assigned the roles of "director" and "matcher". They sat opposite each other at a table separated by a hardboard and were presented with a grid of 21 target objects that were identical but arranged differently on each side (see Figure 6.4.1). In general, "directors" instructed "matchers" to rearrange objects to match their grid by describing details of each object (e.g., position, color, size, or condition). After completion, "matchers" described the grid with detailed features for verification.

CS contexts: CS instructions were played at the beginning of tasks to elicit participants to communicate in CS mode. Participants were instructed to communicate in Spanish and name objects in Chinese or vice versa, eliciting mixed adjective-noun constructions (i.e., Spanish adjectives with Chinese nouns in Spanish ML or Chinese adjectives with Spanish nouns in Chinese ML).

Unilingual contexts: Participants communicated and named objects in either Spanish or Chinese to assess unilingual adjective-noun word order.

Figure 6.4.1. Illustration of the setup of the director-matcher task.



6.4.3 Comprehension task: Acceptability judgment using a two-alternative forced-choice task

6.4.3.1 Materials

Critical sentences

Seventy-two CS sentences (i.e., thirty-six each for Spanish ML and Chinese ML) and thirty-six unilingual sentences (i.e., eighteen each for unilingual Spanish and Chinese) were generated. Nine base sentences were first created in Spanish, each featuring an unilingual adjective-noun construction. These were then translated into Chinese, resulting in nine equivalent Chinese sentences. In both languages, nouns referred to the nine target pictures from the production task, while adjectives described a single property of each picture (color, size, or condition). Each construction included only one adjective.

***CS contexts*:** In CS contexts, target nouns in the nine Spanish base sentences were replaced with their Chinese equivalents, and vice versa for the Chinese base sentences. This resulted in nine Spanish sentences with inserted Chinese nouns (i.e., Chinese noun insertion) and nine Chinese sentences with inserted Spanish nouns (i.e., Spanish noun insertion). Similarly, adjectives in the nine Spanish base sentences were replaced with their Chinese equivalents, and vice versa for the Chinese base sentences. This process resulted in nine Spanish sentences with inserted Chinese adjectives (i.e., Chinese adjective insertion) and nine Chinese sentences with inserted Spanish adjectives(i.e., Spanish adjective insertion). These CS sentences, featuring noun and adjective insertions, were then modified according to four possible code-switching patterns based on the MLF (Myers-Scotton, 1993, 2002) and MP (MacSwan, 1999), namely:

MLF+/MP+: the CS pattern was predicted by both the MLF and the MP
MLF+/MP–: the CS pattern was predicted by the MLF (but not by the MP)
MLF–/MP+: the CS pattern was predicted by the MP (but not by the MLF)
MLF–/MP–: the CS pattern was predicted by neither the MLF nor the MP

This yielded a total of 72 CS sentences, i.e., 36 for the Spanish ML and 36 for the Chinese ML (see Appendix 6.B for a full list of the sentences). Each sentence featured only one instance of code-switching: either a noun or an adjective inserted from the non-matrix language (i.e., Chinese in Spanish ML sentences, or Spanish in Chinese ML sentences; see Table 6.4.3). Sentences representing different CS patterns were compared pairwise in all possible combinations, with the number of pairwise comparisons for each CS base sentence calculated using the formula n*(n-1)/2 (i.e., n is the number of CS patterns for each CS base sentence) (Stadthagen-González et al., 2018). Thus, each CS base sentence generated 6 pairwise comparisons and then generated 108 pair comparisons for all 18 CS base sentences (i.e., 54 each for Chinese ML and Spanish ML).

Unilingual contexts**:** In unilingual contexts, the nine base sentences in both Spanish and Chinese were further modified to reflect adjective positions within adjective-noun constructions.

This yielded a total of 36 unilingual sentences, i.e., 18 in Spanish and 18 in Chinese. These sentences were compared in pairs, with the contrast in the placement of adjectives either before or after nouns. All comparisons only occurred within the same language, yielding a total of 18 pair comparisons, i.e., nine in Spanish and nine in Chinese.

Table 6.4.3. Example critical sentence in different language contexts.

CS contexts			
Chinese as matrix language		**Spanish as matrix language**	
Examples	Predicted acceptability	Examples	Predicted acceptability
我妈妈有一只**淘气的 gato**.	MLF+/MP+ (Pattern A)	Mi mamá tiene un **猫 travieso**.	MLF+/MP+ (Pattern A)
我妈妈有一只 **travieso 猫**.	MLF+/MP– (Pattern B)	Mi mamá tiene un **gato 淘气的**.	MLF+/MP– (Pattern B)
我妈妈有一只**猫 travieso**.	MLF–/MP+ (Pattern C)	Mi mamá tiene un **淘气的 gato**.	MLF–/MP+ (Pattern C)
我妈妈有一只 **gato 淘气的**.	MLF–/MP– (Pattern D)	Mi mamá tiene un **travieso 猫**.	MLF–/MP– (Pattern D)
Unilingual contexts			
Chinese unilingual sentences		**Spanish unilingual sentences**	
我妈妈有一只**淘气的**猫.	Pre-nominal	Mi mamá tiene un gato **travieso**.	Post-nominal
我妈妈有一只猫**淘气的**.	Post-nominal	Mi mamá tiene un **travieso** gato.	Pre-nominal
Translation [English]			
My mom has a naughty cat.			

Filler sentences

Sixty-four CS filler sentences and sixty-four unilingual Chinese filler sentences were included, focusing on classifier contrasts in mixed and unilingual NPs. The outcomes of these filler items fall outside the scope of the present study and will be reported in other studies. Sentences in each mode were compared pairwise, generating 64 comparisons, i.e., 32 for CS mode and 32 for unilingual mode.

6.4.3.2 Procedure

This experiment was administered online using Qualtrics (Qualtrics, LLC, Provo, UT, USA). Participants completed tasks sequentially in CS and unilingual contexts. In CS contexts, they were presented with 108 pair comparisons (i.e., 54 for Chinese ML and 54 for Spanish ML) and instructed to choose the sentence that felt more natural when speaking to other bilinguals, even if both sentences seemed equally natural or unnatural. In unilingual contexts, they were instructed to choose one sentence based on the same criterion as in CS contexts. In each context, participants had to make a choice for each comparison before proceeding to the next, without the option to revisit previous pairs. The experiment followed a by-subject order design, with the order of sentence pairs and sentences within each pair randomized across participants.

6.5 Results and analyses

6.5.1 Production task: director-matcher task

The recorded target constructions from both unilingual and CS contexts were transcribed. Unilingual and mixed adjective-noun constructions displaying identified word order were extracted for further analysis. In CS contexts, a total of 1,260 trials were initially collected, with 630 trials from each matrix language. Of these, 244 trials, i.e., 242 from Spanish ML and 2 from Chinese ML, were excluded due to: (1) unexpected adjective-noun structures, i.e., N + de + prepositional phrases (PP), which differ in underlying syntactic structures and constraints from N + Adj (Bosque & Picallo, 1996), and do not directly represent a syntactic conflict of adjective placements in Spanish and Chinese; (2) or single noun production. This resulted in 1,016 valid trials for analysis: 388 for Spanish ML and 628 for Chinese ML. In unilingual contexts, a total of 1,260 unilingual trials were collected (630 per language), of which 126 Spanish trials (10%) were excluded due to the same reason, yielding 1,134 valid trials, i.e., 504 in Spanish and 630 in Chinese.

6.5.1.1 Unilingual contexts: unilingual sentences

Spanish unilingual sentences: Participants produced a total of 504 trials with Spanish unilingual sentences, featuring Spanish constructions (i.e., Spanish nouns paired with Spanish adjectives). A consistent pattern was observed in these Spanish unilingual sentences:

1. Spanish nouns + Spanish adjectives (postnominal) (504 of 504, 100%)

(1)	el	gato	naranja
	the	cat	orange
	DET	N	ADJ

(from speaker 22M)

Chinese unilingual sentences: A total of 630 unilingual Chinese constructions (i.e., Chinese nouns paired with Chinese adjectives) were obtained. Two distinct patterns were identified:

2. Chinese adjectives + Chinese nouns (prenominal) (627 of 630; 99.52%)

(2)	黑色的	裙子
	black	skirt
	ADJ	N

(from speaker 21D)

3. Chinese nouns + Chinese adjectives (postnominal) (3 of 630; 0.48%)

(3) 裙子 绿色的
skirt green
N ADJ

(from speaker 27D)

The majority of adjective-noun constructions in unilingual Spanish and Chinese aligned with the default word order of each language: postnominal adjectives in Spanish and prenominal adjectives in Chinese. However, an exception was observed in unilingual Chinese, where 3 trials (0.48%) deviated from this pattern, all produced by the same participant (Speaker 27D).

6.5.1.2 CS contexts: CS sentences in Spanish matrix language and Chinese matrix language

Spanish matrix language: For Spanish ML with mixed adjective-noun constructions (i.e., Spanish adjectives with Chinese nouns), participants produced a total of 388 trials with identifiable word order. Four main patterns were observed:

4. Chinese nouns + Spanish adjectives (MLF+/MP+) (361 of 388; 93.04%)

(4) una 苹果 verde
an apple green
ART N ADJ

(e.g., speaker 29D/30M)

5. Spanish adjectives + Chinese nouns (MLF−/MP−) (25 of 388; 6.44%)

(5) un rosa 猪
a pink pig
ART ADJ N

(from speaker 20M)

6. Spanish nouns + Spanish adjectives (no CS-postnominal) (1 of 388; 0.26%)

(6) una patata pequeña
a potato small
ART N ADJ

(from speaker 01D)

7. Spanish adjectives + Spanish nouns (no CS-prenominal) (1 of 388; 0.26%)

(7) un negro cerdo
a black pig
ART ADJ N

(from speaker 09D)

Chinese matrix language: For Chinese ML with mixed adjective-noun constructions (i.e., Chinese adjectives with Spanish nouns), a total of 628 trials were yielded. Three patterns were observed:

8. Chinese adjectives + Spanish nouns (MLF+/MP+) (560 of 628; 89.17%)

(8) 红色的 fresa
red strawberry
ADJ N

(e.g., speaker 07D)

9. Spanish nouns + Chinese adjectives (MLF−/MP−) (63 of 628; 10.03%)

(9) globo 蓝色的
balloon blue
N ADJ

(from speaker 27D)

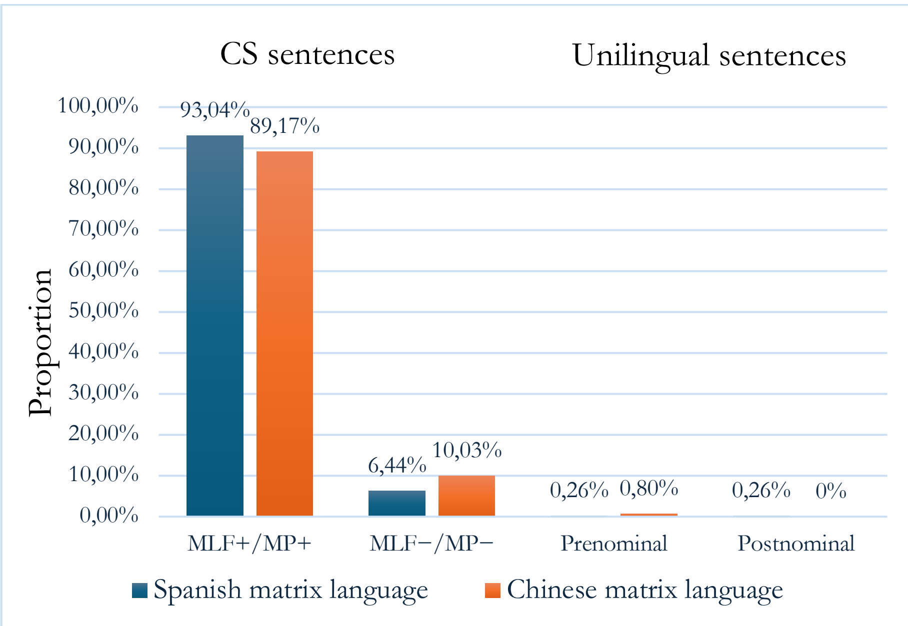
10. Chinese adjectives + Chinese nouns (no CS-prenominal) (5 of 628; 0.8%)

(10) 裂开的 鸡蛋
broken egg
ADJ N

(from speaker 28M)

For both Chinese and Spanish matrix language sentences, the majority of the mixed adjective-noun constructions involved Chinese or Spanish noun insertions, which can be predicted by both the MP and MLF (see Figure 6.5.1). However, neither the MP nor MLF model can account for the occurrence of pattern (5), in which Spanish adjectives were used prenominally, or pattern (9), where Chinese adjectives were used postnominally. Notably, two unilingual Spanish and five unilingual Chinese constructions were produced in CS contexts.

Figure 6.5.1. Distribution of nominal constructions in Spanish and Chinese matrix languages within CS contexts.



6.5.2 Comprehension task: acceptability judgment using a two-alternative forced-choice task and Thurstone's law

6.5.2.1 Unilingual contexts: unilingual sentences

Spanish unilingual sentences: Participants chose postnominal adjectives, which is the natural word order in Spanish, in 93.7% of the trials (253 out of 270 trials). The remaining 6.3% of the trials featured prenominal adjectives, and these occurrences were dispersed across participants and target objects, rather than being concentrated in a single speaker or construction.

Chinese unilingual sentences: Participants showed a strong preference for prenominal adjectives, consistent with the default word order in Chinese, in 98.89% of the trials (267 out of 270 trials). A very small proportion (1.11%, 3 trials) featured postnominal adjectives, produced by three different speakers and involving three different target objects.

6.5.2.2 CS contexts: CS sentences

Participants' responses in CS contexts were first analyzed using Thurstone's law of Comparative Judgment, Case V (Thurstone, 1994). This analysis examines the pairwise comparisons made by participants to generate a ranking of preferences among conditions and a measure of their relative comparison. The measure values are relative to the lowest acceptable pattern, conventionally set to zero (Stadthagen-González et al., 2018). For detailed explanations, refer to Thurstone (1994) and Stadthagen-González et al. (2018). Additionally, confidence intervals (CIs) were calculated using Montag's (2006) method for analyzing paired comparison data. Notably, participants' responses in the Spanish matrix language task and the Chinese matrix language task were analyzed separately.

Spanish matrix language

Table 6.5.1 summarizes the results of Thurstone's measure for each condition in the Spanish matrix language task. Notably, condition B (MLF+/MP–) with a score of 0 indicates that it is the least favored option compared to other conditions. The 95% confidence interval for the Spanish matrix language data was 0.07. For detailed calculation steps of Thurstone's law and confidence interval, please refer to the supplementary materials.

Table 6.5.1. Ranking and Thurstone measures for sentences presented in Spanish as matrix language.

Rank	Pattern	Construction	Type of insertions	Example	Thurstone measure
1	C. MLF–/MP+	ADJ_{CN}+N_{SP}	Adjective insertion	Barry dibujó una 绿色的 fresa.	2.28
2	A. MLF+/MP+	N_{CN}+ADJ_{SP}	Noun insertion	Barry dibujó una 草莓 verde.	2.03
3	D. MLF–/MP–	ADJ_{SP}+N_{CN}	Noun insertion	Barry dibujó una verde 草莓.	0.13
4	B. MLF+/MP–	N_{SP}+ADJ_{CN}	Adjective insertion	Barry dibujó una fresa 绿色的.	0.00

A within-subjects ANOVA was then conducted to statistically analyze the preference for each code-switching pattern. The result revealed a significant effect of different patterns on participants' responses in the Spanish matrix language, $F(3, 87) = 75.50$, $p < .001$, $\eta^2 = .72$. Post-hoc pairwise comparisons using Bonferroni correction indicated all comparisons were significantly different (all p values < .001), except for the contrast between pattern C (i.e., adjective insertion, ADJ_{CN}+N_{SP}) and pattern A (i.e., noun insertion, N_{CN}+ADJ_{SP}), as well as the contrast between patterns D (ADJ_{SP}+N_{CN}) and B (N_{SP}+ADJ_{CN}).

As shown in Table 6.5.1, although Spanish–Chinese bilinguals showed the strongest preference for condition C (ADJ_{CN}+N_{SP}), where adjective insertion is predicted only by the MP, followed by condition A

(N_{CN}+ADJ_{SP}), where noun insertion aligns with both the MLF and MP, post-hoc pairwise comparisons revealed no significant difference between them ($p = 1$, $t = 1.2$). Similarly, the least preferred constructions, condition D (ADJ_{SP}+N_{CN}) and B (N_{SP}+ADJ_{CN}), also showed no significant difference ($p = 1$, $t = 0.64$). Overall, participants showed preferences for patterns involving noun insertions (Condition A, N_{CN}+ADJ_{SP}), where adjectives follow both the matrix language (MLF) and the language of adjectives (MP), as well as adjective insertions (Condition C, ADJ_{CN}+N_{SP}), where adjectives align only with the adjective language.

Chinese matrix language

Following the same procedures, we calculated Thurstone's measure (see Table 6.5.2) and a within-subjects ANOVA for comparing patterns in the Chinese matrix language. The 95% confidence interval in Thurstone's measure was 0.07. A within-subjects ANOVA showed a significant effect of code-switching patterns, $F(3, 87) = 115.50$, $p < .001$, $\eta^2 = .80$, indicating that the different conditions significantly influenced participants' acceptability of adjective-noun constructions in the Chinese matrix language. Post-hoc Bonferroni corrections indicated all comparisons were highly significantly different (all p values < .001), except for the contrasts between patterns B (i.e., MLF+/MP–, ADJ_{SP}+N_{CN}) and D (i.e., MLF–/MP–, N_{SP}+ADJ_{CN}), as well as between patterns B (i.e., MLF+/MP–, ADJ_{SP}+N_{CN}) and C (i.e., MLF–/MP+, N_{CN}+ADJ_{SP}).

Table 6.5.2. Ranking and Thurstone measure for sentences presented in Chinese as matrix language.

Rank	**Pattern**	**Construction**	**Type of insertion**	**Example**	**Thurstone measure**
1	A. MLF+/MP+	ADJ_{CN}+N_{SP}	Noun insertion	巴瑞画了一颗绿色的 fresa.	4.20
2	C. MLF–/MP+	N_{CN}+ADJ_{SP}	Adjective insertion	巴瑞画了一颗草莓 verde.	1.24
3	B. MLF+/MP–	ADJ_{SP}+N_{CN}	Adjective insertion	巴瑞画了一颗 verde 草莓.	0.64
4	D. MLF–/MP–	N_{SP}+ADJ_{CN}	Noun insertion	巴瑞画了一颗 fresa 绿色的.	0.00

Table 6.5.2 shows that Spanish–Chinese bilinguals differed in their acceptability of adjective-noun constructions across four conditions. They showed a marked preference for Condition A, which involved noun insertions (prenominal Chinese adjectives with Spanish nouns) and is predicted by both the MLF and MP, compared to all other conditions (all p values < .001). Regarding condition C, which features adjective insertions (Chinese nouns with postnominal Spanish adjectives), results suggest comparable acceptability to condition B ($p = 0.184$, $t = 2.27$). Note that condition C is predicted only by MP, whereas condition B is predicted only by MLF. Additionally, bilinguals rejected condition D, which contains Spanish nouns with postnominal Chinese adjectives (noun insertions) and is predicted by neither MLF nor MP. Overall, there was a clear preference for pattern A (i.e., MLF+/MP+, ADJ_{CN}+N_{SP}, noun insertion) over all other patterns in the Chinese matrix language.

6.6 Discussion

In this study, we examined how early Spanish–Chinese bilinguals resolved the syntactic difference in adjective-noun placements between Spanish (postnominal adjectives) and Chinese (prenominal adjectives) in mixed constructions. Rather than solely evaluating the predictions of the MLF and MP, we used these models as starting points to examine how well they account for observed code-switching patterns. Our aim was to assess whether the adjective-noun structures produced in Spanish–Chinese code-switching align with either model's predictions, and to consider how the findings might inform or refine our understanding of code-switching patterns.

Regarding the first research question, our findings revealed distinct adjective-noun code-switching patterns in production and comprehension, varying by matrix languages. In the Spanish matrix language, participants predominantly produced noun insertions (N_{CN}+ADJ_{SP}) in production, while they favored both such noun insertions and adjective insertions (ADJ_{CN}+N_{SP}) in comprehension. In the Chinese matrix language, they consistently preferred prenominal Chinese adjectives with Spanish nouns ($ADJ_{CN}+N_{SP}$) in both tasks, also reflecting noun insertion patterns. Overall, bilinguals favored noun insertions across both matrix languages, aligning with previous findings (e.g., Parafita Couto et al., 2015; Van Osch et al., 2023; Vaughan-Evans et al.,2020).

For the second research question, we evaluated the extent to which the MLF and MP offer explanations for adjective placement in adjective-noun constructions among early Spanish–Chinese bilinguals, and how these explanations refine our understanding of code-switching patterns. In production, most constructions aligned with both models in Spanish (93.04%) and Chinese (89.17%) matrix languages, likely due to the dominance of noun insertions. In comprehension, MP predictions could explain more constructions in the Spanish matrix language, in line with the findings observed in Vanden Wyngaerd (2017) in French–Dutch constructions. In the Chinese matrix language, both the MLF and MP are favored, mirroring observations from Balam and Parafita Couto (2019) with Spanish–English bilinguals in Northern Belize. In sum, participants favored noun insertions aligning with both the matrix language (MLF) and the language of adjectives (MP) in the Chinese matrix language, while they preferred noun insertions aligned with both models and adjective insertions that follow the order of the adjective languages (MP) in the Spanish matrix language. Overall, these findings provide insights into how mixed adjective-noun constructions were structured by Spanish–Chinese bilinguals across languages and tasks. However, from a broader perspective, these findings also suggest that, although the two competing theoretical frameworks, the MLF and MP, make partially different predictions regarding preferred CS adjective-noun patterns, both clausal structure and morphosyntactic properties of individual lexical items play roles in constructing mixed adjective-noun constructions among bilinguals. This also implies that no single grammatical constraint from these theoretical frameworks can fully and solely account for observed patterns. Rather, a more comprehensive understanding of the constructions of CS patterns requires considering the combined influence of both clausal structures and morphosyntactic properties of individual lexical items.

Taking a closer look at our data, matrix language and language of adjectives play a crucial role in adjective-noun order within mixed nominal constructions. In both production and comprehension, when adjectives aligned with matrix languages, participants consistently preferred the adjective-noun order of matrix languages. Conversely, when adjectives were inconsistent with matrix languages, participants tended to follow the adjective-order of the language of adjectives. A more detailed inspection of production tasks revealed that the adjective language overwhelmingly matched the matrix language (623/628 in Chinese ML; 386/388 in Spanish ML). This is likely due to the task design, which explicitly instructed participants to name *only* the object in the other language and likely promoted a high occurrence of noun insertions. This aligned with findings in previous studies (e.g., Van Osch et al., 2023). Additionally, the judgment data revealed a sole preference for inserting Spanish nouns in Chinese contexts, whereas both Chinese nouns and adjectives were more acceptable in Spanish contexts. This suggests that Spanish attributive adjectives are less favored than Chinese adjectives in code-switching nominal constructions. One possible explanation for this asymmetry may be attributed to Spanish attributive adjectives functioning as gender carriers, given that Spanish requires gender agreement between nouns and adjectives at both syntactic and morphological levels (Balam & Parafita Couto, 2019). This finding aligns with prior studies showing gender-marked adjectives are infrequently produced in mixed noun-adjective constructions during CS (Parafita Couto & Gullberg, 2019).

A particularly notable finding from the production tasks is that participants employed two types of nominal constructions with postnominal adjectives in Spanish ML in both unilingual and CS contexts: (1) N + Adj (see example 11) and (2) N + *de*-PP (see example 12). The use of both constructions may stem from the fact that both structures function similarly for classificatory adjectives (i.e., those denote a property of nouns, such as color, size, shape, etc.; see Bosque & Picallo, 1996) to modify head nouns. Another plausible explanation for high occurrences of *N + de-PP* constructions may be attributed to structural priming during production, where one speaker's use of constructions would increase the likelihood that another speaker would repeat or prime them to produce similar constructions (see detailed descriptions below).

11. Chinese nouns + Spanish adjectives

(11)	un	猫	gris
	the$_{MASC}$	cat	grey
	Det	N	Adj

12. Chinese nouns + Spanish prepositional phrases

(12)	un	猫	de color blanco
	the$_{MASC}$	cat	of color white
	Det	N	de-PP

As mentioned above, one issue that needs to be mentioned in production tasks is structural priming in code-switching, where one speaker's code-switching behavior would facilitate similar patterns in another (Fricke & Kootstra, 2016). Analyzing spontaneous English–Spanish code-switching dialogues, Fricke and Kootstra (2016) found that bilinguals' behavior would prime each other regarding the tendency and grammatical frame of code-switching (Fricke & Kootstra, 2016). Similarly, Berghoff et al. (2023) showed that lexical repetition enhances code-switching at points of non-shared word order in Afrikaans–English bilinguals, highlighting the interaction between internal (word order) and external (e.g., priming) factors in code-switching. In our study, we observed comparable structure priming in code-switching between "directors" and "matchers" in paired production tasks, as illustrated in Appendix 6.C. During production, "directors" always described object details (e.g., position, color, size, or condition) to guide "matchers" in re-arranging objects, and "matchers" described the same details to verify the match, eliciting mixed adjective-noun constructions from both sides. This interaction seems to facilitate "matchers" to repeat or produce similar code-switching structures as "directors", potentially reducing the opportunity to observe their spontaneous code-switching patterns.

While our study did not fully capture spontaneous speech or naturally occurring code-switching, previous research has demonstrated that experimental data often reflect patterns similar to those found in corpus-based studies. For instance, in well-established bilingual communities like Papiamento–Dutch communities in the Netherlands, bilinguals often favor the minority language (Papiamento) as the matrix language, with insertions from the majority or societal language (Dutch), while the reverse pattern is less frequently accepted (e.g., see corpora data in Parafita Couto & Gullberg, 2019 and semi-experimental data in Van Osch et al., 2023). Although the Spanish–Chinese bilingual community in Barcelona is still relatively young, bilinguals share a relatively stable and homogeneous linguistic background that supports the regular use of Chinese and Spanish. This may contribute to the emergence of shared community norms and promote more widespread and systematic code-switching behavior. Thus, although our data do not offer definitive evidence regarding the directionality of code-switching, we anticipate that bilinguals in this community may exhibit similar patterns to those observed in more established bilingual communities. This hypothesis would have to be investigated in future research.

6.7 Conclusion

This study is the first to empirically examine how mixed adjective-noun constructions are structured in light of syntactic differences in adjective placement among Spanish–Chinese bilinguals. Our findings provide preliminary evidence that the language of adjectives (as predicted by the MP) plays a crucial role in adjective placement within Spanish matrix language sentences, while bilinguals are sensitive to both clausal structure (as predicted by the MLF) and the languages of adjectives (MP) when constructing mixed adjective-noun patterns in Chinese matrix language sentences. Parallel to trends observed in other bilingual communities, we also found a general preference for noun insertions across ML contexts. Additionally, our results highlight the potential influences of syntactic priming and emerging community norms on bilinguals' adjective-noun CS patterns. Together, these results offer new insights into how bilinguals navigate syntactic differences, particularly the interplay between matrix language and adjective language in shaping adjective-noun order in relatively young bilingual communities. This study contributes valuable data from an understudied bilingual population, advancing our understanding of how syntactic constraints shape adjective-noun CS patterns and laying the groundwork for future research on the directionality and constraints of code-switching in this language pair and community.

Acknowledgements

We would like to express our gratitude to the participants for their time and dedication in completing the experiments, as well as to the members of the Center for Brain and Cognition at Pompeu Fabra University (UPF) for their support. Special thanks to Núria Sebastián Gallés, director of the Speech Acquisition and Perception (SAP) group at UPF, for her warm welcome and generous assistance. We also appreciate the help of Francesc Xavier Mayoral Terron and Silvia Blanch Rabat from the Department of Technology at UPF for their support with the experimental setup and guidance during data collection.

Declaration of conflicting interest

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

This work was partially funded by the China Scholarship Council (CSC) NO. 202107720015.

Data Availability Statement

Supplementary materials supporting the experimental design, stimuli pictures and sentences, and findings of this study are openly available in Open Science Framework at https://osf.io/vpbcw/?view_only=556566ad901248c98e75c3e160857846 (view-only link).

Appendix

Appendix 6.A Overview of studies comparing the MLF and MP models regarding mixed adjective-noun nominal constructions (i.e., this form was built based on Bierings et al., 2019).

Studies	Language pairs	Tasks	Findings regarding MP	Findings regarding MLF
Parafita Couto et al. (2015)	Welsh–English (Wales, UK)	• Naturalistic data (corpora) • Elicitation task	• No convincing evidence for supporting the MP	• Support for the MLF (compared to MP)
		• Auditory judgment task	• Inconclusive	• Need more evidence to draw conclusions
Parafita Couto & Gullberg (2019)	Welsh–English (Wales, UK) Spanish–English (Miami, USA) Papiamento–Dutch (The Netherlands)	• Naturalistic data (three corpora)	• Partly support the MP in all language pairs (less than MLF)	• Support for the MLF in all language pairs
Balam & Parafita Couto (2019)	Spanish–English (Northern Belize)	• Naturalistic data (Sociolinguistic interviews)	• Evidence for supporting the MP (less than the MLF)	• Highly support for the MLF
Van Osch et al. (2023)	Spanish–Dutch (The Netherlands) Papiamento–Dutch (The Netherlands)	• Elicitation task	• Support for the MP (most common patterns are noun insertions)	• Support for the MLF (most common patterns are noun insertions)
Bierings et al. (2019)	Kaqchikel–Spanish (Patzún, Guatemala)	• Elicitation task	• The adjective positions contra the prediction of the MP (may due to task-effect, and need more evidence)	• The adjective positions contra the prediction of the MLF (may due to task-effect, and need more evidence)
Vanden Wyngaerd (2017)	French–Dutch (Brussels, Belgium)	• 3-points grammatical judgment task	• Highly support for the MP (more than the MLF)	• Supported for the MLF, less than the MP
Stadthagen-González et al. (2019)	Spanish–English (Mexicans in the U.S.A)	• 5-point Likert judgment task • 2-Alternative forced choice judgment task	• No particular support for the MP, but combined explanation with the MLF	• No particular support for the MLF, but combined explanation with the MP

Olson (2024)	Spanish–English (Online platform)	• 2-Alternative forced choice judgment task	• Replicated the result of Stadthagen-González et al. (2019)	• Replicated the result of Stadthagen-González et al. (2019)
Parafita Couto et al. (2017)	Welsh–English (Wales, UK)	• EEG study (ERPs data, a sentence verification task)	• No convincing support for the MP	• Support for the MLF (more than MP), need complementary evidence
Vaughan-Evans et al. (2020)	Welsh–English (Wales, UK)	• EEG study (Behavioral and ERPs data)	• No convincing support for the MP	• Clear support for the MLF (noun insertions are common patterns)
Pablos et al. (2019)	Papiamento–Dutch (The Netherlands)	• EEG study (ERPs data, comprehension srudy)	• No particular support for the MP	• No particular support for the MLF (no preference between noun-adjective switches)

Appendix 6.B Sample sentences used in the comprehension task in Chapter 6.

Unilingual base sentences			
Target objects	**Spanish**	**Chinese**	**Translation [English]**
gato/猫 [Cat]	Mi mamá tiene un **gato** travieso.	我妈妈有一只淘气的**猫**.	My mom has a naughty **cat**.
globo/气球[balloon]	Hay un **globo** rosa flotando en el cielo.	天空中飘着一个粉色的**气球**.	There is a pink **balloon** floating in the sky.
cerdo/猪 [pig]	En los cuentos de hadas, Wilbur es un **cerdo** inteligente.	在童话里，威尔伯是一只聪明的**猪**。	In fairy tales, Wilbur is a smart **pig**.
guante/手套[glove]	Mary perdió un **guante** verde.	玛丽丢了一只绿色的**手套**.	Mary lost a green **glove**.
huevo/鸡蛋 [egg]	Tom le dio a su hermana un **huevo** cocido.	汤姆给了他妹妹一个煮熟的**鸡蛋**.	Tom gave his sister a cooked **egg**.
patata/土豆[potato)	Jack vio una **patata** estropeada en la cocina.	杰克在厨房里看到了一个变质的**土豆**.	Jack saw a spoiled **potato** in the kitchen.
falda/裙子 [skirt]	Lisa tiene una **falda** hermosa.	丽莎有一条漂亮的**裙子**.	Lisa has a beautiful **skirt**.
fresa/草莓 [strawberry]	Barry dibujó una **fresa** verde.	巴瑞画了一颗绿色的**草莓**.	Barry drew a green **strawberry**.
manzana/苹果[apple]	Blancanieves se comió una **manzana** venenosa.	白雪公主吃了一个有毒的**苹果**.	Snow White ate a poisonous **apple**.
Code-switched sentences [CS contexts]			
Target objects	**Spanish as matrix language**	**Chinese as matrix language**	**Predicted acceptability**
	Mi mamá tiene un 猫 travieso.	我妈妈有一只淘气的 gato.	MLF+/MP+
	Mi mamá tiene un gato 淘气的.	我妈妈有一只 travieso 猫.	MLF+/MP–

gato/猫 [Cat]	Mi mamá tiene un 淘气的 gato.	我妈妈有一只猫 travieso.	MLF–/MP+
	Mi mamá tiene un travieso 猫.	我妈妈有一只 gato 淘气的.	MLF–/MP–
globo/气球[balloon]	Hay un 气球 rosa flotando en el cielo.	天空中飘着一个粉色的 globo.	MLF+/MP+
	Hay un globo 粉色的 flotando en el cielo.	天空中飘着一个 rosa 气球.	MLF+/MP–
	Hay un 粉色的 globo flotando en el cielo.	天空中飘着一个气球 rosa.	MLF–/MP+
	Hay un rosa 气球 flotando en el cielo.	天空中飘着一个 globo 粉色的.	MLF–/MP–
cerdo/猪 [pig]	En los cuentos de hadas, Wilbur es un 猪 inteligente.	在童话里，威尔伯是一只聪明的 cerdo.	MLF+/MP+
	En los cuentos de hadas, Wilbur es un cerdo 聪明的.	在童话里，威尔伯是一只 inteligente 猪.	MLF+/MP–
	En los cuentos de hadas, Wilbur es un 聪明的 cerdo.	在童话里，威尔伯是一只猪 inteligente.	MLF–/MP+
	En los cuentos de hadas, Wilbur es un inteligente 猪.	在童话里，威尔伯是一只 cerdo 聪明的.	MLF–/MP–
guante/手套[glove]	Mary perdió un 手套 verde.	玛丽丢了一只绿色的 guante.	MLF+/MP+
	Mary perdió un guante 绿色的.	玛丽丢了一只 verde 手套.	MLF+/MP–
	Mary perdió un 绿色的 guante.	玛丽丢了一只手套 verde.	MLF–/MP+
	Mary perdió un verde 手套.	玛丽丢了一只 guante 绿色的.	MLF–/MP–
huevo/鸡蛋 [egg]	Tom le dio a su hermana un 鸡蛋 cocido.	汤姆给了他妹妹一个煮熟的 huevo.	MLF+/MP+
	Tom le dio a su hermana un huevo 煮熟的.	汤姆给了他妹妹一个 cocido 鸡蛋.	MLF+/MP–
	Tom le dio a su hermana un 煮熟的 huevo.	汤姆给了他妹妹一个鸡蛋 cocido.	MLF–/MP+
	Tom le dio a su hermana un cocido 鸡蛋.	汤姆给了他妹妹一个 huevo 煮熟的.	MLF–/MP–

patata/土豆[potato)	Jack vio una 土豆 estropeada en la cocina.	杰克在厨房里看到了一个变质的 patata.	MLF+/MP+
	Jack vio una patata 变质的 en la cocina.	杰克在厨房里看到了一个 estropeada 土豆.	MLF+/MP–
	Jack vio una 变质的 patata en la cocina.	杰克在厨房里看到了一个土豆 estropeada.	MLF–/MP+
	Jack vio una estropeada 土豆 en la cocina.	杰克在厨房里看到了一个 patata 变质的.	MLF–/MP–
falda/裙子 [skirt]	Lisa tiene una 裙子 hermosa.	丽莎有一条漂亮的 falda.	MLF+/MP+
	Lisa tiene una falda 漂亮的.	丽莎有一条 hermosa 裙子.	MLF+/MP–
	Lisa tiene una 漂亮的 falda.	丽莎有一条裙子 hermosa.	MLF–/MP+
	Lisa tiene una hermosa 裙子.	丽莎有一条 falda 漂亮的.	MLF–/MP–
fresa/草莓 [strawberry]	Barry dibujó una 草莓 verde.	巴瑞画了一颗绿色的 fresa.	MLF+/MP+
	Barry dibujó una fresa 绿色的.	巴瑞画了一颗 verde 草莓.	MLF+/MP–
	Barry dibujó una 绿色的 fresa.	巴瑞画了一颗草莓 verde.	MLF–/MP+
	Barry dibujó una verde 草莓.	巴瑞画了一颗 fresa 绿色的.	MLF–/MP–
manzana/苹果[apple]	Blancanieves se comió una 苹果 venenosa.	白雪公主吃了一个有毒的 manzana.	MLF+/MP+
	Blancanieves se comió una manzana 有毒的.	白雪公主吃了一个 venenosa 苹果.	MLF+/MP–
	Blancanieves se comió una 有毒的 manzana.	白雪公主吃了一个苹果 venenosa.	MLF–/MP+
	Blancanieves se comió una venenosa 苹果.	白雪公主吃了一个 manzana 有毒的.	MLF–/MP–
Unilingual sentences [Unilingual mode]			
Target objects	**Spanish**	**Chinese**	**Predicted acceptability**
	我妈妈有一只**猫**淘气的.	Mi mamá tiene un **gato** travieso.	Post-nominal

gato/猫 [Cat]	我妈妈有一只淘气的**猫**.	Mi mamá tiene un travieso **gato**.	Pre-nominal
globo/气球[balloon]	天空中飘着一个**气球**粉色的.	Hay un **globo** rosa flotando en el cielo.	Post-nominal
	天空中飘着一个粉色的**气球**.	Hay un rosa **globo** flotando en el cielo.	Pre-nominal
cerdo/猪 [pig]	在童话里，威尔伯是一只**猪**聪明的.	En los cuentos de hadas, Wilbur es un **cerdo** inteligente.	Post-nominal
	在童话里，威尔伯是一只聪明的**猪**.	En los cuentos de hadas, Wilbur es un inteligente **cerdo**.	Pre-nominal
guante/手套[glove]	玛丽丢了一只**手套**绿色的.	Mary perdió un **guante** verde.	Post-nominal
	玛丽丢了一只绿色的**手套**.	Mary perdió un verde **guante**.	Pre-nominal
huevo/鸡蛋 [egg]	汤姆给了他妹妹一个**鸡蛋**煮熟的.	Tom le dio a su hermana un **huevo** cocido.	Post-nominal
	汤姆给了他妹妹一个煮熟的**鸡蛋**.	Tom le dio a su hermana un cocido **huevo**.	Pre-nominal
patata/土豆[potato)	杰克在厨房里看到了一个**土豆**变质的.	Jack vio una **patata** estropeada en la cocina.	Post-nominal
	杰克在厨房里看到了一个变质的**土豆**.	Jack vio una estropeada **patata** en la cocina.	Pre-nominal
falda/裙子 [skirt]	丽莎有一条**裙子**漂亮的.	Lisa tiene una **falda** hermosa.	Post-nominal
	丽莎有一条漂亮的**裙子**.	Lisa tiene una hermosa **falda**.	Pre-nominal
fresa/草莓 [strawberry]	巴瑞画了一颗**草莓**绿色的.	Barry dibujó una **fresa** verde.	Post-nominal
	巴瑞画了一颗绿色的**草莓**.	Barry dibujó una verde **fresa**.	Pre-nominal
manzana/苹果[apple]	白雪公主吃了一个**苹果**有毒的.	Blancanieves se comió una **manzana** venenosa.	Post-nominal
	白雪公主吃了一个有毒的**苹果**.	Blancanieves se comió una venenosa **manzana**.	Pre-nominal

Appendix 6.C Examples of code-switching patterns produced by a paired "director" and "matcher".

Paired director and matcher in the production task in Chinese matrix contexts		
Director (Speaker 19)	Matcher (Spaker 20)	Comparison of two patterns
1. N + de-PP un 手套 de color amarillo Det glove of color yellow	1. N + de-PP un 手套 de color amarillo Det glove of color yellow	same
2. N + de-PP una 裙子 de color negro Det skirt of color black	2. N + Adj una 裙子 negra Det skirt black	different
3. N + de-PP una 草莓 de color blanco Det sreawberry of color white	3. N + de-PP una 草莓 de color blanco Det sreawberry of color white	same
4. N + de-PP un 猫 de color gris Det cat of color grey	4. N + de-PP un 猫 de color gris Det cat of color grey	same
5. N + Adj un 气球 azul Det balloon blue	5. N + de-PP un 气球 de color azul Det balloon of color blue	different
6. N + Adj un 鸡蛋 completo Det egg complete	6. N + Adj un 鸡蛋 completo Det egg complete	same
7. Adj + N un pequeño 土豆 Det small potato	7. Adj + N un pequeño 土豆 Det small potato	same

Chapter 7

General Discussion

This thesis represents the first comprehensive, multi-method investigation into how Spanish–Chinese bilinguals manage lexico-syntactic features across both unilingual and bilingual contexts. By employing a combination of behavioral, neurocognitive, and multi-task paradigms, the research seeks to address the central question of how bilingual individuals process grammatical and structural differences between typologically distinct languages. Specifically, we thoroughly explored how the interplay of linguistic, cognitive, and social factors, as well as task type/demands and community characteristics, shapes bilinguals' processing of grammatical patterns in unilingual and bilingual contexts. By doing so, this thesis therefore advances our understanding on several aspects: first, underscoring how two different linguistic systems are processed, integrated, and interact in bilinguals' minds; second, providing in-depth understanding of the similarities and differences when processing diverse lexico-syntactic features across unilingual and bilingual contexts; third, highlighting bilinguals' shared underlying mechanisms for processing functionally comparable lexico-syntactic features across their languages; fourth, pointing out the importance of exploring bilingualism as discovery science.

More precisely, through a series of (semi-)experimental tasks, we examined how Spanish–Chinese bilingual speakers process and produce key lexico-syntactic features, including grammatical gender, classifiers, and adjective placement in both unilingual and bilingual contexts. This line of inquiry was motivated by the typological differences between Spanish and Chinese, which, despite their typological and grammatical divergence, share comparable lexico-syntactic features. Moreover, research remains limited on how bilinguals navigate such cross-linguistic differences and comparable lexico-syntactic features during language processing and production across different contexts. To address this gap, we investigated five critical issues in depth: first, we examined and identified the gender congruency effect in Spanish NPs among Spanish speakers within unilingual contexts; second, we explored early Spanish–Chinese bilinguals' underlying mechanism of gender assignment in mixed Spanish–Chinese NPs within bilingual contexts; third, we investigated the classifier congruency effect in Chinese NP production within unilingual contexts among Mandarin Chinese speakers and early Spanish–Chinese bilinguals, using behavioral and electrophysiological measures to quantify the similarity and robustness of this effect across the two groups; fourth, we studied the underlying mechanism of classifier assignment among early Spanish–Chinese bilinguals in mixed Chinese–Spanish NPs within bilingual contexts; fifth, we examined how early Spanish–Chinese bilinguals resolve conflicts arising from adjective-noun word order differences across Spanish and Chinese in both unilingual and bilingual contexts.

In this chapter, we revisit the overarching research question that shaped the thesis and provide an integrated summary of the key findings from each chapter. We discuss how the results contribute to a deeper understanding of the processing involved in managing lexico-syntactic structures in both unilingual and bilingual contexts across Spanish and Chinese. Emphasis was placed on how bilinguals navigate grammatical gender, classifiers, and adjective-noun word order conflicts, shedding light on how the two language systems interact and interplay during bilingual language production and comprehension. We conclude this chapter with a discussion of the limitations of this thesis, including methodological constraints and the limited participant scope, and suggest directions for future studies to address these issues.

7.1 Discussion of chapter findings

In **Chapter 2** of this thesis, we explored whether grammatical gender, as a lexico-syntactic feature in Spanish, is used to competitively select determiners in NP production by Spanish speakers, and whether this competition is reflected in a *gender congruency effect*. This was motivated by the conflicting findings in previous studies on Romance languages regarding the presence of the gender congruency effect (see Wang & Schiller, 2019; Sá Leite et al., 2022; and Bürki et al., 2023 for an overview). In this chapter, we therefore employed a PWI-based picture naming task to probe the presence of the gender congruency effect. According to the lexical selection by competition theory (Levelt et al., 1999; Roelofs, 2003), speakers take longer to select and produce a target word when more highly non-target words are activated and compete for selection. Based on this, our working hypothesis was that Spanish speakers would name target pictures faster when distractor words matched the target in grammatical gender (gender-congruent condition), compared to when they did not (gender-incongruent condition). In line with our hypothesis, we predicted shorter naming latencies in gender-congruent conditions than in gender-incongruent conditions. Our results indicated a significant gender congruency effect in Spanish NP production, with participants naming targets significantly faster when paired with gender-congruent distractors than with gender-incongruent ones. As such, we provide clear evidence for the presence of the gender congruency effect in Spanish NP production.

Our behavioral results are important for several reasons: first, the significantly faster naming latencies in gender-congruent conditions suggest the presence of a *gender congruency effect* during gender processing; second, the gender congruency effect implies that competition occurred during the process of gender feature selection, in which different gender values for the target picture and the distractor were activated in the gender-incongruent condition, and gender features of distractors competed for selection at the lemma level; third, the statistically significant longer naming latencies in gender-incongruent conditions also suggests that the competition can be reflected by a naming latencies effect in Spanish NP production. Taken together, this study provides crucial behavioral evidence for the presence of the gender congruency effect in Spanish NP production, contributing to addressing the controversial findings in Romance languages. Moreover, it provides insights into how grammatical gender is processed and produced in unilingual Spanish contexts, laying the foundation for a broad exploration of lexico-syntactic features across languages and contexts.

In **Chapter 3** of this thesis, we shifted the focus to grammatical gender assignment in mixed Spanish–Chinese NPs by early Spanish–

Chinese bilinguals. Having established how grammatical gender is processed and produced in unilingual Spanish NPs in Chapter 2, we were interested in examining how it is processed and assigned in Spanish–Chinese bilingual NPs. Here, we asked the fundamental question of which gender assignment strategy, i.e., a masculine default strategy (e.g., *el 桌子* "the$_{MASC}$ table$_{FEM}$") or a translation equivalent strategy (e.g., *la 桌子* "the$_{FEM}$ table$_{FEM}$", where the Spanish equivalent of "table", *mesa*, is feminine), early Spanish–Chinese bilinguals use when assigning Spanish grammatical gender (masculine or feminine) to Chinese nouns in mixed Spanish–Chinese NPs. Previous studies have suggested that gender assignment strategies vary depending on linguistic, cognitive, and sociolinguistic factors, as well as task type and demands (cf. Bellamy & Parafita Couto, 2022). Accordingly, we predicted that early Spanish–Chinese bilinguals would employ different strategies across tasks, reflecting the combined influence of task type as well as linguistic, cognitive, and sociolinguistic factors. Using a multi-task approach, we found task-dependent variation in gender assignment, with bilinguals employing both masculine default and translation equivalent strategies in production, while showing a clear tendency toward the translation equivalent strategy in comprehension. Moreover, we highlighted that the diverse social network in the community, task demands, metalinguistic awareness, and linguistic variability play crucial roles in shaping bilingual language use.

Our study underscores the importance of documenting variation in bilingual communities for several reasons: first, our study is the first to explore gender assignment strategies in mixed Spanish–Chinese NPs using multiple tasks, offering fresh insights into gender assignments in bilingual speech; second, through explaining gender assignment patterns by early Spanish–Chinese bilinguals, we contribute to an understanding of the cognitive processes underlying gender assignment in bilingual contexts; third, our study provides evidence that variation in gender assignment is shaped not only by linguistic properties but also by task type and community characteristics, emphasizing the need for further research on how these factors interact in bilingual populations; fourth, our findings suggest that the variability in gender assignment strategies not only defines uniformity but also creates a shared norm within the bilingual community. This emphasizes the collective adaptability in language use and highlights bilingualism as an adaptive response to social and cognitive demands. Together, our study provides a crucial piece of evidence for understanding how grammatical gender, as one of the lexico-syntactic features, is navigated in bilingual contexts.

Chapter 4 extends the investigation of lexico-syntactic features by turning to a typologically different language, Mandarin Chinese, which features a classifier system that is comparable to grammatical gender. More precisely, we were interested in examining the processing and production of unilingual Chinese classifier-NPs by early Spanish–Chinese bilingual speakers. In previous literature, it is well-established that classifiers, similar to grammatical gender, undergo a comparable competitive selection process in Chinese NP production by monolingual speakers (e.g., Wang et al., 2019; also see Wang & Schiller, in press and Qian, in press for an overview). This process is evidenced by the *classifier congruency effect*, analogous to the *gender congruency effect* observed in Spanish in Chapter 2 (see an overview of gender congruency effects in Wang & Schiller, 2019 and Sá Leite et al., 2022; classifier congruency effects in Wang et al., 2019 and Huang & Schiller, 2021). However, the critical question we investigated in this chapter is whether bilingual speakers undergo a similar competitive selection process when producing Chinese classifier-NPs, particularly whether this is reflected in behavioral and electrophysiological evidence. Yet, to our knowledge, limited research has investigated how this process unfolds in bilingual speakers, especially using neurocognitive methods. Therefore, our primary research question was whether early Spanish–Chinese bilinguals show a classifier congruency effect in Chinese NP production. To ensure the robustness of our investigation, we included Mandarin Chinese speakers as a comparison group. Using a PWI-based picture-naming task, we examined naming latencies to assess whether the selection of classifiers is behaviorally competitive, and we analyzed electrophysiological responses to determine whether classifier violations elicit electrophysiological correlates such as an N400-like effect. By doing so, we probed the presence of the classifier congruency effect in Chinese NP production by Spanish–Chinese bilingual speakers, both behaviorally and electrophysiologically.

Our behavioral results revealed a robust classifier congruency effect in both groups, which entailed significantly shorter naming latencies when the target picture and distractor word matched in classifiers (classifier-congruent conditions) compared to when they mismatched (classifier-incongruent conditions). This suggested that a competitive selection process was involved in classifier retrieval during NP production in classifier-incongruent conditions. At the neurocognitive level, we found an ERP correlate that was consistent with the N400 component in both groups, which was elicited by classifier-incongruent conditions. Critically, we observed a more negative-going N400-like wave in classifier-incongruent compared to classifier-congruent conditions in both groups. Taken together, our study provides crucial behavioral and electrophysiological evidence for the presence of a classifier congruency effect in Chinese NP production among bilinguals. Moreover, it offers insights into the similarities in the underlying mechanisms involved in classifier processing and selection across both groups in unilingual Chinese NP production. As such, obtaining a clearer picture of how classifiers are processed in unilingual Chinese NPs by bilinguals could be a critical milestone in understanding how they manage lexico-syntactic features in unilingual contexts.

Chapter 5 continues the investigation of classifier selection in Mandarin Chinese by early Spanish–Chinese bilingual speakers and extends it to Chinese–Spanish NP in bilingual contexts. This chapter was built directly on Chapter 3, with a shifted focus on the classifier assignment in mixed Chinese–Spanish NPs. Thus, we asked an identical question as in Chapter 3 to examine which classifier assignment strategies early Spanish–Chinese bilinguals employ when assigning classifiers to Spanish nouns in mixed NPs, such as a translation equivalent strategy (e.g., *一棵 árbol*, [one + specific-classifier-ke1 + tree], "one tree", where the expected-specific classifier for the Chinese equivalent of “tree”, 树 /shu4/, is 棵 /ke1/, “a specific classifier used for plants”) or a default strategy (e.g., 一个 *árbol*, [one + general-classifier-ge4 + tree], “one tree”, where the general classifier 个 /ge4/ is used as default in Chinese). Employing a multi-task approach parallel to Chapter 3, we observed a remarkably similar task-dependent variation in classifier assignment among Spanish–Chinese bilinguals, mirroring the grammatical gender assignment patterns reported in Chapter 3. This was evidenced by their use of both default and translation equivalent strategies in production, while they favored the translation equivalent strategy in comprehension.

Crucially, this study revealed that the mixed strategies bilinguals use for classifier assignment closely mirror the patterns observed in grammatical gender assignment in mixed Spanish–Chinese NPs within the same bilingual community (see Chapter 3). These parallels highlight the roles of diverse social networks in the community, task demands, metalinguistic awareness, and linguistic variability in influencing bilingual language use. Moreover, our study suggests that exposure to diverse language practices promotes a high degree of adaptability and flexibility in how bilinguals process grammatical features. Together, this study provides key evidence that bilinguals rely on comparable underlying mechanisms for processing and assigning grammatical gender and classifiers in mixed NPs, thereby advancing our understanding of how Spanish–Chinese bilinguals manage comparable lexico-syntactic features across languages in bilingual contexts.

In **Chapter 6** of this thesis, we shift focus from language-specific lexico-syntactic features (grammatical gender and classifiers) to a shared feature, adjective placement, which exists in both Spanish and Chinese but is realized differently. Specifically, we evaluated how syntactic constraints, proposed by two models (the MLF and MP), shape bilinguals’ code-switched adjective-noun patterns and code-switching behaviors. Alongside this aim, we also examined whether noun insertion, which was documented as the preferred pattern in previous studies (e.g., Parafita Couto et al., 2015; Vanden Wyngaerd, 2017; Van Osch et al., 2023), is preferred by

Spanish–Chinese bilinguals. Previous studies yielded inconclusive results regarding which model could fully explain grammatical patterns in adjective placement in code-switched speech (see Section 1.2.3 and Chapter 6 for detailed descriptions). Building on this, we extended the investigation of adjective placement to an understudied language pair and bilingual population in both unilingual and bilingual contexts. In doing so, we examined the code-switching patterns that early Spanish–Chinese bilinguals favor in production and comprehension. Moreover, we also investigated how the MLF and MP frameworks inform adjective placement in mixed adjective-noun constructions, thereby refining our understanding of code-switching patterns. Employing a multi-task approach, we observed that adjective-noun patterns vary across matrix languages and language contexts. Specifically, our results revealed a consistent preference for the expected default word orders in unilingual contexts (prenominal in Chinese, postnominal in Spanish). In bilingual contexts, Spanish matrix language data supported MLF predictions in production, but MP better explained both production and comprehension patterns. In contrast, Chinese matrix language data supported both models. Additionally, noun insertions were preferred over adjective insertions, aligning with trends in other bilingual communities.

This study, to our knowledge, is the first to empirically examine adjective placement conflicts in Spanish–Chinese code-switching, providing new insight into how the interplay of task type, language mode, and grammatical constraints proposed by the two models shapes mixed adjective-noun constructions in bilingual production and comprehension. Our findings provide evidence that, although the two models make competing predictions regarding adjective placement, both matrix language structure (MLF) and morphosyntactic properties of individual lexical items (MP) contribute to constructing mixed adjective-noun constructions among Spanish–Chinese bilinguals. These results challenge the singular predictions proposed by each model individually and suggest that no single grammatical constraint from either model can fully explain the observed code-switching patterns. Instead, a more comprehensive understanding of mixed adjective-noun constructions requires acknowledging the combined influence of both clausal structures and morphosyntactic properties of individual lexical items. Together, our study contributes valuable data from an understudied language combination and bilingual community, offering new insight into how grammatical constraints shape bilingual code-switching patterns and complements existing research on adjective placement in bilingual speech. Moreover, this study fills a crucial gap in our understanding of how typologically distinct languages interact at the shared lexico-syntactic level, where shared grammatical features exist in both languages but are realized differently, highlighting the complex interplay between structural constraints and bilingual processing mechanisms.

7.2 The broader picture

Taking all chapters together, this thesis lays the groundwork for a new line of inquiry into bilingual lexico-syntactic processing by presenting the first comprehensive, multi-method investigation of grammatical gender, classifiers, and the linear order of the adjectives and nouns in Spanish–Chinese bilinguals. To achieve this, we first examined Spanish speakers and Mandarin Chinese speakers, not only to establish baseline patterns, but also to gain a more nuanced understanding of how lexico-syntactic features are navigated in unilingual contexts. This foundation enables a more robust examination of how Spanish–Chinese bilinguals navigate these lexico-syntactic features in bilingual contexts. We also employed a multi-task approach, including an elicitation task, a repetition task, and a forced-choice acceptability judgment task, combined with behavioral and neural measures, such as a picture-naming task and electrophysiological measures, to examine several critical issues of bilingual lexico-syntactic processing across languages and contexts.

To address the first critical issue outlined in the Introduction and Chapter 2, we employed a behavioral picture-naming task to investigate whether grammatical gender is competitively selected during NP production in unilingual Spanish. By doing this, we learnt that gender agreement between Spanish nouns and other elements (e.g., determiners and adjectives) is manifested by the reflection of the gender congruency effect in NP production. In other words, gender congruency between nouns and determiners (i.e., whether or not the nouns and determiners agree in gender) serves as a key factor in eliciting the gender congruency effect in Spanish speakers' production. This finding has directly demonstrated how grammatical gender is processed in unilingual contexts, establishing the foundation for the subsequent investigation of grammatical gender, particularly for exploring how grammatical gender is processed and assigned in mixed NPs within bilingual contexts.

After establishing the processing of grammatical gender in unilingual Spanish NPs, we then turn to the second critical issue, which concerns the assignment of grammatical gender in mixed NPs within bilingual contexts. To tackle this issue, we employed a multi-task approach to investigate how early Spanish–Chinese bilinguals assign gender to Chinese nouns in mixed NPs during production and comprehension, and to identify the gender assignment strategies they employ. We observed task-dependent variability in gender assignment and identified varied gender assignment strategies across tasks. These findings imply that the interplay of social network, task type, and cognitive demands plays a crucial role in shaping bilinguals' gender assignment patterns, highlighting the flexible and adaptable nature of bilingual language processing. Combined with the investigation of grammatical gender processing in unilingual contexts, these two studies provide comprehensive insights into how grammatical gender is processed and produced in both unilingual and bilingual contexts.

Regarding the third issue, whether classifiers, as another lexico-syntactic feature, are activated and competitively selected during unilingual Chinese NP production, we found that both Mandarin Chinese speakers and Spanish–Chinese bilinguals were indeed sensitive to classifier violations at the behavioral and neurocognitive level. This sensitivity is reflected in the presence of a classifier congruency effect, observed not only in the naming speed of producing classifier-NPs during picture-naming tasks, but also in the neurocognitive correlates of classifier processing, specifically an N400-like effect elicited by classifier violations. This is a robust finding across different populations of speakers in unilingual contexts, which directly implies the characterization of classifier processing from the behavioral and neurocognitive perspectives. Critically, this study demonstrated striking parallels between grammatical gender (Spanish, see Chapter 2) and classifiers (Chinese) in terms of congruency effects, underscoring that, despite typological distinctions, bilinguals appear to employ shared underlying mechanisms for processing functionally comparable lexico-syntactic features across languages. Moreover, this study starts a broader line of inquiry into the investigation of lexico-syntactic features. It establishes a robust empirical foundation for investigating classifier processing and assignment in mixed NPs within bilingual contexts.

With respect to the fourth critical issue, which parallels the investigation of grammatical gender assignment in Chapter 3, we examined classifier assignment strategies in mixed NPs preferred by Spanish–Chinese bilinguals from the same community. We observed comparable task-dependent variability in classifier assignment strategies, closely resembling those used for grammatical gender. This suggests bilinguals may apply similar underlying mechanisms when navigating structurally different but functionally comparable lexico-syntactic features. This finding is significant because it demonstrates that bilinguals do not rigidly adhere to one assignment strategy, but instead flexibly adapt their language processing and production based on task demands, sociolinguistic context, and the specific lexico-syntactic features involved. In particular, we found that the consistent influence of social network, task type, and cognitive demands across both grammatical gender and classifier assignment may enable bilinguals to flexibly adapt their linguistic choices in bilingual language use. Critically, this finding has direct implications for understanding the underlying mechanisms involved in processing lexico-syntactic features in bilingual contexts. It sheds light on how the flexibility of bilingual grammatical systems interacts with cognitive demand and sociolinguistic influences to shape bilingual language behavior.

In terms of the fifth critical issue, we shift our attention to the syntactic conflict of linear order between adjectives and nouns across Spanish and Chinese, which exists in both languages but is realized differently. To tackle this issue, we employed a multi-task approach to examine how bilinguals structure adjective-noun constructions based on grammatical constraints proposed by two theoretical models, the MLF and MP, within both unilingual and bilingual contexts. Research on this topic involving Spanish–Chinese languages remains particularly limited, making this study both critical and valuable. Critically, our finding challenges the idea that a single theoretical model (MLF or MP) fully explains code-switching patterns. Instead, our results strongly suggest that, although the two models make competing predictions regarding adjective placement, both matrix language structure (MLF) and the morphosyntactic properties of individual lexical items (MP) interact and contribute to mixed constructions, leading to a more fine-grained understanding of syntactic integration. This finding moves the field beyond a binary "either/or" debate between the two models. Moreover, this critical investigation offers new insights into bilingual language processing. It complements the study of grammatical gender and classifiers by revealing how bilinguals negotiate and integrate overlapping lexico-syntactic systems. As such, we demonstrate that the linear order of adjectives and nouns is an ideal lens that provides nuanced insights into bilingual language processing. It completes a crucial piece in understanding how such features are processed and integrated across languages and contexts.

By pooling several strands of empirical evidence in this thesis, we contribute novel insights into processing and integrating lexico-syntactic features across typologically distinct languages, Spanish and Chinese, within both unilingual and bilingual contexts. This work offers comprehensive accounts of how early Spanish–Chinese bilinguals navigate grammatical gender, classifiers, and adjective-noun word order across languages, thereby enriching the literature with data from an underexplored bilingual population. Moreover, the findings provide important building blocks to characterize how these cross-linguistic lexico-syntactic features integrate and interact in bilinguals' minds, advancing our understanding of bilingual grammatical representation and processing.

7.3 Open science and bilingualism as discovery science

To continue the efforts on *open science* and *bilingualism as discovery science* in previous studies, which makes published findings of scientific research on bilingualism more robust, reproducible, and replicable, we also put special emphasis on making all our studies, related materials, and data analysis scripts in this thesis fully transparent, openly available, and easily accessible through the Open Science Framework (OSF, Foster & Deardorff, 2017). Moreover, we provide detailed descriptions of participants' full social background and language profile, task designs, stimulus selection criteria, the rationale for statistical analytical choices, and data analysis procedures across all chapters of this thesis. This level of transparency enhances the accessibility and replicability of our studies, contributing meaningfully to ongoing Open Science efforts. Beyond this, we continued the view of bilingualism as an ongoing discovery process in this thesis. Bilingualism is a complex and evolving phenomenon that is still far from fully defined or understood (Navarro-Torres et al., 2021), and it should be studied as a science of discovery. With this in mind, we designed our research by identifying convergent patterns across diverse studies and applying *variety* throughout all chapters in this thesis. To achieve this, we employed diverse research ideas and multi-method approaches, combined with detailed participant characterizations and accessible, transparent materials, rather than relying on simplified or uniform methodologies used in prior research. This approach not only prompts a more incremental and nuanced understanding of bilingualism but also ensures greater reliability and replicability of research findings. To sum up, *open science* and *bilingualism as discovery science* should be regarded as critical factors for ensuring the reliability, replicability, and robustness of future research on bilingualism.

7.4 Limitations and future research

While this thesis provides fresh insights into the process and production of lexico-syntactic features and adjective-noun word order patterns among Spanish–Chinese bilinguals, several limitations should be acknowledged. First and foremost, this thesis focuses on a specific bilingual population, early Spanish–Chinese bilinguals in Barcelona, Spain. While this population offers a valuable case to investigate bilingual language processing across two typologically distinct languages, the findings should be interpreted with caution in terms of their broader applicability. Bilingual communities vary widely in terms of sociolinguistic contexts, and factors such as language dominance, frequency and patterns of code-switching, community language practices, and the quantity and quality of input in each language can differ substantially depending on geographic location and social environment. These variables may significantly influence how bilinguals process and produce lexico-syntactic features. Therefore, future research is needed to determine whether the patterns observed in this study are consistent across other Spanish–Chinese bilingual populations, particularly those residing in different regions, immersed in different community norms, or shaped by distinct language acquisition trajectories. To do so, future research should conduct systematic replication studies across diverse Spanish–Chinese communities, such as in different regions or with varied acquisition trajectories, to test the generalizability of the findings.

Second, the investigation of grammatical gender congruency effects in unilingual Spanish NPs was limited to only Spanish speakers born and raised in Spanish-speaking countries without Chinese proficiency. This methodological choice was determined by the constraints imposed by the COVID-19 pandemic, which prevented the recruitment and testing of bilingual participants. As a result, the study does not offer insight into how early Spanish–Chinese bilinguals process grammatical gender in unilingual Spanish contexts, which is an important dimension that could further illustrate whether bilinguals exhibit similar processing mechanisms relative to Spanish speakers. Moreover, this study only included behavioral measures, without incorporating neurocognitive measures such as EEG. The exclusion of EEG measures was primarily due to time constraints and limited prior experience with electrophysiological data collection and analysis. These limitations restricted our ability to explore the temporal dynamics and underlying neurocognitive correlates of gender processing in greater depth during language production. To address this limitation, future research should include early Spanish–Chinese bilinguals in these unilingual Spanish tasks with both behavioral and neurocognitive measures to fully understand their processing mechanisms compared to monolinguals.

Third, studies presented in this thesis primarily employ (semi-)experimental tasks, which, while offering high levels of control and replicability, may not fully reflect the complexity and variability of naturalistic language use. This is particularly relevant for understanding how bilingual speakers engage in code-switching and resolve morphosyntactic conflicts, such as grammatical gender assignment, classifier selection, or the linear word order of adjectives and nouns in spontaneous communicative contexts. The reliance on elicited production and judgment comprehension tasks means that the findings may not generalize to bilinguals' everyday language practices. The absence of conversational or spontaneous speech data limits our ability to evaluate how these lexico-syntactic features and syntactic differences are processed and negotiated in spontaneous and natural interactions. Thus, integrating longitudinal and corpus-based methods, alongside experimental designs, is essential to capture the full spectrum of spontaneous language use and evaluate the ecological validity of the findings.

Fourth, the interpretation of bilinguals' code-switching behaviors in this thesis may have been influenced by inconsistencies in participants' self-reported language use. Variability in individual understandings of what constitutes code-switching likely introduced inconsistencies. Such discrepancies can complicate the accurate measurement of switching frequency, patterns, and contextual usage. To enhance the validity and consistency of future analyses, researchers should develop and use more objective and systematic observational methods for assessing code-switching behavior in natural settings, such as structured video recordings of spontaneous interactions or corpus-based annotations, ideally in combination with social network analysis to capture the influence of interpersonal and community-level factors.

Taken together, this thesis offers the first exploration of how lexico-syntactic features are processed and produced by early Spanish–Chinese bilinguals. The findings provide valuable insights into how bilinguals navigate these "conflict sites", such as the presence or absence of grammatical gender and classifier system, and the differences in adjective placement across their two languages, and how these features are processed and produced in different language contexts. While this thesis acknowledges several limitations, it nonetheless lays an important foundation for future research in this field. To build on these findings, future research should address these limitations by recruiting more diverse participant samples, integrating naturalistic data and neuroimaging methods, and employing more systematic and observational methods for assessing bilingual language behaviors to enhance validity and theoretical depth of subsequent research.

References

Aikhenvald, A. Y. (2000). *Classifiers: A Typology of Noun Categorization Devices*. Oxford University Press. https://doi.org/10.1093/oso/9780198238867.001.0001

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, *19*(6), 716–723. https://doi.org/10.1109/TAC.1974.1100705

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

Allan, K. (1977). Classifiers. *Language*, *53*(2), 285–311.

Balam, O. (2016). Semantic Categories and Gender Assignment in Contact Spanish: Type of Code-Switching and its Relevance to Linguistic Outcomes. *Journal of Language Contact*, *9*(3), 405–435. https://doi.org/10.1163/19552629-00903001

Balam, O., Lakshmanan, U., & Parafita Couto, M. D. C. (2021). Gender Assignment Strategies Among Simultaneous Spanish/English Bilingual Children from Miami, Florida. *Studies in Hispanic and Lusophone Linguistics*, *14*(2), 241–280. https://doi.org/10.1515/shll-2021-2045

Balam, O., & Parafita Couto, M. del C. (2019). Adjectives in Spanish/English code-switching: Avoidance of grammatical gender in bi/multilingual speech. *Spanish in Context*, *16*(2), 194–216.

Barr, D. J. (2013). Random effects structure for testing interactions in linear mixed-effects models. *Frontiers in Psychology*, *4*. https://doi.org/10.3389/fpsyg.2013.00328

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

Beatty-Martínez, A. L., & Dussias, P. E. (2017). Bilingual experience shapes language processing: Evidence from codeswitching. *Journal of*

Memory and Language, *95*, 173–189. https://doi.org/10.1016/j.jml.2017.04.002

Beatty-Martínez, A. L., & Dussias, P. E. (2019). Revisiting Masculine and Feminine Grammatical Gender in Spanish: Linguistic, Psycholinguistic, and Neurolinguistic Evidence. *Frontiers in Psychology*, *10*, 751. https://doi.org/10.3389/fpsyg.2019.00751

Beatty-Martínez, A. L., Valdés Kroff, J. R., & Dussias, P. E. (2018). From the field to the lab: A converging methods approach to the study of codeswitching. *Languages*, *3*(2), 19.

Belke, E., Meyer, A. S., & Damian, M. F. (2005). Refractory effects in picture naming as assessed in a semantic blocking paradigm. *The Quarterly Journal of Experimental Psychology Section A*, *58*(4), 667–692. https://doi.org/10.1080/02724980443000142

Bellamy, K., & Parafita Couto, M. C. (2022). Chapter 2. Gender assignment in mixed noun phrases: State of the art. In D. Ayoun (Ed.), *Studies in Bilingualism* (Vol. 63, pp. 13–48). John Benjamins Publishing Company. https://doi.org/10.1075/sibil.63.02bel

Bellamy, K., Parafita Couto, M., & Stadthagen-Gonzalez, H. (2018). Investigating Gender Assignment Strategies in Mixed Purepecha–Spanish Nominal Constructions. *Languages*, *3*(3), 28. https://doi.org/10.3390/languages3030028

Bellamy, K., & Wichers Schreur, J. (2022). When semantics and phonology collide: Gender assignment in mixed Tsova-Tush–Georgian nominal constructions. *International Journal of Bilingualism*, *26*(3), 257–285. https://doi.org/10.1177/13670069211039559

Beltrán Antolín, J. (2006). El modo español en las diásporas de Asia Oriental. *Las Diasporas de Asia Oriental En Europa Occidental, Barcelona: CIDOB*, 101–124.

Beltrán Antolín, J. (2008). Los retos de la inserción social. Un caso de estudio asiático. *Sistemas Culturales Multiétnicos y Derecho de Integración, Madrid, Consejo General Del Poder Judicial, Colección Cuadernos de Derecho Judicial*, *20*, 177–210.

Beltrán Antolín, J., & López, A. S. (2013). Del restaurante chino al bar autóctono. Evolución del empresariado de origen chino en España y su compleja relación con la etnicidad. In *Retos y estrategias del empresariado étnico. Estudios de caso de empresarios latinos en los Estados Unidos y empresarios inmigrantes en España* (pp. 85–108).

Berger, A., & Kiefer, M. (2021). Comparison of different response time outlier exclusion methods: A simulation study. *Frontiers in Psychology*, 2194. https://doi.org/10.3389/fpsyg.2021.675558

Berghoff, R., Gullberg, M., & Kootstra, G. J. (2023). Structural priming of code-switches in non-shared-word-order utterances: The effect of lexical repetition. *Bilingualism: Language and Cognition*, *26*(4), 670–683.

Bialystok, E., Craik, F. I., & Luk, G. (2008). Lexical access in bilinguals: Effects of vocabulary size and executive control. *Journal of Neurolinguistics*, *21*(6), 522–538.

Bialystok, E., Craik, F. I., & Luk, G. (2012). Bilingualism: Consequences for mind and brain. *Trends in Cognitive Sciences*, *16*(4), 240–250.

Bialystok, E., Kroll, J. F., Green, D. W., MacWhinney, B., & Craik, F. I. (2015). Publication bias and the validity of evidence: What's the connection? *Psychological Science*, *26*(6), 944–946.

Bialystok, E., & Luk, G. (2012). Receptive vocabulary differences in monolingual and bilingual adults. *Bilingualism: Language and Cognition*, *15*(2), 397–401.

Bialystok, E., Luk, G., Peets, K. F., & Yang, S. (2010). Receptive vocabulary differences in monolingual and bilingual children. *Bilingualism: Language and Cognition*, *13*(4), 525–531.

Bierings, E., Parafita Couto, M. C., & Mateo Pedro, P. (2019). Contrasting code-switching theories: Insights from Kaqchikel-Spanish code-switched nominal constructions. *Proceedings of form and analysis in Mayan linguistics, 5*. https://hdl.handle.net/1887/136824

Bisang, W. (in press). Nominal classification from a typological perspective. In *Oxford Handbook of Gender and Classifiers*. Oxford University Press.

Blackford, T., Holcomb, P. J., Grainger, J., & Kuperberg, G. R. (2012). A funny thing happened on the way to articulation: N400 attenuation despite behavioral interference in picture naming. *Cognition*, *123*(1), 84–99.

Bloem, I., & La Heij, W. (2003). Semantic facilitation and semantic interference in word translation: Implications for models of lexical access in language production. *Journal of Memory and Language*, *48*(3), 468–488.

Blokzijl, J., Deuchar, M., & Parafita Couto, M. C. (2017). Determiner asymmetry in mixed nominal constructions: The role of grammatical factors in data from Miami and Nicaragua. *Languages*, *2*(4), 20.

Boersma, P., & Weenink, D. (2019). *Praat: Doing Phonetics by Computer (6.0.49) [Computer software]*. https://www.fon.hum.uva.nl/praat/

Bornkessel-Schlesewsky, I., & Schlesewsky, M. (2008). An alternative perspective on "semantic P600" effects in language comprehension. *Brain Research Reviews*, *59*(1), 55–73.

Bosque, I., & Picallo, C. (1996). Postnominal adjectives in Spanish DPs1. *Journal of Linguistics*, *32*(2), 349–385.

Brysbaert, M. (2021). Power considerations in bilingualism research: Time to step up our game. *Bilingualism: Language and Cognition*, *24*(5), 813–818.

Bull, W. E. (1950). Spanish adjective position: Present rules and theories. *Hispania*, *33*(4), 297–303.

Bull, W. E. (1984). *Spanish for teachers: Applied linguistics*. Krieger Publishing Company.

Bullock, B. E., & Toribio, A. J. E. (2009). *The Cambridge handbook of linguistic code-switching*. Cambridge University Press.

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

Bürki, A., van den Hoven, E., Schiller, N., & Dimitrov, N. (2023). Cross-linguistic differences in gender congruency effects: Evidence from meta-analyses. *Journal of Memory and Language*, *131*, 104428.

Cantone, K. F., & MacSwan, J. (2009). Adjectives and word order. *Multidisciplinary Approaches to Code Switching*, 243–278.

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

Chambers, C. D. (2013). Registered reports. *A New Publishing Initiative at Cortex*. *Cortex*, *49*(3), 609–610.

Chao, L. L., Haxby, J. V., & Martin, A. (1999). Attribute-based neural substrates in temporal cortex for perceiving and knowing about objects. *Nature Neuroscience*, *2*(10), 913–919.

Chao, Y. (1968). A grammar of spoken Chinese. In *A grammar of spoken Chinese*. University of California Press.

Chao, Y. R. (1948). *Mandarin primer: An intensive course in spoken Chinese*. Harvard University Press.

Chen, P., Bobb, S. C., Hoshino, N., & Marian, V. (2017). Neural signatures of language co-activation and control in bilingual spoken word comprehension. *Brain Research*, *1665*, 50–64.

Chomsky, N. (1995). *The minimalist program*. Cambridge, MA: MIT Press.

Chou, C.-J., Huang, H.-W., Lee, C.-L., & Lee, C.-Y. (2014). Effects of semantic constraint and cloze probability on Chinese classifier-noun agreement. *Journal of Neurolinguistics*, *31*, 42–54.

Chwilla, D. J., Brown, C. M., & Hagoort, P. (1995). The N400 as a function of the level of processing. *Psychophysiology*, *32*(3), 274–285.

Cidian, H. L. (1988). Hanyu Liangci Cidian [A dictionary of Chinese classifiers]. *Fuzhou: Fujian Renmin Chubanshe*.

Cinque, G. (1994). On the evidence for partial N-Movement in the Romance DP "in Cinque, G. Koster, J. Pollock, J.-Y. Rizzi, L. and R. Zanuttini (eds.) Paths Towards Universal Grammar: Studies in

Honour of Richard Kayne. *Studies in Honor of Richard S. Kayne. Washington, Georgetown University Press*, 185–217.

Cinque, G. (1999). *Adverbs and functional heads: A cross-linguistic perspective*. Oxford University Press.

Cinque, G. (2005). Deriving Greenberg's Universal 20 and its exceptions. *Linguistic Inquiry*, *36*(3), 315–332.

Cisneros, M., Rodríguez-González, E., Bellamy, K., & Couto, M. C. P. (2023). Gender strategies in the perception and production of mixed nominal constructions by New Mexico Spanish-English bilinguals. *Isogloss. Open Journal of Romance Linguistics*, *9*(2), 1–30.

Corbett, G. G. (1991). *Gender*. Cambridge University Press.

Corbett, G. G. (2012). Gender. Cambridge University Press; Cambridge Core. https://doi.org/10.1017/CBO9781139166119

Corbett, G. G. (2013). *The expression of gender*. De Gruyter.

Costa, A., Miozzo, M., & Caramazza, A. (1999a). Lexical selection in bilinguals: Do words in the bilingual's two lexicons compete for selection? *Journal of Memory and Language*, *41*(3), 365–397.

Costa, A., & Santesteban, M. (2004). Lexical access in bilingual speech production: Evidence from language switching in highly proficient bilinguals and L2 learners. *Journal of Memory and Language*, *50*(4), 491–511.

Costa, A., Sebastián-Gallés, N., Miozzo, M., & Caramazza, A. (1999b). The Gender Congruity Effect: Evidence from Spanish and Catalan. *Language and Cognitive Processes*, *14*(4), 381–391. https://doi.org/10.1080/016909699386275

Costa, A., Strijkers, K., Martin, C., & Thierry, G. (2009). The time course of word retrieval revealed by event-related brain potentials during overt speech. *Proceedings of the National Academy of Sciences*, *106*(50), 21442–21446.

Coulson, S., King, J. W., & Kutas, M. (1998). Expect the unexpected: Event-related brain response to morphosyntactic violations. *Language and Cognitive Processes*, *13*(1), 21–58.

Cruz, A. (2021). A syntactic approach to gender assignment in Spanish–English bilingual speech. *Glossa: A Journal of General Linguistics*, *6*(1). https://doi.org/10.16995/glossa.5878

Cruz, A. (2023). Linguistic factors modulating gender assignment in Spanish–English bilingual speech. *Bilingualism: Language and Cognition*, *26*(3), 580–591.

Cubelli, R., Lotto, L., Paolieri, D., Girelli, M., & Job, R. (2005). Grammatical gender is selected in bare noun production: Evidence from the picture–word interference paradigm. *Journal of Memory and Language*, *53*(1), 42–59. https://doi.org/10.1016/j.jml.2005.02.007

Damian, M. F., & Martin, R. C. (1999). Semantic and phonological codes interact in single word production. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *25*(2), 345–361. https://doi.org/10.1037/0278-7393.25.2.345

Davies, M. (2016). Corpus del Español: Two billion words, 21 countries. *Corpus Del Español: Two Billion Words*, *21*.

De Bot, K. (1992). A Bilingual Production Model: Levelt's Speaking Model Adapted. *Applied Linguistics*, *13*(1), 1–24.

De Bot, K., & Schreuder, R. (1993). Word production and the bilingual lexicon. *The Bilingual Lexicon*, *191*, 214.

Dell'Acqua, R., Sessa, P., Peressotti, F., Mulatti, C., Navarrete, E., & Grainger, J. (2010). ERP evidence for ultra-fast semantic processing in the picture–word interference paradigm. *Frontiers in Psychology*, *1*, 177.

Delogu, F., Brouwer, H., & Crocker, M. W. (2019). Event-related potentials index lexical retrieval (N400) and integration (P600) during language comprehension. *Brain and Cognition*, *135*, 103569.

Denny, J. P. & University of Western Ontario. Dept. of Psychology. (1976). *What are Noun Classifiers Good For?* University of Western Ontario. https://books.google.co.jp/books?id=FNkPnAEACAAJ

Desmet, T., & Duyck, W. (2007). Bilingual language processing. *Language and Linguistics Compass*, *1*(3), 168–194.

Deuchar, M. (2005). Congruence and code-switching in Welsh. *Bilingualism: Language and Cognition*, *8*(3), 1–15.

Deuchar, M. (2006). Welsh-English code-switching and the Matrix Language Frame model. *Lingua*, *116*(11), 1986–2011.

Deuchar, M. (2012). Code Switching. In *The Encyclopedia of Applied Linguistics*. John Wiley & Sons, Ltd. https://doi.org/10.1002/9781405198431.wbeal0142

Dijkstra, T., & van Heuven, W. J. (1998). The BIA model and bilingual word recognition. In *Localist connectionist approaches to human cognition* (pp. 189-225). Psychology Press.

Duñabeitia, J. A., Crepaldi, D., Meyer, A. S., New, B., Pliatsikas, C., Smolka, E., & Brysbaert, M. (2018). MultiPic: A standardized set of 750 drawings with norms for six European languages. *Quarterly Journal of Experimental Psychology*, *71*(4), 808–816. https://doi.org/10.1080/17470218.2017.1310261

Eddington, D. (2002). Spanish Gender Assignment in an Analogical Framework. *Journal of Quantitative Linguistics*, *9*(1), 49–75. https://doi.org/10.1076/jqul.9.1.49.8482

Erbaugh, M. S. (1986). Taking stock: The development of Chinese noun classifiers historically and in young children. *In Colette G. Craig, Ed. Noun Classes and Nominalization. Amsterdam: John Benjamins.*, 399–436.

Erbaugh, M. S. (2002). Classifiers are for specification: Complementary functions for sortal and general classifiers in Cantonese and Mandarin. *Cahiers de Linguistique Asie Orientale*, *31*(1), 33–69.

Erbaugh, M. S. (2006). Chinese classifiers: Their use and acquisition. In *The Handbook of East Asian Psycholinguistics* (pp. 39–51). Cambridge University Press.

Finocchiaro, C., Alario, F.-X., Schiller, N., Costa, A., Miozzo, M., & Caramazza, A. (2011). Gender congruency goes Europe: A cross-linguistic study of the gender congruency effect in Romance and Germanic languages. *Italian Journal of Linguistics/Rivista Di Linguistica*, *23*(2), 161–198.

Foster, E. D., & Deardorff, A. (2017). Open science framework (OSF). *Journal of the Medical Library Association: JMLA*, *105*(2), 203.

Fricke, M., & Kootstra, G. J. (2016). Primed codeswitching in spontaneous bilingual dialogue. *Journal of Memory and Language*, *91*, 181–201.

Friederici, A. D., Hahne, A., & Mecklinger, A. (1996). Temporal structure of syntactic parsing: Early and late event-related brain potential effects. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *22*(5), 1219.

Friederici, A. D., Pfeifer, E., & Hahne, A. (1993). Event-related brain potentials during natural speech processing: Effects of semantic, morphological and syntactic violations. *Cognitive Brain Research*, *1*(3), 183–192.

Fuller, J. M., & Lehnert, H. (2000). Noun phrase structure in German-English codeswitching: Variation in gender assignment and article use. *International Journal of Bilingualism*, *4*(3), 399–420. https://doi.org/10.1177/13670069000040030601

García-Pentón, L., Fernandez Garcia, Y., Costello, B., Duñabeitia, J. A., & Carreiras, M. (2016a). "Hazy" or "jumbled"? Putting together the pieces of the bilingual puzzle. *Language, Cognition and Neuroscience*, *31*(3), 353–360.

García-Pentón, L., Fernandez Garcia, Y., Costello, B., Duñabeitia, J. A., & Carreiras, M. (2016b). The neuroanatomy of bilingualism: How to turn a hazy view into the full picture. *Language, Cognition and Neuroscience*, *31*(3), 303–327.

Gil, D. (2005). Numeral Classifiers. *The World Atlas of Language Structures, 226-229 (2005).*

Glaser, W. R. (1992). Picture naming. *Cognition*, *42*(1–3), 61–105. https://doi.org/10.1016/0010-0277(92)90040-O

Glaser, W. R., & Düngelhoff, F.-J. (1984). The time course of picture-word interference. *Journal of Experimental Psychology: Human Perception and Performance*, *10*(5), 640–654. https://doi.org/10.1037/0096-1523.10.5.640

Gollan, T. H., Fennema-Notestine, C., Montoya, R. I., & Jernigan, T. L. (2007). The bilingual effect on Boston Naming Test performance. *Journal of the International Neuropsychological Society*, *13*(2), 197–208.

Gollan, T. H., Montoya, R. I., Fennema-Notestine, C., & Morris, S. K. (2005). Bilingualism affects picture naming but not picture classification. *Memory & Cognition*, *33*, 1220–1234.

González-Vilbazo, K. (2005). Die syntax des code-switching. Esplugisch: Sprachwechsel an der deutschen schule Barcelona. *Unpublished PhD Dissertation, Universität Zu Köln*, *76*.

Green, D. W. (1986). Control, activation, and resource: A framework and a model for the control of speech in bilinguals. *Brain and Language*, *27*(2), 210–223.

Greenberg, J. H. (1972). Numeral Classifiers and Substantival Number: Problems in the Genesis of a Linguistic Type. *Working Papers on Language Universals, No. 9*.

Greenham, S. L., Stelmack, R. M., & Campbell, K. B. (2000). Effects of attention and semantic relation on event-related potentials in a picture-word naming task. *Biological Psychology*, *55*(2), 79–104.

Grosjean, F. (1982). *Life with two languages: An introduction to bilingualism*. Harvard University Press.

Grosjean, F. (1992). Another view of bilingualism. In *Advances in psychology* (Vol. 83, pp. 51–62). Elsevier.

Grüter, T., Lew-Williams, C., & Fernald, A. (2012). Grammatical gender in L2: A production or a real-time processing problem? *Second Language Research*, *28*(2), 191–215. https://doi.org/10.1177/0267658312437990

Gullberg, M., Indefrey, P., & Muysken, P. (2009). Research techniques for the study of code-switching. In *The Cambridge Handbook of Linguistic Code-switching* (pp. 21–39). Cambridge University Press.

Guo, X. (2002). *现代汉语量词用法词典:* (3rd ed.). 语文出版社. https://books.google.nl/books?id=vSNYAAAACAAJ

Hagoort, P., Brown, C., & Groothusen, J. (1993). The syntactic positive shift (SPS) as an ERP measure of syntactic processing. *Language and Cognitive Processes*, *8*(4), 439–483.

Hagoort, P., & Brown, C. M. (2000). ERP effects of listening to speech compared to reading: The P600/SPS to syntactic violations in spoken sentences and rapid serial visual presentation. *Neuropsychologia*, *38*(11), 1531–1549.

Harris, J. W. (1991). The Exponence of Gender in Spanish. *Linguistic Inquiry*, *22*(1), 27–62. JSTOR.

He, Y. (2024). *Chinese Migration and Their Families in Spain: A Case Study on Barcelona.* [Doctoral dissertation, Universitat de Barcelona]. https://hdl.handle.net/2445/212887

Heim, S., Friederici, A. D., Schiller, N. O., Rüschemeyer, S.-A., & Amunts, K. (2009). The determiner congruency effect in language production investigated with functional MRI. *Human Brain Mapping*, *30*(3), 928–940. https://doi.org/10.1002/hbm.20556

Hernandez, A. E., Martinez, A., & Kohnert, K. (2000). In search of the language switch: An fMRI study of picture naming in Spanish–English bilinguals. *Brain and Language*, *73*(3), 421–431.

Hockett, C. F. (1958). A course in modern linguistics. *Language Learning*, *8*(3-4), 73-75.

Hopp, H. (2016). Learning (not) to predict: Grammatical gender processing in second language acquisition. *Second Language Research*, *32*(2), 277–307. https://doi.org/10.1177/02676583156 24960

Hu, Q. (1993). *The acquisition of Chinese classifiers by young Mandarin-speaking children*. Boston University.

Huang, S., & Schiller, N. O. (2021). Classifiers in Mandarin Chinese: Behavioral and electrophysiological evidence regarding their representation and processing. *Brain and Language*, *214*, 104889. https://doi.org/10.1016/j.bandl.2020.104889

Instituto Cervantes. (2010). *Diplomas de Español como Lengua Extranjera (DELE)*. Diplomas de Español Como Lengua Extranjera(DELE). http://diplomas.cervantes.es/

Instituto Nacional de Estadística (INE). (2016). Estadística del Padrón Continuo a 1 de enero de 2016. *Datos a Nivel Nacional*.

Ioannidis, J. P. (2005). Why most published research findings are false. *PLoS Medicine*, *2*(8), e124.

Jake, J. L., Myers-Scotton, C., & Gross, S. (2002). Making a minimalist ap proach to codeswitching work: Adding the Matrix Language. *Bilingualism: Language and Cognition*, *5*(01). https://doi.org/10.1017/S1366728902000147

Kim, A., & Osterhout, L. (2005). The independence of combinatory semantic processing: Evidence from event-related potentials. *Journal of Memory and Language*, *52*(2), 205–225.

Kolk, H. H., Chwilla, D. J., Van Herten, M., & Oor, P. J. (2003). Structure and limited capacity in verbal working memory: A study with event-related potentials. *Brain and Language*, *85*(1), 1–36.

Kreher, D. A., Holcomb, P. J., & Kuperberg, G. R. (2006). An electrophysiological investigation of indirect semantic priming. *Psychophysiology*, *43*(6), 550–563.

Królikowska, M., Bierings, E., Beatty-Martínez, A. L., Navarro-Torres, C., Dussias, P. E., & Parafita Couto, M. del C. (2019). *Gender-assignment strategies within the bilingual determiner phrase: Four Spanish-English communities examined*. 9–11.

Kuperberg, G. R., Sitnikova, T., Caplan, D., & Holcomb, P. J. (2003). Electrophysiological distinctions in processing conceptual relationships within simple sentences. *Cognitive Brain Research*, *17*(1), 117–129.

Kutas, M., & Federmeier, K. D. (2011). Thirty years and counting: Finding meaning in the N400 component of the event-related brain potential (ERP). *Annual Review of Psychology*, *62*(1), 621–647.

Kutas, M., & Hillyard, S. A. (1980). Reading senseless sentences: Brain potentials reflect semantic incongruity. *Science*, *207*(4427), 203–205.

La Heij, W. (1988). Components of Stroop-like interference in picture naming. *Memory & Cognition*, *16*(5), 400–410. https://doi.org/10.3758/BF03214220

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

Lakoff, G. (1987). *Women, fire, and dangerous things: What categories reveal about the mind*. Chicago : University of Chicago Press.

Lau, E. F., Phillips, C., & Poeppel, D. (2008). A cortical network for semantics:(de) constructing the N400. *Nature Reviews Neuroscience*, *9*(12), 920–933.

Leckey, M., & Federmeier, K. D. (2019). Electrophysiological methods in the study of language processing. *The Oxford Handbook of Neurolinguistics, Oxford University Press*. https://doi.org/10.1093/ oxfordhb/9780190672027.013.3.

Levelt, W. J. M. (1992). Accessing words in speech production: Stages, processes and representations. *Cognition*, *42*(1–3), 1–22. https://doi.org/10.1016/0010-0277(92)90038-J

Levelt, W. J. M., Roelofs, A., & Meyer, A. S. (1999). A theory of lexical access in speech production. *Behavioral and Brain Sciences*, *22*(01). https://doi.org/10.1017/S0140525X99001776

Li, C. N., & Thompson, S. A. (1981). *Mandarin Chinese: A Functional Reference Grammar*. University of California Press.

Li, X. (2013). Numeral Classifiers in Chinese: The Syntax-Semantics Interface (1st ed., Vol. 250). In *Numeral Classifiers in Chinese*. De Gruyter Mouton.

Liceras, J. M., Fernández Fuertes, R., Perales, S., Pérez-Tattam, R., & Spradlin, K. T. (2008). Gender and gender agreement in bilingual native and non-native grammars: A view from child and adult functional–lexical mixings. *Lingua*, *118*(6), 827–851. https://doi.org/10.1016/j.lingua.2007.05.006

Linde-Domingo, J., Treder, M. S., Kerrén, C., & Wimber, M. (2019). Evidence that neural information flow is reversed between object perception and object reconstruction from memory. *Nature Communications*, *10*(1), 179.

López, L. (Ed.). (2020). Building the Case for 1Lex: Gender in Code-Switching. In *Bilingual Grammar: Toward an Integrated Model* (pp. 76–115). Cambridge University Press; Cambridge Core. https://doi.org/10.1017/9781108756181.005

MacSwan, J. (1999). *A Minimalist Approach to Intrasentential Code- Switching*. Routledge. https://doi.org/10.4324/9781315053158

MacSwan, J. (2005). Codeswitching and generative grammar: A critique of the MLF model and some remarks on "modified minimalism." *Bilingualism: Language and Cognition*, *8*(1), 1–22.

Marian, V., Blumenfeld, H. K., & Kaushanskaya, M. (2007). The Language Experience and Proficiency Questionnaire (LEAP-Q): Assessing language profiles in bilinguals and multilinguals. *Journal of Speech, Language, and Hearing Research*, *50*(4), 940-967.

McRae, K., Cree, G. S., Seidenberg, M. S., & McNorgan, C. (2005). Semantic feature production norms for a large set of living and nonliving things. *Behavior Research Methods*, *37*(4), 547–559.

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

Miozzo, M., Costa, A., & Caramazza, A. (2002). The absence of a gender congruency effect in romance languages: A matter of stimulus onset asynchrony? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *28*(2), 388–391. https://doi.org/10.1037/0278-7393.28.2.388

Montag, E. D. (2006). Empirical formula for creating error bars for the method of paired comparison. *Journal of Electronic Imaging*, *15*(1), 010502–010502.

Montes-Alcalá, C., & Lapidus Shin, N. (2011). Las keys versus el key: Feminine gender assignment in mixed-language texts. *Spanish in Context*, *8*(1), 119–143. https://doi.org/10.1075/sic.8.1.06mon

Munarriz-Ibarrola, A., Ezeizabarrena, M.-J., Arrazola, V. D., & Parafita Couto, M. C. (2022). Gender assignment strategies and L1 effects in the elicited production of mixed Spanish-Basque DPs. *Linguistic Approaches to Bilingualism*, *12*(6), 778–815. https://doi.org/10.1075/lab.20016.mun

Münte, T. F., Heinze, H.-J., Matzke, M., Wieringa, B. M., & Johannes, S. (1998). Brain potentials and syntactic violations revisited: No evidence for specificity of the syntactic positive shift. *Neuropsychologia*, *36*(3), 217–226.

Myers, J., Gong, S.-P., Shen, Z.-G., & Min-Hsiung, C.-Y. (1999). *The semantic content of the general classifier in Mandarin*. International Association of Chinese Linguistics Eighth Annual Conference, Melbourne, Australia.

Myers, J., & Tsay, J. (2000). *The acquisition of the default classifier in Taiwanese.* Proceedings of the 7th International Symposium of Chinese Languages and Linguistics. 87–106.

Myers-Scotton, C. (1993). *Duelling Languages: Grammatical Structure in Codeswitching.* Oxford University Press. https://doi.org/10.1093/oso/9780198240594.001.0001

Myers-Scotton, C. (2002). *Contact linguistics: Bilingual encounters and grammatical outcomes*. Oxford University Press.

Navarro-Torres, C. A., Beatty-Martínez, A. L., Kroll, J. F., & Green, D. W. (2021). Research on bilingualism as discovery science. *Brain and Language*, *222*, 105014.

Neath, A. A., & Cavanaugh, J. E. (2012). The Bayesian information criterion: Background, derivation, and applications. *WIREs Computational Statistics*, *4*(2), 199–203. https://doi.org/10.1002/wics.199

Olson, D. J. (2022). The Bilingual Code-Switching Profile (BCSP) Assessing the reliability and validity of the BCSP questionnaire. *Linguistic Approaches to Bilingualism*, *14(3)*, 400–433.

Olson, D. J. (2024). Bilingual language experience and code-switching acceptability judgments: A constructive replication of the work by Stadthagen-González et al.(2019), Balam et al.(2020), and Stadthagen-González et al.(2018). *International Journal of Bilingualism*, 13 670069241298766.

O'Rourke, P. (2007). The gender congruency effect in bare noun production in Spanish. *Coyote Papers: Working Papers in Linguistics (Pp. 66–89), Linguistic Theory at the University of Arizona, Vol. 15*. http://hdl.handle.net/10150/126406

Osterhout, L., & Holcomb, P. J. (1992). Event-related brain potentials elicited by syntactic anomaly. *Journal of Memory and Language*, *31*(6), 785–806.

Osterhout, L., & Nicol, J. (1999). On the distinctiveness, independence, and time course of the brain responses to syntactic and semantic anomalies. *Language and Cognitive Processes*, *14*(3), 283–317.

Otheguy, R., & Lapidus, N. (2003). An adaptive approach to noun gender in New York contact Spanish. *A Romance Perspective on Language Knowledge and Use*, 209–229.

Pablos, L., Parafita Couto, M. C., Boutonnet, B., De Jong, A., Perquin, M., de Haan, A., & Schiller, N. O. (2019). Adjective-noun order in Papiamento-Dutch code-switching. *Linguistic Approaches to Bilingualism*, *9*(4–5), 710–735.

Paolieri, D., Lotto, L., Leoncini, D., Cubelli, R., & Job, R. (2011). Differential effects of grammatical gender and gender inflection in bare noun production: Gender and morphology in picture naming. *British Journal of Psychology*, *102*(1), 19–36. https://doi.org/10.1348/000712610X496536

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

Paquot, M., & Callies, M. (2020). Promoting methodological expertise, transparency, replication, and cumulative learning: Introducing new manuscript types in the International Journal of Learner

Corpus Research. *International Journal of Learner Corpus Research*, *6*(2), 121–124.

Parafita Couto, M., Bellamy, K., Ameka, F., Cabrelli, J., Chaouch-Orozco, A., González Alonso, J., Pereira Soares, S., Puig-Mayenco, E., & Rothman, J. (2023). Theoretical linguistic approaches to multilingual code-switching. *The Cambridge Handbook of Third Language Acquisition and Processing*, 403–436.

Parafita Couto, M. C., Bellamy, K., & Valdés Kroff, J. R. (in press). Nominal classification in code-switching. In *The Oxford Handbook of Gender and Classifiers*. Oxford University Press.

Parafita Couto, M. C., Boutonnet, B., Hoshino, N., Webb-Davies, P., Deuchar, M., & Thierry, G. (2017). Testing alternative theoretical accounts of code-switching using event-related brain potentials: A pilot study on Welsh-English. In *Bilingualism and minority languages in Europe: Current trends and Developments* (pp. 240–254). Cambridge Scholars Publishing.

Parafita Couto, M. C., Deuchar, M., & Fusser, M. (2015). How do Welsh-English bilinguals deal with conflict? Adjective-noun order resolution. *Code-Switching between Structural and Sociolinguistic Perspectives*, 65–84.

Parafita Couto, M. C., & Gullberg, M. (2019). Code-switching within the noun phrase: Evidence from three corpora. *International Journal of Bilingualism*, *23*(2), 695–714.

Parafita Couto, M. C., Munarriz, A., Epelde, I., Deuchar, M., & Oyharçabal, B. (2016). Gender conflict resolution in Spanish–Basque mixed DPs. *Bilingualism: Language and Cognition*, *19*(4), 834–853.

Parafita Couto, M. C., & Stadthagen-González, H. (2019). *El book or the libro?* Insights from acceptability judgments into determiner/noun code-switches. *International Journal of Bilingualism*, *23*(1), 349–360. https://doi.org/10.1177/1367006917728392

Paul, W. (2010). Adjectives in Mandarin Chinese. *Adjectives: Formal Analyses in Syntax and Semantics*, *153*, 115.

Polio, C. (1994). Non-native speakers' use of nominal classifiers in Mandarin Chinese. *Journal of the Chinese Language Teachers Association*, *29*(3), 51–66.

Poplack, S. (1980). Sometimes I'll start a sentence in Spanish Y TERMIN O EN ESPAÑOL: toward a typology of code-switching. *Linguistics*, *18*(7–8). https://doi.org/10.1515/ling.1980.18.7-8.581

Poplack, S., & Meechan, M. (1998). Introduction: How languages fit together in codemixing. *International Journal of Bilingualism*, *2*(2), 127–138.

Poplack, S., Pousada, A., & Sankoff, D. (1982). Competing influences on gender assignment: Variable process, stable outcome. *Lingua*, *57*(1), 1–28.

Puhacheuskaya, V. (2021). *The "Semantic P600": A Brief Review*. University of Alberta. https://doi.org/10.13140/RG.2.2.26903.9 1042

Qian, Z. (in press). The psycholinguistics of classifiers. In *Oxford Handbook of Gender and Classifiers*. Oxford University Press.

R Core Team (2020). *R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. [Computer software]*. (n.d.). [Computer software]. https://www.R-project.org/.

Real Academia Española:Diccionario de la lengua española [Dictionary of the Spanish Language] (23. ª ed). (2014). <https://dle.rae.es>

Rekun, O., & Meir, N. (2024). Two gender systems in a bilingual mind: A study of gender assignment in code-switched Russian-Hebrew adjective-noun phrases. *Ampersand*, 100189. https://doi.org/10.1016/j.amper.2024.100189

Roberts, P. M., Garcia, L. J., Desrochers, A., & Hernandez, D. (2002). English performance of proficient bilingual adults on the Boston Naming Test. *Aphasiology*, *16*(4–6), 635–645.

Robles-Llana, P. (2018). Cultural identities of children of Chinese migrants in Spain: A critical evaluation of the category 1.5 generation. *Identity*, *18*(2), 124–140.

Roca. (1989). The organisation of grammatical gender. *Transactions of the Philological Society*, *87*(1), 1–32.

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

Roelofs, A. (1993). Testing a non-decompositional theory of lemma retrieval in speaking: Retrieval of verbs. *Cognition*, *47*(1), 59–87.

Roelofs, A. (1998). Lemma selection without inhibition of languages in bilingual speakers. *Bilingualism: Language and Cognition*, *1*(2), 94–95.

Roelofs, A. (2003). Goal-referenced selection of verbal action: Modeling attentional control in the Stroop task. *Psychological Review*, *110*(1), 88.

Rosinski, R. R. (1977). Picture-word interference is semantically based. *Child Development*, 643–647.

Rosinski, R. R., Golinkoff, R. M., & Kukish, K. S. (1975). Automatic Semantic Processing in a Picture-Word Interference Task. *Child Development*, *46*(1), 247. https://doi.org/10.2307/1128859

Sá Leite, A. R., Haro, J., Comesaña, M., & Fraga, I. (2021). Of Beavers and Tables: The Role of Animacy in the Processing of Grammatical Gender Within a Picture-Word Interference Task. *Frontiers in Psychology*, *12*, 661175. https://doi.org/10.3389/fpsyg.2021.661175

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. https://doi.org/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. https://doi.org/10.1016/j.cognition.2022.105060

Schiller, N. O. (2013). Psycholinguistic approaches to the investigation of grammatical gender. In G. G. Corbett (Ed.), *The Expression of Gender* (pp. 161–190). De Gruyter. https://doi.org/10.1515/9783110307337.161

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

Schiller, N. O., & Caramazza, A. (2006). Grammatical gender selection and the representation of morphemes: The production of Dutch diminutives. *Language and Cognitive Processes*, *21*(7–8), 945–973. https://doi.org/10.1080/01690960600824344

Schiller, N. O., & Costa, A. (2006). Different selection principles of freestanding and bound morphemes in language production. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *32*(5), 1201–1207. https://doi.org/10.1037/02787393.32.5.1201

Schiller, N. O., & Sá Leite, A. R. (in press). The psycholinguistics of grammatical gender. In N. O. Schiller & T. Kupisch (Eds.), *Oxford Handbook of Gender and Classifiers* (pp. XXX-XXX). New York, Oxford: Oxford University Press.

Schneider, W., Eschman, A., & Zuccolotto, A. (2002). *E-prime User's Guide*. Pittsburgh: Psychology Software Tools, Inc.

Schriefers, H. (1993). Syntactic processes in the production of noun phrases. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *19*(4), 841–850. https://doi.org/10.1037/02787393.19.4.841

Schriefers, H., & Jescheniak, J. D. (1999). Representation and processing of grammatical gender in language production: A review. *Journal of Psycholinguistic Research*, *28*, 575–600.

Schriefers, H., Meyer, A. S., & Levelt, W. J. (1990). Exploring the time course of lexical access in language production: Picture-word interference studies. *Journal of Memory and Language*, *29*(1), 86–102.

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. https://doi.org/10.1037/0278-7393.26.6.1368

Seifart, F. (2010). Nominal classification. *Language and Linguistics Compass*, *4*(8), 719–736.

Severens, E., Lommel, S. V., Ratinckx, E., & Hartsuiker, R. J. (2005). Timed picture naming norms for 590 pictures in Dutch. *Acta Psychologica*, *119*(2), 159–187. https://doi.org/10.1016/j.actpsy.2005.01.002

Shao, J. (1993). 量词的语义分析及其与名词的双向选择. *中国语文*, *3*(289), 2.

Sönning, L., & Werner, V. (2021). The replication crisis, scientific revolutions, and linguistics. *Linguistics*, *59*(5), 1179–1206.

Stadthagen-González, H., López, L., Parafita Couto, M. C., & Párraga, C. A. (2018). Using two-alternative forced choice tasks and Thurstone's law of comparative judgments for code-switching research. *Linguistic Approaches to Bilingualism*, *8*(1), 67–97.

Stadthagen-González, H., Parafita Couto, M. C., Párraga, C. A., & Damian, M. F. (2019). Testing alternative theoretical accounts of code-switching: Insights from comparative judgments of adjective–noun order. *International Journal of Bilingualism*, *23*(1), 200–220.

Starreveld, P., & La Heij, W. (2004). Phonological facilitation of grammatical gender retrieval. *Language and Cognitive Processes*, *19*(6), 677–711. https://doi.org/10.1080/01690960444000061

Terker, A. (1985). On Spanish adjective position. *Hispania*, *68*(3), 502–509.

Teschner, R. V., & Russell, W. M. (1984). The gender patterns of Spanish nouns: An inverse dictionary-based analysis. *Hispanic Linguistics*, *1*(1), 115–132.

Thurstone, L. L. (1994). A law of comparative judgment. *Psychological Review*, *101*(2), 266.

Torres Cacoullos, R., & Vélez Avilés, J. (2024). Mixing adjectives: A variable equivalence hypothesis for bilingual word order conflicts. *Linguistic Approaches to Bilingualism*, *14*(5), 609–639.

Tzeng, O. J., Chen, S., & Hung, D. L. (1991). The classifier problem in Chinese aphasia. *Brain and Language*, *41*(2), 184–202.

Valdés Kroff, J. R. (2016). Mixed NPs in Spanish-English bilingual speech: Using a corpus-based approach to inform models of sentence processing. In R. E. Guzzardo Tamargo, C. M. Mazak, & M. C. Parafita Couto (Eds.), *Issues in Hispanic and Lusophone Linguistics* (Vol. 11, pp. 281–300). John Benjamins Publishing Company. https://doi.org/10.1075/ihll.11.12val

Van Casteren, M., & Davis, M. H. (2006). Mix, a program for pseudorandomization. *Behavior Research Methods*, *38*(4), 584–589. https://doi.org/10.3758/BF03193889

Van Herten, M., Kolk, H. H., & Chwilla, D. J. (2005). An ERP study of P600 effects elicited by semantic anomalies. *Cognitive Brain Research*, *2*(22), 241–255.

Van Osch, B., Parafita Couto, M. C., Boers, I., & Sterken, B. (2023). Adjective position in the code-switched speech of Spanish and Papiamento heritage speakers in the Netherlands: Individual differences and methodological considerations. *Frontiers in Psychology*, *14*, 1136023.

Vanden Wyngaerd, E. (2017). The adjective in Dutch-French codeswitching: Word order and agreement. *International Journal of Bilingualism*, *21*(4), 454–473.

Vaughan-Evans, A., Parafita Couto, M. C., Boutonnet, B., Hoshino, N., Webb-Davies, P., Deuchar, M., & Thierry, G. (2020). Switchmate! An electrophysiological attempt to adjudicate between competing accounts of adjective-noun code-switching. *Frontiers in Psychology*, *11*, 549762.

Vicente-Saez, R., & Martinez-Fuentes, C. (2018). Open Science now: A systematic literature review for an integrated definition. *Journal of Business Research*, *88*, 428–436.

Vinther, T. (2002). Elicited imitation: A brief overview. *International Journal of Applied Linguistics*, *12*(1), 54–73.

Voeten, C. (2019). *permutes: Permutation tests for time series data (1.0) [Computer software].*

Von Grebmer zu Wolfsthurn, S., Pablos, L., & Schiller, N. O. (2021). Noun-phrase production as a window to language selection: An ERPstudy. *Neuropsychologia*, *162*, 108055. https://doi.org/10.1016/j.neuropsychologia.2021.108055

Wang, J., Witteman, J., & Schiller, N. (submitted). Processing Mandarin Chinese classifiers as a lexico-syntactic feature during noun phrase production. *Submitted*. https://ssrn.com/abstract=5247074

Wang, L., Guo, J., Bi, Y., & Shu, H. (2006). Classifier congruency effect in the production of noun phrases. *Studies of Psychology and Behavior*, *4*(1), 34.

Wang, M., Chen, Y., & Schiller, N. O. (2019). Lexico-syntactic features are activated but not selected in bare noun production: Electrophysiological evidence from overt picture naming. *Cortex*, *116*, 294–307. https://doi.org/10.1016/j.cortex.2018.05.014

Wang, M., & Schiller, N. O. (2019). A Review on Grammatical Gender Agreement in Speech Production. *Frontiers in Psychology, 9,* 2754. https://doi.org/10.3389/fpsyg.2018.02754

Wang, M., & Schiller, N. O. (in press). The neurolinguistics of classifiers. In *Oxford Handbook of Gender and Classifiers*. Oxford University Press.

Wang, S., & Schiller, N. O. (*submitted*). Classifier processing in L1 and L2 noun phrase production: Behavioral and electrophysiological evidence. *Submitted*.

Wang, W. S. (1973). The Chinese language. *Scientific American*, *228*(2), 50–63.

Wang, Y., Witteman, J., & Schiller, N. O. (2024). The role of animacy in language production: Evidence from bare noun naming. *Quarterly Journal of Experimental Psychology*, 17470218241281868.

White, L., Valenzuela, E., Kozlowska–Macgregor, M., & Leung, Y.-K. I. (2004). Gender and number agreement in nonnative Spanish. *Applied Psycholinguistics, 25*(01). https://doi.org/10.1017/S0142716404001067

Wicha, N. Y., Moreno, E. M., & Kutas, M. (2003). Expecting gender: An event related brain potential study on the role of grammatical gender in comprehending a line drawing within a written sentence in Spanish. *Cortex, 39*(3), 483–508.

Wu, R., Parafita Couto, M. C., & Schiller, N. (in press). Variation in Gender Assignment Strategies in Mixed Spanish-Chinese Noun Phrases: Insights from a Multilingual Community in Barcelona. *International Journal of Bilingualism*.

Wu, R., & Schiller, N. O. (2023). Gender Congruency Effects in Spanish: Behavioral Evidence from Noun Phrase Production. *Brain Sciences*, *13*(4), 696.

Xun et al. (2016). 大数据背景下 BCC 语料库的研制. In *BCC* (Vols. 1–03). http://bcc.blcu.edu.cn/

Yan, X., Lei, Y., & Shih, C. (2020). A corpus-driven, curriculum-based Chinese elicited imitation test in US universities. *Foreign Language Annals*, *53*(4), 704–732.

Yang, R. (1998). Chinese Bare Nouns as Kind-Denoting Terms. *RuLing Papers*, *1*(1), 247-288.

Zhang, H. (2007). Numeral classifiers in mandarin Chinese. *Journal of East Asian Linguistics*, *16*(1), 43–59.

Zhang, N. N. (2013). Classifier structures in Mandarin Chinese (vol. 263). *Berlin and Boston, MA: Walter de Gruyter*.

Zhang, Y., Zhang, J., & Min, B. (2012). Neural dynamics of animacy processing in language comprehension: ERP evidence from the interpretation of classifier–noun combinations. *Brain and Language*, *120*(3), 321–331.

Zhou, X., Jiang, X., Ye, Z., Zhang, Y., Lou, K., & Zhan, W. (2010). Semantic integration processes at different levels of syntactic hierarchy during sentence comprehension: An ERP study. *Neuropsychologia*, *48*(6), 1551–1562.

Zhu, X., Damian, M. F., & Zhang, Q. (2015). Seriality of semantic and phonological processes during overt speech in Mandarin as revealed by event-related brain potentials. *Brain and Language*, *144*, 16–25.

Summary

This dissertation investigates how bilingual speakers process lexico-syntactic features, including grammatical gender, classifier systems, and the linear order of adjectives and nouns, across Spanish and Chinese in both unilingual and bilingual contexts. The central focus is on how early Spanish–Chinese bilinguals, particularly those residing in Barcelona, Spain, process and produce grammatical gender in Spanish and classifiers in Mandarin Chinese. It also examines how these bilinguals resolve syntactic conflicts arising from differences in adjective placement across the two languages, where Spanish typically places adjectives postnominally while Chinese places them prenominally. Drawing on behavioral and, where relevant, electrophysiological data, and employing a multi-task approach that integrates elicitation, repetition, and acceptability judgment tasks, this dissertation provides a comprehensive account of bilingual morphosyntactic processing in both experimental and semi-naturalistic contexts. The findings contribute to broader discussions in bilingualism and psycholinguistics by illuminating the mechanisms through which bilinguals reconcile distinct grammatical systems in language production and comprehension.

This dissertation is divided into seven chapters.

Chapter 1 introduces the research background and the core research questions and provides a concise overview of the main content of each chapter.

Chapter 2 begins by examining how grammatical gender, as a lexico-syntactic feature, is processed during NP production in unilingual Spanish. Using a picture-word interference (PWI) paradigm with Spanish speakers, we tested whether gender congruency and semantic relatedness affect picture naming latencies. The goal was to determine if grammatical gender is activated and competitively selected during speech production and whether this process results in a gender congruency effect. This chapter is important because it clarifies the inconsistent findings of previous resear-

ch on Spanish NP production and provides empirical evidence for the gender congruency effect. It also sets the foundation for later chapters by establishing how a key lexico-syntactic feature, i.e., grammatical gender, is processed in a unilingual context.

Chapter 3 shifts from unilingual Spanish contexts to bilingual Spanish–Chinese contexts. It focuses on how early Spanish–Chinese bilinguals assign Spanish grammatical gender to Chinese nouns when processing and producing mixed Spanish–Chinese NPs. By using a multi-task approach, this study identifies which gender categories (i.e., masculine or feminine) in Spanish are most often assigned to Chinese nouns and what strategies bilinguals use during this process. The findings suggest that gender assignment in bilinguals is modulated by task demands. Specifically, in language production, bilinguals predominantly employed the masculine default strategy alongside the translation equivalent strategy. In contrast, during language comprehension, participants showed a stronger reliance on the translation equivalent strategy. These results underscore the adaptive nature of bilingual language processing, demonstrating that the selection of gender assignment strategies is context-sensitive and dynamically shaped by the specific cognitive requirements of the task. This chapter contributes to our understanding of how bilinguals adapt language-specific features (i.e., grammatical gender in Spanish) to non-gendered items (i.e., Chinese nouns), especially across typologically different languages. It also links back to Chapter 2 by exploring whether the mechanisms observed among Spanish speakers with unilingual contexts are comparable to those among bilingual speakers within bilingual contexts.

In **Chapter 4**, we shift attention to another language in the bilingual language pair, Mandarin Chinese, and investigate how another type of abstract lexico-syntactic feature, the numeral classifier, is processed and produced in Chinese NP production. This study used picture-naming tasks and ERP (electrophysiological) measures to investigate whether classifiers are competitively selected during Mandarin Chinese NP production by both Mandarin Chinese speakers and early Spanish–Chinese bilinguals. Specifically, it investigated participants' naming latencies, assessing the effects of classifier congruency and semantic relatedness between target pictures and distractor words, as well as the potential elicitation of specific ERP components. The findings identified a robust classifier congruency effect in both groups, supporting the idea that classifiers as lexico-syntactic features are competitively selected during NP production. This chapter parallels the findings on grammatical gender in Chapter 2 by focusing on the processing of lexico-syntactic features, while extending the investigation to a different language system. It further reinforces the parallels between classifier selection in Chinese and grammatical gender assignment in Spanish, suggesting shared underlying selection mechanisms.

In **Chapter 5**, we extend the investigation of classifier processing into bilingual contexts by examining how early Spanish–Chinese bilinguals assign classifiers to Spanish nouns in mixed Chinese–Spanish NPs. Using the same multi-task approach as in Chapter 3, this chapter explores how bilinguals assign classifiers to Spanish nouns during code-switching and identifies the strategies they use in classifier assignment. The results indicate that the strategies employed by bilinguals in classifier assignment are largely consistent with those reported in Chapter 3. The results provide important insight into how classifier assignment is managed during code-switching and how structural compatibility and language experience shape the assignment. Together with Chapter 4, this chapter completes the analysis of classifier use in both unilingual and bilingual contexts, in parallel to how Chapters 2 and 3 addressed grammatical gender. Furthermore, the classifier assignment strategies identified here show notable parallels with the grammatical gender assignment strategies discussed in Chapter 3, reinforcing the broader patterns in bilingual morphosyntactic processing.

Chapter 6 moves beyond language-specific lexico-syntactic features to language-shared lexico-syntactic features, focusing on adjective-noun order, which is shared by both Spanish and Chinese but realized differently in each language. Spanish features postnominal adjectives such as *manzana roja* 'apple red', while Chinese places adjectives prenominally like *红色的苹果* 'red apple'. Specifically, this study examines how Spanish–Chinese bilinguals navigate adjective-noun word order conflict and assesses how predictions from the Matrix Language Frame (MLF) model and Minimalist Program (MP)-based approaches account for adjective–noun patterns in Spanish–Chinese code-switched constructions. Findings indicate that production data align with MLF predictions when Spanish served as the matrix language, but MP accounts better for overall production and comprehension patterns. In Chinese matrix contexts, both models were supported. This chapter is the first to empirically examine how grammatical constraints, derived from the MLF and MP, interact with Spanish–Chinese CS patterns, contributing to a growing body of research on adjective placement in bilingual speech. Moreover, our findings challenge the idea that a single theoretical model fully explains code-switching patterns; instead, they strongly suggest that both matrix language structure (MLF) and the features of individual lexical items (MP) interact and contribute to mixed constructions, leading to a more fine-grained understanding of syntactic integration. This chapter builds on the previous chapters by shifting the focus from language-specific to language-shared but conflicting lexico-syntactic features, making them a valuable site for studying how bilinguals manage syntactic differences across languages and providing a deeper understanding of how bilinguals resolve grammatical differences across languages in production and comprehension.

Chapter 7 revisits the central research questions and synthesizes the main findings of the preceding chapters. It further elaborates on the theoretical and empirical implications of all studies, critically acknowledges their limitations, and outlines potential directions for future research.

Samenvatting

Dit proefschrift onderzoekt hoe tweetalige sprekers lexicosyntactische kenmerken, zoals grammaticaal geslacht, classificatorensystemen en de volgorde van adjectieven en zelfstandige naamwoorden, hanteren in zowel eentalige als tweetalige contexten van het Spaans en het Chinees. De studie richt zich met name op vroege Spaans–Chinees tweetaligen, in het bijzonder degenen die in Barcelona wonen, en onderzoekt hoe zij grammaticaal geslacht in het Spaans en numerieke classificatoren in het Mandarijn verwerken en produceren. Daarnaast wordt onderzocht hoe deze tweetaligen syntactische conflicten oplossen die voortkomen uit verschillen in de positie van adjectieven in beide talen, waarbij het adjectief in het Spaans doorgaans na het zelfstandig naamwoord staat, terwijl het in het Chinees ervoor staat. Het onderzoek combineert gedragsmatige en, waar relevant, elektrofysiologische gegevens, en maakt gebruik van een multitaskaanpak met elicitatie-, herhalings- en aanvaardbaarheidsoordeeltaken. Zo biedt het proefschrift een uitgebreid beeld van morfosyntactische verwerking bij tweetaligen in zowel experimentele als seminatuurlijke contexten. De bevindingen dragen bij aan bredere discussies binnen het tweetaligheidsonderzoek en de psycholinguïstiek, door inzicht te geven in de mechanismen waarmee tweetaligen verschillende grammaticale systemen integreren in taalproductie en -begrip.

Het proefschrift is opgebouwd uit zeven hoofdstukken.

Hoofdstuk 1 introduceert de onderzoeksachtergrond en de centrale onderzoeksvragen en biedt een beknopt overzicht van de inhoud van de afzonderlijke hoofdstukken.

Hoofdstuk 2 onderzoekt hoe grammaticaal geslacht, als lexicosyntactisch kenmerk, verwerkt wordt tijdens de productie van naamwoordgroepen in eentalig Spaans. Er wordt gebruikgemaakt van een picture-word interference (PWI)-paradigma om te testen of geslachtscongruentie en semantische gerelateerdheid de benoemlatentie beïnvloeden bij Spaanstalige deelnemers. Het doel is om vast te stellen of grammaticaal geslacht geactiveerd en competitief geselecteerd wordt tijdens spraakproductie en of dit leidt tot een geslachtscongruentie-effect. Dit hoofdstuk is van belang omdat het de inconsistente bevindingen uit eerder onderzoek naar de productie van Spaanse naamwoordgroepen verheldert en empirisch bewijs levert voor het geslachtscongruentie-effect. Bovendien vormt het de basis voor de latere hoofdstukken door aan te tonen hoe een cruciaal lexicosyntactisch kenmerk, het grammaticale geslacht, in een eentalige context wordt verwerkt.

Hoofdstuk 3 verschuift de focus van eentalige Spaanse contexten naar tweetalige Spaans–Chinese contexten. Het richt zich op hoe vroege Spaans–Chinese tweetaligen Spaans grammaticaal geslacht toewijzen aan Chinese zelfstandige naamwoorden bij het verwerken en produceren van gemengde Spaans–Chinese nominale woordgroepen. Door gebruik te maken van een multitaskbenadering onderzoekt deze studie welke Spaanse geslachtscategorieën (mannelijk of vrouwelijk) het vaakst worden toegewezen aan Chinese zelfstandige naamwoorden en welke strategieën tweetaligen tijdens dit toewijzingsproces toepassen. De bevindingen suggereren dat de toewijzing van grammaticaal geslacht bij tweetaligen wordt beïnvloed door de taakvereisten. Specifiek gebruikten tweetaligen tijdens taalproductie voornamelijk de mannelijke standaardstrategie, aangevuld met de vertaal equivalente strategie. Daarentegen vertoonden de deelnemers tijdens taalbegrip een duidelijkere afhankelijkheid van de vertaalequivalente strategie. Deze resultaten benadrukken het adaptieve karakter van tweetalig taalverwerking en laten zien dat de selectie van geslachtstoewijzingsstrategieën contextafhankelijk is en dynamisch wordt bepaald door de specifieke cognitieve eisen van de taak. Dit hoofdstuk draagt bij aan ons begrip van hoe tweetaligen taalspecifieke kenmerken (d.w.z. grammaticaal geslacht in het Spaans) toepassen op niet-genderde elementen (d.w.z. Chinese zelfstandige naamwoorden), vooral in het kader van typologisch verschillende talen. Bovendien sluit het aan bij Hoofdstuk 2 door te onderzoeken in hoeverre de mechanismen die bij eentalige

Spaanstalige sprekers worden waargenomen, vergelijkbaar zijn met die bij tweetalige sprekers in tweetalige contexten.

Hoofdstuk 4 verschuift de aandacht naar de andere taal binnen het tweetalige taalpaar, het Mandarijn, en onderzoekt hoe een ander type abstract lexicosyntactisch kenmerk, de numerale classificator, wordt verwerkt en geproduceerd bij de productie van Mandarijnse nominale woordgroepen. Deze studie maakt gebruik van picture-namingtaken in combinatie met ERP (elektrofysiologische) metingen om te onderzoeken of classificatoren competitief worden geselecteerd tijdens de productie van Mandarijnse nominale woordgroepen door zowel moedertaalsprekers Mandarijn als vroege Spaans–Chinese tweetaligen. Specifiek worden de reactietijden van de deelnemers onderzocht, waarbij de effecten van congruentie van classificatoren en semantische verwantschap tussen doelafbeeldingen en afleiderwoorden worden beoordeeld, evenals het potentiële optreden van specifieke ERP-componenten. De bevindingen tonen een robuust classificator-congruentie-effect in beide groepen, wat ondersteunt dat classificatoren als lexicosyntactische kenmerken competitief worden geselecteerd tijdens de productie van nominale woordgroepen. Dit hoofdstuk vertoont parallellen met de bevindingen over grammaticaal geslacht in Hoofdstuk 2 door de focus te leggen op de verwerking van lexicosyntactische kenmerken, terwijl het onderzoek wordt uitgebreid naar een ander taalsysteem. Het versterkt bovendien de parallellen tussen classificatorselectie in het Mandarijn en grammaticaal geslachtstoewijzing in het Spaans, wat wijst op gedeelde onderliggende selectiemechanismen.

In **Hoofdstuk 5** wordt het onderzoek naar de verwerking van classificatoren uitgebreid naar tweetalige contexten door te onderzoeken hoe vroege Spaans–Chinese tweetaligen classificatoren toewijzen aan Spaanse zelfstandige naamwoorden in gemengde Chinese Spaanse nominale woordgroepen. Met gebruik van dezelfde multitaskbenadering als in Hoofdstuk 3 onderzoekt deze studie hoe tweetaligen classificatoren toewijzen aan Spaanse zelfstandige naamwoorden tijdens code-switching en identificeert de strategieën die zij daarbij hanteren. De resultaten geven aan dat de strategieën die tweetaligen bij classificatortoewijzing toepassen grotendeels overeenkomen met die in Hoofdstuk 3 gerapporteerd. De bevindingen bieden belangrijke inzichten in hoe classificatortoewijzing wordt beheerd tijdens code-switching en hoe structurele compatibiliteit en taalervaring de toewijzing beïnvloeden. Samen met Hoofdstuk 4 voltooit dit hoofdstuk de analyse van het gebruik van classificatoren in zowel eentalige als tweetalige contexten, parallel aan de manier waarop Hoofdstukken 2 en 3 grammaticaal geslacht behandelden. Bovendien vertonen de hier geïdentificeerde strategieën voor classificatortoewijzing opvallende parallellen met de strategieën voor grammaticaal geslachtstoewijzing die in Hoofdstuk 3 zijn besproken, wat de bredere patronen in tweetalige morfosyntactische verwerking versterkt.

Hoofdstuk 6 gaat verder dan taalspecifieke lexicosyntactische kenmerken en richt zich op taalgedeelde lexicosyntactische kenmerken, met name de volgorde van adjectief en substantief. Deze volgorde komt zowel in het Spaans als in het Chinees voor, maar wordt in elke taal op een andere manier gerealiseerd. In het Spaans staan adjectieven meestal achter het zelfstandig naamwoord, zoals *manzana roja* ('rode appel'), terwijl in het Chinees het adjectief vóór het zelfstandig naamwoord staat, zoals *红色的苹果* ('rode appel'). Dit onderzoek onderzoekt hoe Spaans–Chinese tweetaligen omgaan met het woordvolgordeconflict tussen adjectief en substantief. Het evalueert in welke mate voorspellingen van het Matrix Language Frame (MLF)-model en op het Minimalist Program (MP) gebaseerde benaderingen de patronen van adjectief–substantiefcombinaties in Spaans–Chinees codeswitchen kunnen verklaren. De resultaten tonen aan dat de productiegegevens overeenkomen met de voorspellingen van het MLF wanneer het Spaans fungeert als matrixtaal, terwijl het MP beter verklaart hoe de algemene productie- en begrippatronen verlopen. In Chinese matrixtaalcontexten werden beide modellen ondersteund. Dit hoofdstuk is het eerste dat empirisch onderzoekt hoe grammaticale beperkingen, afgeleid van het MLF en het MP, interageren met Spaans–Chinees codeswitchgedrag, en draagt zo bij aan het groeiende onderzoek naar adjectiefplaatsing in tweetalige spraak. Bovendien betwisten de bevindingen het idee dat één enkel theoretisch model codeswitchpatronen volledig kan verklaren; in plaats daarvan suggereren ze sterk dat zowel de structuur van de matrixtaal (MLF) als de eigenschappen van individuele lexicale items (MP) met elkaar interageren en bijdragen aan gemengde constructies, wat leidt tot een meer verfijnd begrip van syntactische integratie. Dit hoofdstuk bouwt voort op de voorgaande hoofdstukken door de focus te verleggen van taalspecifieke naar taalgedeelde maar conflicterende lexicosyntactische kenmerken, die een waardevolle invalshoek vormen om te onderzoeken hoe tweetaligen syntactische verschillen tussen talen hanteren. Het biedt zo een dieper inzicht in hoe tweetaligen grammaticale verschillen tussen talen oplossen in zowel productie als begrip.

Hoofdstuk 7 keert terug naar de centrale onderzoeksvragen en synthetiseert de belangrijkste bevindingen van de voorgaande hoofdstukken. Daarnaast worden de theoretische en empirische implicaties besproken, de beperkingen van de studies erkend en suggesties gedaan voor toekomstig onderzoek.

摘要

本论文探讨了双语者在单语与双语语境中如何处理与协调西班牙语与汉语中的词汇句法特征，包括语法性别、量词系统以及形容词与名词的语序。研究的核心主要考察西班牙语－汉语双语者（尤其是居住在西班牙巴塞罗那的群体）在语言加工与产出过程中如何处理西班牙语中的语法性别与汉语中的量词。同时，本论文还探讨双语者如何应对两种语言在形容词与名词语序上的差异：西班牙语通常将形容词置于名词之后，而汉语则置于名词之前。本论文结合行为数据与电生理数据，并采用包括诱发、重复和可接受性判断任务在内的多任务范式，系统考察双语者在实验性与半自然语境下的词汇句法加工过程。此外，本论文还揭示了双语者在语码转换过程中如何处理与应对跨语言的语法差异，重点涉及语法性别赋值、汉语量词选择以及形容词与名词语序差异等。本论文的研究结果不仅为双语研究与心理语言学领域提供了新的视角，也深化了我们对双语者在语言理解与产出中协调处理不同语法系统机制的理解。

本论文共由七章组成。

第一章主要介绍了研究背景和核心研究问题，并简明概述了各章节的主要研究内容。

第二章聚焦语法性别这一词汇句法特征在西班牙语单语语境下的加工与产出机制。本研究旨在探讨语法性别在言语产出过程中是否会被激活并参与竞争性选择，以及这一过程是否会诱发“性别一致性效应”。具体来说，本研究采用图–词干扰（PWI）实

验范式，考察西班牙语母语者在图片命名任务中的反应时是否受到图片与干扰词之间的性别一致性与语义相关性的影响。本章梳理并整合了以往关于西班牙语名词短语产出中“性别一致性效应”的研究分歧，并通过实验证据证实了该效应的存在。本研究还阐释了语法性别作为词汇句法特征在西班牙语单语语境下的加工机制，从而为后续章节奠定理论基础。

第三章聚焦语法性别在双语语境下的加工与产出机制，旨在探讨西班牙语–汉语双语者在理解和产出西班牙语–汉语混合名词短语时，如何将西班牙语的语法性别应用于非性别范畴的汉语名词。通过多任务实验，本研究揭示了双语者在语法性别赋值过程中更常使用的性别范畴（阳性或阴性）及其赋值策略。研究结果表明，性别赋值存在实验任务依赖性的差异。在语言产出中，双语者主要采用了阳性默认策略以及翻译对应策略。在语言理解中，受试者则更倾向于采用翻译对应策略。本研究突显了双语语言加工的灵活性，表明性别赋值策略会根据不同任务的要求发生变化。本章与第二章形成呼应，将语法性别的研究从西班牙语单语语境拓展至西班牙语-汉语双语语境，并进一步考察西班牙语母语者中观察到的语法性别加工机制是否同样适用于双语者。

第四章聚焦量词在汉语单语语境下的加工与产出机制，旨在考察汉语母语者和西班牙语–汉语双语者在汉语单语语境下如何加工与产出汉语量词。本研究结合图片命名任务与事件相关电位（ERP）技术，探讨汉语母语者与西班牙语–汉语双语者在产出汉语名词短语时，汉语量词否会被激活并进行竞争性选择。具体来说，本研究考察汉语母语者和西班牙语-汉语双语者在图片命名任务中的反应时，评估图片与干扰词之间的量词一致性与语义相

关性是否对反应时产生影响，并观察其是否会诱发特定的 ERP 成分。研究结果表明，两组受试者均表现出显著的“量词一致性效应”, 支持了量词作为词汇句法特征在名词短语产出过程中存在竞争性选择的观点。本章与第二章关于语法性别的研究相呼应，将研究从西班牙语语法性别扩展至汉语量词系统，同时揭示了汉语量词与西班牙语语法性别在单语语言产出过程中呈现出相似模式，表明两者可能依赖共通的选择与加工机制。

第五章将量词研究拓展至双语语境，旨在探讨西班牙语–汉语双语者在理解和产出汉语–西班牙语混合名词短语时，如何将汉语量词赋予西班牙语名词。通过采用与第三章相同的多任务方法，本研究揭示了双语者在语码转换过程中主要使用的量词赋值策略。研究结果表明，双语者在量词赋值过程中采用的策略与第三章所述基本一致。本章与第四章相辅相成，进一步完善了对量词在单语与双语语境下处理与加工机制的系统研究，同时也与第二、三章关于语法性别加工机制的探讨形成呼应。此外，本章所揭示的量词赋值策略与第三章的语法性别赋值策略呈现出高度一致性，这一结果进一步说明了词汇句法特征在双语加工过程中可能依赖于共通机制。

第六章聚焦于西班牙语和汉语中同属存在，但在实现形式上存在差异的句法结构，即形容词–名词语序。西班牙语通常将形容词置于名词之后，例如 *manzana roja* (“苹果红色的”)，而汉语则相反，如*红色的苹果* (“红色的苹果”)。本研究旨在考察西班牙语–汉语双语者在西班牙语–汉语混合名词短语中如何处理由形容词–名词语序差异引发的冲突。同时，本研究还评估矩阵语言框架模型 (MLF) 与最简方案 (MP) 所提出的句法约束，在多大程度上

能够预测西班牙语–汉语双语者在混合名词短语中的语序选择。研究结果显示，当西班牙语作为矩阵语言时，语言产出数据与矩阵语言框架模型 (MLF) 的预测一致，而最简方案 (MP) 则能更好地解释整体的语言产出和理解数据。在汉语矩阵语言语境中，两种模型的预测在语言产出和理解数据中均得到证实。本章首次通过实证研究探讨了源自矩阵语言框架模型 (MLF) 和最简方案 (MP) 的语法限制如何与西班牙–汉语语码转换模式相互作用，为双语者在语言产出与理解中形容词–名词语序的研究提供了新的实证依据。此外，我们的研究结果挑战了单一理论模型能够完全解释语码转换模式的观点；相反，实证证据表明，矩阵语言框架模型 (MLF) 和最简方案 (MP) 存在交互作用，共同影响形容词–名词语码混合构式，从而为双语句法整合提供了更细致的理解。本章在前几章研究的基础上，将研究重心从语言特有的词汇句法特征转向语言共有的特征，为研究双语者如何协调处理跨语言句法差异提供了理想切入点, 从而深入揭示双语者在语言产出与理解中如何应对跨语言的语法差异。

第七章回顾了核心研究问题，并对前几章的主要研究发现进行了归纳与总结。在此基础上，本章进一步深入探讨了本研究的理论与实证意义，审慎反思其局限性，并提出未来研究的拓展方向。

Curriculum Vitae

Ruixue Wu was born on November 8, 1993, in Shangqiu, located in the northern Chinese province of Henan. She completed her high school education in her hometown before pursuing higher education. In 2017, she earned a Bachelor's degree (BA) in English from Henan Normal University (China). She then obtained a Master's degree (MA) in General and Applied Linguistics, with a specialization in English Language, from the Australian National University (ANU, Australia) in 2019. In March 2021, she was admitted as a PhD candidate at the Leiden University Center for Linguistics (LUCL), Leiden University (The Netherlands). The same year, she was awarded a scholarship by the Chinese Scholarship Council (CSC) to support her doctoral studies. During her PhD, she was supervised by Prof. dr. Niels O. Schiller and Dr. Maria del Carmen Parafita Couto on her PhD project entitled *"The Role of Lexico-Syntactic Features in Noun Phrase Production and Comprehension: Insights from Spanish and Chinese in Unilingual and Bilingual Contexts".* In September 2022 and May 2024, she joined the Speech Acquisition and Perception (SAP) research group at Universitat Pompeu Fabra (UPF, Spain) as a pre-doctoral visiting researcher. Her 2024 visit was supported by the Leiden University Fund (LUF). Her research primarily focuses on the processing of lexico-syntactic features, specifically grammatical gender in Spanish, classifiers in Chinese, and the linear order of adjective and noun in both languages, in language production within monolingual contexts, as well as in code-switched speech within bilingual contexts. Her work draws on behavioral and electrophysiological methods, alongside multi-task approaches, to examine how early Spanish-Chinese bilingual speakers navigate these features across typologically distinct grammatical systems. All her research efforts and findings have been brought together in the present thesis.