The Acquisition of Gender

Crosslinguistic perspectives

EDITED BY
Dalila Ayoun

STUDIES IN BILINGUALISM
63

JOHN BENJAMINS
PUBLISHING COMPANY

The Acquisition of Gender

Studies in Bilingualism (SiBil)

ISSN 0928-1533

The focus of this series is on psycholinguistic and sociolinguistic aspects of bilingualism. This entails topics such as childhood bilingualism, psychological models of bilingual language users, language contact and bilingualism, maintenance and shift of minority languages, and socio-political aspects of bilingualism.

For an overview of all books published in this series, please see *benjamins.com/catalog/sibil*

Editors

Advisory Editorial Board

Volume 63

The Acquisition of Gender. Crosslinguistic perspectives
Edited by Dalila Ayoun

The Acquisition of Gender

Crosslinguistic perspectives

Edited by

Dalila Ayoun
University of Arizona

John Benjamins Publishing Company
Amsterdam / Philadelphia

The paper used in this publication meets the minimum requirements of the American National Standard for Information Sciences – Permanence of Paper for Printed Library Materials, ANSI Z39.48-1984.

DOI 10.1075/sibil.63

Cataloging-in-Publication Data available from Library of Congress:
LCCN 2021041994 (PRINT) / 2021041995 (E-BOOK)

ISBN 978 90 272 1009 8 (HB)
ISBN 978 90 272 5839 7 (E-BOOK)

John Benjamins Publishing Company · https://benjamins.com

Table of contents

About the contributors

Fabrizio Arosio is Associate Professor in Linguistics and Psycholinguistics in the Department of Psychology at the University of Milano-Bicocca. His research focuses on how language conveys temporal information and on how short-term memory resources and executive functions influence the processing of linguistic information. In his studies, he makes use of implicit and explicit behavioral measures with both typically and atypically developing children as well as with adults.

Dalila Ayoun is Professor of French Linguistics and Second Language Acquisition & Teaching in the Department of French & Italian at the University of Arizona in Tucson. Her research interests lie in the second language acquisition of morpho-syntax (tense-aspect-mood/modality, grammatical gender), theoretical and applied French linguistics from a generative/minimalist perspective. Recent publications include two edited volumes (Ayoun, D. 2015. *The Acquisition of the Present*. Amsterdam: John Benjamins; Ayoun, D., Celle, A. & Lansari, L. 2018. *Tense, Aspect, Modality and Evidentiality. Crosslinguistic perspectives*. Amsterdam: John Benjamins) as well as articles published in *Languages*, *Italica* and *Journal of French Language Studies*.

Kate Bellamy is a Marie Curie Postdoctoral Research Fellow at the National Centre for Scientific Research (France). Her research project focuses on gender assignment in mixed nominal constructions among bilingual communities in Mexico and Georgia. In addition to code-switching and grammatical gender, Kate also conducts research into the Purepecha language from language-internal and language contact perspectives. Her recent publications includes a co-edited volume on bilingualism in the Hispanic and Lusophone world (John Benjamins, 2017) as well as articles in journals such as *Languages, Language Typology and Universals* and*Language Learning*.

Ivo Boers is a research assistant at the Heritage Linguistics Lab (HERLING) of the Leiden University Centre of Linguistics (LUCL). His research interests lie in heritage linguistics, more specifically in contact-induced change and code-switching.

Jasmijn Bosch is an early stage researcher in the MultiMind project and a PhD student at the University of Milan-Bicocca. Her research focuses on language processing and literacy development in different groups of bilingual children, as well as multilingualism in educational settings.

Mathilde Chailleux was an early stage researcher in the MultiMind project and a PhD student at the University of Milan-Bicocca. Her project aimed to find markers of developmental dyslexia in bilingual children, with a focus on language processing and rhythmic abilities.

Alejandro Cuza is Professor of Spanish and Linguistics at Purdue University. His primary research area centers on the psycholinguistic nature and dynamics of second language acquisition, heritage speaker bilingualism and child bilingual development. More specifically, he investigates the effects of linguistic and psycholinguistic factors in the acquisition of Spanish morphosyntax and semantics in heritage speakers, bilingual children, and second language learners.

Peter Ecke is Professor of German and Second Language Acquisition and Teaching in the Department of German Studies at the University of Arizona. His research interests include second and third language acquisition, vocabulary acquisition, language attrition, bilingual speech including lexical errors and word finding problems, the effects of study abroad, and the learning and teaching of German in the United States.

Amanda Edmonds is Associate Professor at the Université Paul-Valéry Montpellier 3, where she teaches courses on linguistics, second language acquisition, and language didactics. Her research examines the acquisition of additional languages, with a focus on lexis and morphosyntax. Recent publications include a co-edited special issue on methodological challenges in SLA (*Journal of French Language Studies*) and co-authored articles that have appeared in the journals *Applied Linguistics*, the *Journal of the European Second Language Association*, and *Language Learning*.

Yulia Esaulova received her PhD in Psychology from Heidelberg University (Germany) in 2014. She conducted psycholinguistic studies on gender biases at the University of Duisburg-Essen (Germany), the University of Sussex (UK), the University of Fribourg (Switzerland), and the University of Potsdam (Germany). She now carries out research using eye-tracking and other experimental methods at the University of Cologne (Germany). Her research scope covers native and non-native language comprehension and production in speakers of English, French, German, and Standard Arabic.

Janet Grijzenhout is Professor of English Linguistics at Leiden University. Her research interests are in phonological theory, morphology, first and second language acquisition, language change during the lifespan, historical linguistics and comparative linguistics. Her research groups focus on the morphophonology of West-Germanic languages, the history of Germanic sound systems, multilingualism and bilingual language acquisition. Recent publications include an edited volume (Caon, L., Elenbaas M. & Grijzenhout J. 2021. *Language Use, Usage Guides and Linguistic Norms*, Cambridge: Cambridge Scholars Publishing) and articles published in *Journal of Phonetics* and *Linguistic Approaches to Bilingualism*.

Maria Teresa Guasti is a researcher in language acquisition with about 160 publications. She has researched first and early second language acquisition, language impairments, and dyslexia. Her contributions span over several domains (syntax, semantics and pragmatics) and adopt a strong cross-linguistic perspective. Her research has been supported by several national and international grants (Cost Action, ITN, ERC-Synergy).

Aarnes Gudmestad is Associate Professor at Virginia Polytechnic Institute and State University. She specializes in second language acquisition and sociolinguistics, and her research explores the intersection of the two fields. Her current projects address issues pertaining to morphosyntactic structures (e.g. the subjunctive-indicative contrast, subject expression, future-time reference, grammatical gender) in Spanish and French. She is the co-editor of *Critical Reflections on Data in Second Language Acquisition* (John Benjamins, 2018) and*Interpreting Language-Learning Data* (Language Science Press, 2020). She has published articles in journals such as *Applied Linguistics, Language Learning, Journal of French Language Studies*, and-*Studies in Second Language Acquisition*.

Stefano Maranzana is Lecturer of Italian and French in the Department of World Languages and Literatures at the Southern Methodist University in Dallas. His research interests focus on the acquisition of Italian grammatical gender, captioned video in listening comprehension as well as Virtual Reality. Recent publications include Ayoun, D., Maranzana, S. (2020). Italian Grammatical Gender: A Corpus Study from a Second Language Acquisition Perspective. *Italica, 97*(2), 206–333, and Berti, M., Maranzana, S., Monzingo, J. (2020). Virtual Reality in the Language Classroom: A Look at Students' Attitudes and Beliefs. *International Journal of Computer-Assisted Language Learning and Teaching (IJCALLT)*, *1*, 1–30.

Thomas A. Metzger is Visiting Assistant Professor of Statistics with the Department of Statistics at The Ohio State University in Columbus, Ohio. His research focuses on Bayesian model selection, statistical software, and hierarchical modeling in linguistics.

M. Carmen Parafita Couto is Senior Lecturer at the Leiden University Center for Linguistics. Her main line of research focuses on uncovering the linguistic patterns and strategies used by multilinguals as they switch seamlessly between their languages. Her research has appeared in journals such as *Bilingualism, Language and Cognition, Linguistic Approaches to Bilingualism* or*International Journal of Bilingualis*m. She has co-edited several special issues and thematic volumes and is also editor in chief of the open-access book series *Current Issues in Bilingualism* (Language Science Press) and associate editor of *Isogloss. Open Journal of Romance Linguistics*.

Liliana Sánchez is a Professor of Spanish Linguistics in the department of Hispanic and Italian Studies at the University of Illinois Chicago. She has published *Bilingualism in the Spanish-speaking world* with Jennifer Austin & Maria Blume (Cambridge University Press, 2015), *The Morphology and Syntax of Topic and Focus: Minimalist Inquiries in the Quechua Periphery* (John Benjamins, 2010), and *Quechua-Spanish Bilingualism. Interference and Convergence in Functional Categories* (John Benjamins, 2003) as well as articles in journals such as *Bilingualism: Language and Cognition*, *Glossa, International Journal of Bilingualism*, *Linguistic Approaches to Bilingualism*, *Lingua*, *Probus*, and *Studies in Second Language Acquisition*.

LeAnne Spino is an Assistant Professor of Spanish and the Proficiency Coordinator for the Department of Modern and Classical Languages and Literatures at the University of Rhode Island. She specializes in second language acquisition and her research centers primarily around the development of language proficiency and language processing. She has published articles in journals such as *The Language Educator*, *Language Learning* and *TESOL Quarterly*.

Bo Sterken is an MA student Linguistics and Latin American studies, as well as a research trainee at Leiden University. Her research interests lie in bilingual and second language acquisition (especially of grammatical gender), heritage linguistics and code-switching.

Deniz Tat is Assistant Professor of Turkish at the Leiden Institute for Area Studies. She is also a Marie Curie research fellow (2019–2022) at UiT The Arctic University of Norway for her project PATH "Preservation and Adaptation in Turkish as a Heritage Language," examining a series of morphophonological, lexical and syntactic features of the Northwestern dialect of Anatolian Turkish as it has been spoken by a community of heritage speakers in a small Dutch town and comparing these directly to changes happening in the dialect's original context in Turkey.

Brechje van Osch is a Postdoctoral researcher in the AcqVA Aurora Center at the Department of Language and Culture at UiT. Her research focuses on bilingualism, with a particular interest in heritage language acquisition. Her current post doc project looks at cross-linguistic influence effects by comparing bilingual populations of several different language combinations, using a variety of behavioral and online methods. Recent publications include articles in *Languages* and *Heritage Language Journal*.

Lisa von Stockhausen is Professor of Language and Cognition, Department of Psychology, University of Duisburg-Essen. Her psycholinguistic research focusses on cognitive and linguistic representations of gender, and how language contributes to shaping, reflecting and maintaining social reality. Another research interest are cognitive mechanisms underlying effects of mindfulness (cf. von Stockhausen & Funke (2019). The nature of language. In *The Psychology of Human Thought*, Sternberg & Funke (eds), pp. 199–212. Heidelberg: Heidelberg University Publishing; Wimmer et al. (2020). Mindfulness training for improving attention regulation in university students: is it effective? and Do yoga and homework matter? *Frontiers in Psychology*, 11, p. 719.

Jia'en Yee is an early stage researcher in the MultiMind project and a PhD student at Universiti Putra Malaysia. Her project focuses on the role of multilingualism and biliteracy on brain structure, function and cognition.

CHAPTER 1

Introduction

Gender, *geslag, Geschlecht, γένος*, लिंग, 性别

Dalila Ayoun
University of Arizona

What is gender?

Gender as a morphosyntactic feature is arguably "an endlessly fascinating linguistic category" (Corbett, 2014: 1). One may even say it is among "the most puzzling of the grammatical categories" (Corbett, 1991: 1) that has raised probing questions from various perspectives even if the recent literature on its complexity has muddied the waters with mixed characterizations such as "historical junk" (Trudgill, 2011: 156) or a perfect example of a mature phenomenon (Dahl, 2004; cited in Di Garbo, Olsson, & Wälchli, 2019).

The word *gender* itself has been used as a grammatical term since the 14th century to refer to noun classes labelled as feminine, masculine or neuter as in most Indo-European languages such as French or German. It is a common category in the world languages which display an astounding variety. Most languages display semantic and/or formal gender systems with various degrees of opacity and complexity. Closely related languages such as Scandinavian languages use different gender systems: Norwegian distinguishes between the masculine, feminine and neuter, while in Swedish and Danish, the masculine and feminine have merged into a common lexical gender (Comrie, 1999). However, dialectal varieties may differ with regard to gender. Thus, out of Norwegian's two written standards, Nynorsk and Bokmål, all feminine nouns may take masculine agreement in Bokmål, which means that this written variety uses only common and neuter, while the three-gender system remains in almost all dialects of spoken Norwegian (Lohndal & Westergaard, 2016; Venås, 1993). The collapse of the feminine and masculine genders into common gender is a well known fact in Germanic languages and dialects such as Dutch and Danish (Jahr, 1998; Nesse, 2002).

Nigerian Fula, a Senegambian language spoken in Central and Western Africa by over 65 million people, displays around twenty genders, depending on the dialect

https://doi.org/10.1075/sibil.63.01ayo

(e.g. Arnott, 1967; Breedveld, 1995), while others have none such as Hungarian (Hetzron, 2001), Hawaiian (Elbert & Pukui, 1979) or Indonesian (Sneddon, 1996).

Several of the languages represented in the present volume do not grammaticalize gender like Turkish and Mandarin Chinese (Corbett, 1991). Papiamento is an Iberian-lexifier creole language spoken on the ABC islands (i.e. Aruba, Bonaire, Curaçao) and in the Netherlands through its diaspora, without grammatical gender assignment or agreement, and its invariable adjectives typically end in *-o* or *-u* (Kouwenberg, 2007). Modern English only displays semantic gender, having lost its three-gender system (masculine, feminine, neuter) from Old to Middle English (Curzan, 2003).

The remaining languages – Italian, Spanish, German, French, Brazilian Portuguese – all exhibit formal gender systems with both lexical gender and grammatical gender. The former is a semantically motivated gender based on the sex of the animate referent, while the latter is said to be arbitrary because it lacks semantic motivation, or as Deutsch and Dank (2009: 116) put it, "grammatical gender for inanimate nouns has no conceptual basis, and does not go through a process of feature marking in mapping the conceptual notion onto a linguistic code".

German is a three-gender system language, while Romance languages have binary systems. German gender assignment is opaque with a complex interplay of phonological and morphological rules with varying degrees of reliability and numerous exceptions (Corbett, 1991; Zubin & Köpcke, 1984). The morphological rules based on affixes are more reliable (Hager, 2014). French can be described as having a formal gender system with a weak semantic component because only 10.5% of nouns have a semantic gender (Séguin, 1969). Tucker, Lambert, and Rigault (1977) contend that a comprehensive and complex system of morpho-phonological and semantic rules can account for gender assignment, but it is fraught with exceptions, highly variable predictive values and multiple orthographies (Ayoun, 2010).

Gender marking is more transparent in Spanish since 99.87% of nouns ending in *-o* are masculine, while 96.3% of nouns ending in *-a* are feminine. Phonological gender assignment rules for other vowels and consonants are highly reliable (Teschner & Russell, 1984), although they fail to account for epicene nouns or grammatical homonyms (Harris, 1991).

Italian, another Romance language, is quite different from Spanish. Its vocalic endings are portmanteau morphemes conflating gender and number: about 70% of inanimate nouns follow the predictable, transparent pattern of *-o*/*-i* for single/plural masculine and *-a*/*-e* for singular/plural feminine nouns (Thornton, Iacobini, & Burani, 1997). The nouns that fall outside this pattern are said to be phonologically opaque and are considered as exceptions. Moreover, only five vowels express the combinations of two genders (feminine, masculine) and number (singular, plural) and none is clearly transparent (Ayoun & Maranzana, 2020; this volume).

Portuguese's semantic and grammatical binary gender system appears to display a transparent pattern of feminine words ending in *-a* and masculine words ending in *-o*, but words ending in stressed *-á* and *-ema/-oma* are usually masculine, while words ending in *-ção, -são* are feminine, among several other endings, almost all with exceptions. Compound nouns in Portuguese and French are typically masculine even when both nouns are feminine as in *guarda-roupa* 'wardrobe' or *tête à tête* 'face to face', and both languages have grammatical homonyms as well such as *livre* 'book-MASC, pound-FEM' or *aide* 'aid-MASC, help-FEM' in French.

There are thus distinct differences even between languages from the same family such as Romance languages, although Foundalis (2002) that selected 84 common nouns in 14 Indo-European languages found that the closer languages are in the family tree, the more they tend to agree on gender. Crosslinguistic differences are bound to create difficulties for second language learners who may be confronted with a completely different gender system than in their first language that may not even exhibit one to begin with. And yet, its acquisition will be crucial to achieve a target-like second language grammar.

Chapters in this volume

The first three chapters present critical reviews in three different areas – gender assignment in mixed noun phrases, subtle gentle biases and gender acquisition in child and adult heritage speakers of Spanish – while the next six chapters present new empirical evidence in the acquisition of gender by bilingual children, adult L2/L3 learners and heritage speakers of various languages.

Bellamy and **Parafita Couto** (Chapter 2) present a state-of-the-art review of the research on grammatical gender assignment in mixed noun phrases, one of the most frequent constructions for code-switching by bilingual speakers. Mixed noun phrases are composed of a noun in one language and a gender-marked determiner and/or adjective from another language as in *los pants rojos* 'the red pants'. Three main gender assignment strategies are described: (a) using a translation equivalent also referred to as synonymic gender, gender copy or analogical gender among other terms (Montes-Alcalá & Lapidus Shin, 2011); (b) a shape-based strategy also known as phonological analogy, but the code-switched pair can be produced in oral or written mode; and (c) a default strategy that appears to be the most common and consists of using a single gender for most code-switched nouns regardless of their morpho-phonological and semantic properties, or the gender of their translation equivalent. Bellamy and Parafita Couto systematically review the linguistic and non-linguistic factors that contribute to the choice of one (or more) gender assignment strategy citing examples from various language pairs (e.g. Spanish-English,

the most common one, but also Russian-English, Spanish-German, Tsova-Tush-English). Future research should include different language pairs particularly from languages exhibiting different gender systems (i.e. non-binary, ternary, etc), as well as various empirical tools to collect naturalistic data in communities where code-switching is a common occurrence.

In the third chapter reviewing psychological and psycholinguistic studies, **Esaulova** and **von Stockhausen** uncover subtle gender biases and stereotypes embedded in a wide range of linguistic phenomena from gender asymmetries to case systems and reflect upon whether they shape cognitive representations. Among the linguistic phenomena reviewed are linguistic intergroup and expectancy bias, negation bias, irony bias, male bias in generic masculine and gender-fair alternatives and gender mismatch effects. Data are typically elicited with questionnaires, language perception experiments or descriptions of visual representations of specific positive and/or negative situations involving different groups. For instance, irony may be used to express whether information is (in)consistent with expectations given a specific situation involving stereotypical actors or not. Another example may be how generic masculine forms influence people's perception and behavior such as job-seekers who would not consider applying for positions that were advertised using stereotypical phrases, although they were qualified (Bem & Bem, 1973). Or *girls are as good as boys at math* shows that even an intended expression of gender equality expresses a subtle bias (Chesnut & Markman, 2018). Esaulova and von Stockhausen point out that however pervasive gender biases may be, uncovering them is no easy task requiring different empirical approaches depending on the nature of the bias. Future studies should focus on a wider variety of languages and more specifically on which factors may determine the strength of a bias in gender-related contexts. Their findings may have important implications for neutralizing their negative social consequences.

Cuza and **Sánchez** (Chapter 4) set out to review empirical studies on the acquisition of grammatical gender in child and adult heritage speakers of Spanish in contact with English as the majority language who tend to overgeneralize the masculine to feminine nouns (e.g. **un*-MASC *llave*-FEM *chiquita*-FEM 'a small key'), only the adjective (e.g. *una*-FEM *nube*-FEM **rojo*-MASC 'a red cloud'), or both the determiner and the adjective (e.g. **un*-MASC *nube*-FEM *rojo*-MASC 'a red cloud')(Cuza & Pérez-Tattam, 2015). The authors also consider to which extent crosslinguistic influence from English as the majority language, the age of onset of bilingualism and patterns of language activation and use can account for morphosyntactic changes in heritage language grammars such as gender assignment. Are the differences between heritage speaker grammars and monolingual speaker grammars due to incomplete language acquisition during childhood, language attrition later on in life or grammar restructuring in general (reconfiguration of gender features in

particular)? Cuza and Sánchez contend that the Bilingual Alignment Approach (Sánchez, 2019) is well suited to account for the characteristics observed in heritage speaker grammars. It suggests that lexical frequency, linguistic patterns and proficiency lead to a reorganization of morphosyntactic features from both languages creating new configurations or new bilingual alignments that may become permanent. Future research from other heritage speaker grammars may offer additional supporting evidence.

In **Ayoun** and **Maranzana** (Chapter 5), a cross-sectional study investigates the ability of Anglophone learners of Italian (n = 87) to assign the appropriate grammatical gender to individual nouns in a written, computerized gender assignment task. The stimuli were a subset of isolated nouns drawn from a written corpus of current magazine and newspaper articles, thus exemplifying the type of input instructed learners are likely to receive (Ayoun & Maranzana, 2020). The fact that Italian differs from other Romance in that its vocalic endings are portmanteau morphemes conflating gender and number features complicates the task of L2 learners particularly when their L1 lacks grammatical gender as it is the case for English. After a descriptive account detailing how grammatical gender is morphologically expressed, Ayoun and Maranzana adopt a Minimalist perspective and the view that features and functional categories are in principle acquirable by adult learners because L2 acquisition is constrained and guided by Universal Grammar. According to the Feature Reassembly Hypothesis, L2 learners first need to detect how the features of gender and number are expressed, and then reassemble them (Lardiere, 2008, 2009). Results show L2 learners performed poorly while NS controls performed as expected, with strong effects for transparency, gender and number, but not for suffixed nouns. Proficiency was not a good predictor of performance suggesting that these participants were not yet able to handle the complexity and ambiguity of Italian vocalic endings. Future research with more advanced learners would answer the question of whether L2 learners' performance may improve at higher proficiency levels.

Chapter 6 by **Ecke** focuses on the oral production of gender marked determiners by L3 German learners (n = 123)(L1 Spanish, advanced L2 English). In addition to an opaque three-way gender system, German displays a complex case (nominative, genitive, dative, accusative) system that applies to nouns and determiners. For instance, the indefinite article *ein* is both masculine and neuter singular for the nominative, but only neuter singular for the accusative. Determiners are acquired late and constitute poor cues to the grammatical gender of nouns while phonological and morphological cues in the form of nominal endings vary in reliability and have exceptions (Köpcke & Zubin, 1983, 1984). The results show that 70.6% of noun phrases were accurately produced with a gender-marked determiner when appropriate, with a better accuracy for masculine and neuter than feminine as well

as for indefinite versus definite articles. The Parasitic Model of L2 and L3 vocabulary acquisition (e.g. Ecke, 2015) that assumes that cross-linguistic influence (CLI) plays a role in the form of already established lexical connections, was used to interpret the results. Similarly to the Spanish-English bilingual speakers in Chapter 2, these beginning L3 German learners appear to adopt the gender of a lexical equivalent and to be subjected to frequency effects rather than make use of formal gender cues.

Edmonds, **Gudmestad** and **Metzger** (Chapter 7) adopt a form-based approach within a variationist framework to explore the linguistic and extralinguistic factors that may determine the use of feminine as opposed to masculine determiners and adjectives regardless of accuracy. A variationist framework propounds that (inter) languages exhibit systematic variation defined as the use of two (or more) forms to express the same function or meaning in a given context. The oral production data – semi-guided interviews and picture narrations – were obtained from Anglophone young adult learners of L2 French (n = 20) before, during, and after an academic year spent in France (LANGSNAP corpus; Mitchell, Tracy-Ventura, & McManus, 2017). The transcribed data were coded following 14 independent variables such as syllable distance, noun gender and noun-final phoneme. The authors analyzed the data with the objective of exploring variable gender marking with feminine as the marked form and masculine as the default form. Six factors were found to be significant and are compared to a previous accuracy-based analysis (Edmonds et al., 2020) : "noun gender, noun-initial phoneme, noun frequency, modifier type, syllable distance, and noun class". The authors conclude that a form-based approach and an accuracy-based approached are complementary rather than exclusive and may inform the acquisition of gender in novel ways.

In Chapter 8, **Spino** explores how L2 Spanish learners (n = 25) and Spanish NSs (n = 27) process sentences embedded with grammatical gender agreement violations (i.e. determiner-noun and noun-adjective) with an eye-tracking experiment during which they read 148 experimental sentences and answered a comprehension question after each one. They then completed a post-reading questionnaire designed to assess their awareness of the gender agreement violations, a vocabulary posttest, a background questionnaire and a proficiency test. The results of the eye-tracking experiment indicate that both L2 learners and Spanish NSs were sensitive to determiner-noun agreement violations, but only the Spanish NSs were also sensitive to noun-adjective agreement violations as measured by reading times for grammatical vs ungrammatical stimuli. The answers to the post-reading questionnaire revealed important differences between the L2 learners and the NSs. Although both groups noticed errors, 93% of NS reported noticing determiner-noun violations and noun-adjective violations versus 64% and 52%, respectively, for the L2 learners. The L2 learners' failure to notice grammatical violations is interpreted as a lack of explicit awareness although such violations are explicitly taught in foreign

language classrooms. Other empirical options for researchers interested in the activation of explicit knowledge are discussed.

Chapter 9 by **van Osch**, **Boers**, **Grijzenjout**, **Parafita Couto**, **Sterken** and **Tat** sets out to explore the role of cross-linguistic influence (CLI) from three different heritage languages – Turkish, Papiamento, and Spanish – onto Dutch as the majority language in the Netherlands. Dutch and Spanish exhibit grammatical gender in binary systems (common and neuter for Dutch, masculine and feminine for Spanish), but Turkish and Papiamento do not. The authors are interested in the linguistic and extralinguistic variables that may play a role in gender assignment accuracy as well as the strategies (pre)adolescent and adult heritage speakers (n = 58) of these three different languages may adopt. A Director-Matcher task was used to collect oral data consisting of nominal phrases (determiner, noun adjective) in unilingual modes (Dutch and a heritage language) and in two code-switched modes (Dutch-heritage language and vice-versa). The results indicate that all heritage speakers had a tendency to overgeneralize the common gender in the Dutch unilingual mode, but the Spanish speakers' performance was more accurate that the performance of the Papiamento and Turkish speakers. Most participants used common gender in code-switching mode as well, but some used a translation equivalent of the noun in Dutch or produced an uninflected adjective. The age of onset, amount of input and language use turned out to be significant factors.

Bosch, **Chailleux**, **Foppolo**, **Guasti** and **Arosio** (Chapter 10) examine the online processing of grammatical gender and number in Mandarin-Italian bilingual children (n = 25) who are compared to monolingual Italian children (n = 32). Mandarin differs from Italian in several ways: it lacks determiners, gender and number (bare nouns are unspecified being neither singular nor plural, although numerals and quantifiers serve as nominal classifiers and express number) as well as morphosyntactic agreement. This original language pair provides an interesting testing ground for how early child learners may process a morphologically rich language like Italian when their L1 is so different. This study investigates how Mandarin-Italian bilingual children may use grammatical gender and number marking on articles to anticipate upcoming nouns as measured by a visual world eye-tracking paradigm. Participants heard sentences in Italian composed of an introduction ('now find'), a definite article and a target noun; they saw two images and had to select the appropriate one. There were 3 experimental conditions: early gender (predictable), early number (predictable), late (unpredictable). The results show that monolingual children used articles to anticipate nouns with nearly identical patterns for gender and number, while bilingual children were considerably slower in the gender condition than in the number condition. L2 proficiency (particularly in lexical knowledge), which had a significant effect on gender processing in the bilingual group, and transfer are offered as explanations to these differential results.

References

Arnott, D. W. (1967). Some reflections on the content of individual classes in Fula and Tiv. In M. André (Ed.), *La classification nominale dans les langues négro-africaines* (pp. 45–74). Éditions du Centre national de la recherche scientifique.

Ayoun, D. (2010). Corpus data. Shedding the light on French grammatical gender…or not. *Eurosla Yearbook*, 10, 119–141. https://doi.org/10.1075/eurosla.10.08ayo

Ayoun, D., & Maranzana, S. (2020). Italian grammatical gender: A corpus study from a second language perspective. *Italica*, 97(2), 306–333.

Bem, S. L., & Bem, D. J. (1973). Does sex-biased job advertising aid and abet sex discrimination. *Journal of Applied Social Psychology*, 3(1), 6–18. https://doi.org/10.1111/j.1559-1816.1973.tb01290.x

Breedveld, J. O. (1995). *Form and meaning in Fulfulde: A morphophonological study of Maasinankoore*. Research School CNWS.

Chestnut, E. K., & Markman, E. M. (2018). "Girls are as good as boys at math" implies that boys are probably better: A study of expressions of gender equality. *Cognitive Science*, 42, 2229–2249. https://doi.org/10.1111/cogs.12637

Comrie, B. (1999). Grammatical gender systems: A linguist's assessment. *Journal of Psycholinguistic Research*, 28, 457–466. https://doi.org/10.1023/A:1023212225540

Corbett, G. G. (1991). *Gender*. Cambridge University Press. https://doi.org/10.1017/CBO9781139166119

Corbett, G. G. (Ed.). (2014). *The expression of gender*. De Gruyter.

Curzan, A. (2003). *Gender shifts in the history of English*. Cambridge University Press. https://doi.org/10.1017/CBO9780511486913

Cuza, A., & Perez-Tattam, R. (2015). Grammatical gender selection and phrasal word order in child heritage Spanish: A feature re-assembly approach. *Bilingualism: Language and Cognition*, 19(3), 332–350.

Dahl, Ö. (2004). *The growth and maintenance of linguistic complexity*. John Benjamins. https://doi.org/10.1075/slcs.71

Deutsch, A., & Dank, M. (2009). Conflicting cues and competition between notional and grammatical factors in producing number and gender agreement: Evidence from Hebrew. *Journal of Memory and Language*, 60, 112–143. https://doi.org/10.1016/j.jml.2008.07.001

Di Garbo, F., Olsson, B., & Wälchli, B. (Eds). (2019). *Grammatical gender and linguistic complexity, Vol. I: General issues and specific studies*. Language Science Press.

Ecke, P. (2015). Parasitic vocabulary acquisition, cross-linguistic influence, and lexical retrieval in multilinguals. *Bilingualism: Language and Cognition*, 18, 145–162. https://doi.org/10.1017/S1366728913000722

Edmonds, A., Gudmestad, A., & Metzger, T. (2020). A longitudinal study of grammatical gender marking in French as an additional language. *Applied Linguistics*, 41(5), 733–755. https://doi.org/10.1093/applin/amz036

Elbert, S. H., & Pukui, M. K. (1979). *Hawaiian Grammar*. University of Hawai'i Press. https://doi.org/10.1515/9780824840792

Foundalis, H. E. (2002). Evolution of gender in Indo-European languages. *Proceedings of the Annual Meeting of the Cognitive Science Society*, 24.

Hager, M. (2014). *Der Genuserwerb bei mehrsprachig aufwachsenden Kindern*. Bergische Universität Wuppertal.

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

Hetzron, R. (2001). Hungarian. In J. Garry & C. Rubino (Eds.), *Facts about the world's languages. An encyclopedia of the world's languages: Past and present* (pp. 314–316). H. W. Wilson.

Jahr, E. H. (1998). Sociolinguistics in historical language contact: The Scandinavian languages and Low German during the Hanseatic period. In E. H. Jahr (Ed.), *Language change: Advances in historical sociolinguistics* (pp. 119–130). Mouton de Gruyter.

Köpcke, K.-M., & Zubin, D. A. (1983). Die kognitive Organisation der Genuszuweisung zu den einsilbigen Nomen der deutschen Gegenwartssprache. *Zeitschrift für Germanistische Linguistik*, 11(2), 166–182. https://doi.org/10.1515/zfgl.1983.11.2.166

Köpcke, K.-M., & Zubin, D. (1984). Sechs Prinzipien für die Genuszuweisung im Deutschen: Ein Beitrag zur natürlichen Klassifikation. *Linguistische Berichte*, 93, 26–50.

Kouwenberg, S. (2007). Papiamento (Creole Spanish/Portuguese). In J. Holm & P. Patrick (Eds.), *Comparative creole syntax* (pp. 307–332). Battlebridge.

Lardiere, D. (2008). Feature-assembly in second language acquisition. In J. Liceras, H. Zobl, & H. Goodluck (Eds.), *The role of formal features in second language acquisition* (pp. 106–140). Lawrence Erlbaum Associates.

Lardiere, D. (2009). Some thoughts on the contrastive analysis of features in second language acquisition. *Second Language Research*, 25(2), 173–227. https://doi.org/10.1177/0267658308100283

Lohndal, T., & Westergaard, M. (2016). Grammatical gender in American Norwegian heritage language: Stability or attrition? *Frontiers in Psychology*, 7, 1–15. https://doi.org/10.3389/fpsyg.2016.00344

Mitchell, R., Tracy-Ventura, N., & McManus, K. (2017). *Anglophone students abroad: Identity, social relationships, and language learning*. Routledge. https://doi.org/10.4324/9781315194851

Nesse, A. (2002). *Språkkontakt Mellom Norsk og Tysk I Hansatidens Bergen* [Language Contact between Norwegian and German in Bergen during the Hanseatic Time]. Novus.

Sánchez, L. (2019). Bilingual alignments. *Languages*, 4(82), 1–24. https://doi.org/10.3390/languages4040082

Séguin, H. (1969). Les marques du genre dans le lexique français écrit contemporain: Compilation des cas et essai de classement (Unpublished PhD dissertation). Université de Montréal.

Sneddon, James N. (1996). *Indonesian: A comprehensive grammar*. Routledge.

Teschner, R., & Russell, W. M. (1984). The gender patterns of Spanish nouns: An inverse dictionary-based analysis. *Hispanic Linguistics*, 1, 115–132.

Thornton, A. M., Iacobini, C., & Burani, C. (1997). *BDVDB. Una base di dati per il vocabolario di base della lingua italiana* [BDVDB. A database for the basic vocabulary of Italian], (2nd ed.). Bulzoni.

Trudgill, P. (2011). *Sociolinguistic typology: Social determinants of linguistic complexity*. Oxford University Press.

Tucker, G., Lambert, W., & Rigault, A. (1977). *The French speaker's skill with grammatical gender: An example of rule-governed behavior*. Mouton. https://doi.org/10.1515/9783110805413

Venås, K. (1993). On the choice between two written standards in Norway. In E. H. Jahr (Ed.), *Language conflict and language planning* (pp. 263–278). Mouton de Gruyter. https://doi.org/10.1515/9783110886580.263

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

PART I

Review chapters

CHAPTER 2

Gender assignment in mixed noun phrases

State of the art

Kate Bellamy and M. Carmen Parafita Couto
Langues et civilisations à tradition orale – Centre national de la recherche scientifique / Leiden University Centre for Linguistics

Noun phrases (NPs) constitute one of the most frequent sites where bilingual speakers code-switch. In this chapter we aim to provide a comprehensive overview of research into grammatical gender assignment in mixed NPs, namely NPs containing a noun from one language and a gender-agreeing element from another. We outline the three main gender assignment strategies observed in mixed NPs: (i) translation equivalent, (ii) shape-based, and (iii) default, and in which language pairs, using which tasks, they have been observed. We discuss how the order of acquisition of the gendered and non-gendered language, language dominance, task type, and community norms combine with linguistic properties to modulate gender assignment patterns. Findings suggest that bilinguals who learned a gendered language first seem to prefer the translation equivalent strategy. Since insufficient data from a wide variety of language dyads is currently available, we are cautious about offering further generalisations. Nonetheless, we highlight findings suggesting that more habitual code-switchers prefer the masculine default strategy. In order to progress beyond the current state of the art, we suggest that the field needs to expand into more language dyads, as well as extend existing research on individual language dyads, using multiple methodologies and in communities differing in their code-switching frequency.

Keywords: code-switching, mixed noun phrases, grammatical gender, gender assignment, bilingualism

1. Introduction

In recent years, the field of linguistics has seen a surge in interest concerning how bilinguals process (e.g. Valdés Kroff, 2012; Pablos et al., 2019; Vaughan-Evans et al., 2020), produce (e.g. Parafita Couto, Deuchar, & Fusser, 2015; Balam, 2016; Parafita Couto et al., 2015; Valdés Kroff, 2016; Blokzijl, Deuchar, & Parafita Couto,

https://doi.org/10.1075/sibil.63.02bel

2017; Parafita Couto & Gullberg, 2019) and evaluate (e.g. Anderson & Toribio, 2017; Parafita Couto & Stadthagen-González, 2017) mixed nouns phrases. Much of this research has focused on how grammatical gender is assigned in mixed noun phrases, one of the most frequent switch sites in the bilingual speech of some communities (Eichler, Hagen, & Müller, 2012, p. 237; cf. Deuchar & Quay, 1999; Ezeizabarrena, 2009).

We define grammatical gender in terms of agreement, namely as "classes of nouns reflected in the behaviour of associated words" (Corbett, 1991, p. 1, following Hockett, 1958, p. 231). When we refer to the gender of a noun, we are using this as shorthand for the gender assigned to its agreeing elements, such as determiners or adjectives. Equally, when we speak of a gender system, or how many gender classes a particular language possesses, we are referring to these agreement classes.

Mixed noun phrases contain minimally two elements, a gendered element from one language and a noun from another language, which may or may not possess a gender system. Gender can be marked on the determiner, as in (1a), or on a different modifier element in the absence of a determiner, such as an adjective (1b), or on both, depending on the grammatical categories present in the languages in question. Gender may occur in only one of the languages in the dyad, as in (1a–b), or in both (1c). There are also cases where the noun comes from the gendered language and the determiner from the non-gendered language, e.g. Spanish-English the *casa* 'the house', but we will not consider these here (for discussion see Blokzijl et al., 2017; Parafita Couto & Gullberg, 2019; see also Section 2).

(1) a. *la* house 'the.FEM house' (*Spanish*-English; López, 2020, p. 100)
b. *d-aqqo^{n}* mdinar 'big.$_{D}$ river' (*Tsova-Tush*–Georgian; Bellamy & Wichers Schreur, in press)
c. *el* gürtel 'the.MASC belt.MASC' (*Spanish*-German; López, 2020, p. 108)

Mixed noun phrases such as those in (1a–c) constitute intraclausal (as opposed to interclausal) switches, which are encountered in the code-switched speech of many, but not all, bilingual communities (Deuchar, 2020; cf. Muysken, 2000, who refers to intra-clausal switching as code-mixing and inter-clausal as code-switching). We define code-switching here following Deuchar (2012, p. 1) as "an activity which may be observed in the speech (or writing) of bilinguals who go back and forth between their two languages in the same conversation".

The precise points within the clause at which speakers switch, as well as which language contributes the different elements (i.e. function vs. content words) are constrained by various linguistic and extralinguistic factors (e.g. Myers-Scotton & Jake, 2000; Parafita Couto et al., 2014; Beatty-Martínez, Valdés Kroff, & Dussias, 2018). Code-switching researchers are especially interested in switch points that constitute 'conflict sites' between the two languages, namely features which contrast

in the two languages, such as the presence or absence of gender, or the order of noun and adjective (see Poplack & Meechan, 1998).

Mixed noun phrases can therefore be considered a rich vein from which patterns of bilingual speech can be tapped. The key question with specific reference to gender in mixed NPs is: how do speakers assign a particular gender category to the other-language noun? Further to this, why do speakers opt for a particular gender assignment strategy over another? Which factors modulate this choice? (e.g. Munarriz-Ibarrola et al., 2021).

There are numerous studies, stretching back over a century, focusing on how borrowings are integrated into a gendered language, often in relation to immigrant communities (e.g. Flom, 1903; Haugen, 1969; Reed, 1942; Weinreich, 1953). Borrowing and code-switching are often considered to be related phenomena, with the former constituting the outcome of the latter (e.g. Backus, 2015; Gardner-Chloros, 2009; Myers-Scotton, 2002). However, differentiating between the two remains a divisive issue in bilingualism research, especially in relation to single word insertions, the focus of this chapter (see Deuchar, 2020).

We consider as code-switches those lexical items that do not appear in the monolingual dictionary of the language into which they have been inserted or, in the absence of an authoritative dictionary, do not form part of the linguistic repertoire of monolinguals (where they exist) in the same region (compare also Muysken's (2000) notion of "listedness"). Some would classify these items as 'nonce words' or 'nonce borrowings', depending on how they pattern with respect to the recipient language and, by extension, established loanwords (Poplack, 2018, p. 7). The authors of the studies presented in this overview define their single-word insertions as code-switches, therefore we will not delve into the thorny borrowing vs. code-switching debate any further.

The ever-expanding body of studies on gender assignment produces a wide variety of data, using multiple methods and analyses, from various language pairs in different regions and countries, leading to a situation where we seemingly know a lot about specific communities, but much less about the overarching picture. The findings reported here stem from four main data collection methods: (i) naturalistic production, stored as corpora, where participants are recorded speaking freely, often in the absence of the researcher; (ii) guided production, either alone or in pairs, where participants react to stimuli provided by the researcher, such as the Frog Story or 'map task' (see also Gullberg, Indefrey, & Muysken, 2009); (iii) judgement tasks, where speakers are asked to provide evaluations in either written or spoken form to mixed language material; and (iv) online psycholinguistic methods, such as Event Related Potentials, which tap into the processing of mixed NPs.

In this chapter we aim to provide a comprehensive, up-to-date overview of our knowledge to date, taking stock of the various language dyads studied and

methodologies adopted.[1] Its structure is as follows: In Section 2 we present the evidence supporting the three principal gender assignment strategies encountered in the literature, in both adult and child language. The same study may appear in various sections since it is rare for a speaker community to use only one gender assignment strategy. Section 3 presents the extralinguistic factors that play a role in gender assignment, namely type of bilingual, type of task, and community norms. A discussion of how the aforementioned linguistic and extralinguistic factors modulate gender assignment preferences, as well as the lacunae in the body of knowledge is presented in Section 4, while Section 5 concludes with recommendations for future research.

2. Linguistic factors

For languages possessing a grammatical gender system, the assignment of gender to inserted other-language nouns[2] is obligatory, irrespective of whether or not the donor language also has gender (Stolz, 2008, p. 400). Corbett (1991, p. 81) claims that gender is assigned to borrowings "in essentially the same way as are native words", that is, according to semantic, or a combination of semantic and formal (phonological and/or morphological) principles (Pensalfini & Meakins, 2019; Poplack, Pousada, & Sankoff, 1982; Violin-Wigent, 2006).

Early typologies of gender assignment strategies focused on how borrowings are integrated into a recipient language, but importantly acknowledged that there was no single way to do this, even within one language. In relation to English nouns in Pennsylvania German, Reed (1942, pp. 25–26) identified four gender assignment strategies, which form the basis for subsequent typologies (notably Arndt, 1970; Ibrahim, 1973; Stolz, 2008).

First are nouns that have taken over the gender of the German nouns they displace (their *translation equivalent* in the terms adopted in Section 2.1), akin to Weinreich's (1953) "semantic association with a displaced native equivalent" (cited in Arndt, 1970, pp. 244–245), and Stolz's (2009) "gender copy" for dyads

1. We cannot claim complete exhaustivity of the literature on the topic but, having searched in English, French, German and Spanish, we consider this review to cover a significant proportion of studies published to date.

2. Deuchar (2020) refers to single-word code-switches as 'lone other-language items' (LOLIs), Aaron (2015) labels them 'singletons' while Poplack (2018) would call them 'nonce words' or 'nonce borrowings' depending on their status. We do not opt for a single term, following the conventions used by different authors, but often refer to them as 'other-language nouns' or 'insertions'.

comprising two gendered languages (see Section 2.1.1). Second are nouns ending in a type of suffix that normally characterizes a particular gender in German ("morphological analogy" as per Weinreich (1953), and could likely be subsumed under our *shape-based strategy* in Section 2.2). Third are nouns whose gender is determined by the 'natural' sex of an animate (often restricted to humans),[3] a strategy claimed to supersede all others in gender assignment (Montes-Alcalá & Lapidus Shin, 2011, p. 120; but see Balam, 2016; Valdés Kroff, 2016 for counter-evidence). Fourth are nouns that have been given the feminine gender, because the English definite article [ði:/ðə] phonetically resembles the German feminine definite article *die* [di:]. Weinreich (1953) replaces this final type with a category based on the relative productiveness of the genders in the recipient language (most closely associated with a *default strategy*, see Section 2.3).

In relation to German, Arndt (1970, p. 245) adds three additional influences: graphic analogy (which can reasonably be associated with our *shape-based strategy*), homonymity (also referring to a *translation equivalent*), and assignment by semantic categories (related again to the *translation equivalent*, see also Montés Alcalá & Lapidus Shin, 2011 for a discussion of gender assignment and its relation to hyperonymy). These early typologies are infrequently cited in contemporary literature, but in them we can recognise the three main gender assignment strategies that have been identified in the literature on code-switching in mixed NPs and which are presented in detail in this chapter.

In a mixed noun phrase, the determiner and the noun belong to different languages. It has been claimed that bilingual speakers prefer mixed noun phrases where the determiner comes from their gendered language, such as the Spanish-English ***la*** *table*, to those where the determiner is genderless, such as the English-Spanish *the* ***mesa*** (e.g. López, 2020, pp. 88–89; Liceras el al., 2008 for production only, but see Blokzijl et al., 2017; Parafita Couto & Stadthagen-González, 2017; Parafita Couto & Gullberg, 2019 for counter-evidence).

Formal accounts supporting this prediction appeal to feature checking and the interpretability of the gender feature in each language. Liceras et al. (2008, p. 828) cite the "intrinsic Gender feature of the Spanish Noun and the intrinsic Gender Agreement feature of the Spanish Determiner", both of which are absent in English, as exemplified in the pair *the table* and *la*.FEM *mesa*. Similarly, Moro Quintanilla (2014) accounts for this asymmetry in terms of the "presence of an uninterpretable gender feature on the Spanish determiner, as opposed to its absence on the English determiner" (p. 222).

3. Corbett (1991) refers to this process as semantic assignment.

However, neither Liceras et al. (2008) nor Moro Quintanilla (2014) look outside of the noun phrase for an explanation of why these patterns occur. Indeed, various studies have demonstrated that the language of the determiner is congruent with that of the matrix language of the clause, as in 'my mom got the *manguera* (hosepipe)' (from the Bangor-Miami corpus, Deuchar et al., 2014), where the matrix language provides key functional categories, such as finite verb morphology (Myers-Scotton, 2002; see also Blokzijl et al., 2017; Eichler et al., 2012; Herring et al., 2010; Parafita Couto & Gullberg, 2019). The choice of language of the determiner as opposed to the choice of gender of the determiner are separate issues, and should therefore be treated separately from an empirical perspective.

To reiterate, in this chapter we focus on mixed noun phrases where the determiner (or other gender-bearing element in the absence of the category of determiner) is marked for gender and the other-language noun may or may not originate from a gendered language. In cases where it does, then gender may or may not be predictable from the semantics and/or form of that noun.

Let us now turn to the three gender assignment strategies that have been attested in code-switching contexts, which we term: (i) translation equivalent, (ii) shape-based, and (iii) default.

2.1 Translation equivalent strategy

Also referred to in previous literature as rhyme analogy, synonymic gender, gender copy, analogical gender (see references in Montes-Alcalá & Lapidus Shin, 2011), as well as the "analogical criterion" (Liceras et al., 2008), speakers using this strategy assign gender to an other-language insertion on the basis of the gender of its translation equivalent in the recipient (or matrix) language. Various studies have found a preference for this strategy, including in language dyads where only one language has gender, and others where both languages possess a gender system.

2.1.1 *Dyads with one gendered language*

Perhaps most recently, Bellamy and Wichers Schreur (2021) find an overwhelming preference for the translation equivalent strategy amongst Tsova-Tush – Georgian early sequential bilingual adults (n = 12) in Georgia. Tsova-Tush possesses a five-way gender system, comprising masculine, feminine, and three 'neuter genders', labelled B, D and J for the form of the agreement markers.[4] This preference is

4. Masculine and feminine genders include only masculine and feminine humans respectively, while the remaining genders are both animate and inanimate, from a variety of semantic domains (see Wichers Schreur (2021) for an in-depth discussion).

almost at ceiling in semi-naturalistic speech elicited via a director-matcher task, but less pronounced in a forced-choice acceptability judgement task. Note also that the translation equivalent strategy was applied to all instances of code-switched noun phrases in a corpus of Tsova-Tush speech (Wichers Schreur, 2018), but the number of tokens is too low (n = 5) to be instructive. Example (2) demonstrates how the Georgian noun (in italics) receives J gender agreement on the adjective since the Tsova-Tush translation equivalent *sango* 'yard' is a J gender noun.

(2) j-axxe^{n} *ezǒ*
J-long yard
'long yard'

Relatedly, Liceras et al. (2008) demonstrate that their L1 Spanish-L2 English adult participants in Spain (n = 72) also prefer the translation equivalent strategy in a Likert scale judgement task, although their other experimental groups opt for a default strategy in both spontaneous speech and judgements (see Section 3). Late sequential L1 Spanish-L2 English bilinguals (n = 35) prefer the translation equivalent strategy over any other in an acceptability judgement task, in contrast to L1 English speakers with an intermediate level of L2 Spanish (n = 43), who showed no clear preference (Klassen & Liceras, 2017).

L1 Spanish speakers (five adults and eight children) in Spain also display a preference for the same strategy in both a 4-point acceptability judgement task and a gap-filling production task (Gómez Carrero, 2015, p. 36). In a forced-switch elicitation task, L1 Spanish-L2 Basque and Spanish-dominant 2L1 Basque-Spanish speakers (n = 17) also relied predominantly on the translation equivalent strategy (Munarriz-Ibarrola et al., 2021). Similarly, 12 early young bilingual Basque-Spanish bilinguals also prefer the translation equivalent strategy over a shape-based strategy in an adapted version of Liceras et al.'s (2012) English-Spanish reading acceptability judgement task (Iriondo, 2017).

On the basis of a corpus of naturalistic spoken and written bilingual speech, some Estonian nouns inserted into Russian are assigned gender on a "semantic" basis, such as *tolmulapp* 'duster', which takes feminine gender marking in line with its Russian translation equivalent *trjapka* (Zabrodskaya, 2009, p. 369). Notably, the majority of speakers in this corpus were sequential bilinguals, with Russian as their L1. Finally, while the naturalistic corpus data from German-English bilinguals (n = 20) in the USA contains evidence for multiple gender assignment strategies, the translation equivalent emerges as a strong tendency (Füller & Lehnert, 2000). It is noteworthy that these final two corpus studies do not show a preference for a default strategy; the translation equivalent seems particularly prevalent in experimental tasks in some communities.

2.1.2 *Dyads with two gendered languages*

Studies including two gendered languages are less numerous than those including just one. Cantone and Müller (2008) find that four Italian-German speaking children aged 1;8-5 encode on the determiner (which is most frequently Italian) the gender of the noun (which is in the other language), irrespective of its gender in the same language. Stolz (2009) refers to this process as 'gender copy' in relation to Romance loanwords (as identified in the dictionary) in Maltese. The same children also, more frequently, provide mixed NPs in which the noun takes the gender of the switched noun rather than that of its equivalent, where they differ (Cantone & Müller, 2000, p. 821). Similarly, Spanish-German bilinguals in Barcelona assign gender to code-switched masculine and feminine nouns on the basis of their gender, not that of the translation equivalent, as in (3a–b) where Spanish determiners are italicised.

(3) a. *la* Hose 'the.FEM trousers(f)' (cf. el pantalón – masculine)
b. *el* Gürtel 'the.MASC belt(m)' (cf. la cintura – feminine)
(adapted from López, 2020, p. 108)

The mixed NPs in (3a–b) clearly demonstrate that the Spanish determiner is assigned the same gender as that of the German noun. Neuter nouns, for which there is no analogical gender category in Spanish, are also assigned a masculine determiner, such as *el brötchen* 'the.MASC bun.NEUT' (González-Vilbazo, 2005, cited in López, 2020).

Treffers-Daller (1993) analysed naturalistic and elicited production data from 34 Dutch-French bilinguals in Brussels and found a strong preference for an analogical strategy based on the gender of the noun in the original language. The French gender system distinguishes masculine and feminine gender, partly overlapping with the Brussels Dutch system, which, unlike Netherlands Dutch, distinguishes three genders (masculine, feminine and neuter). The Dutch-French bilinguals assigned the gender of French nouns to French noun insertions in Dutch. Nouns that are, for example, masculine in French received Dutch masculine gender when inserted in Dutch rather than the gender of the Dutch translation equivalent.

Nonetheless, the way in which the translation equivalent strategy emerges in mixed noun phrases between gendered languages, including those with more extensive gender systems has yet to be treated. How gender would map between two systems differing in extent and internal assignment principles, including how a default would be assigned (if at all) are empirical questions worth pursuing (see Section 4). We now turn to the next strategy.

2.2 Shape-based strategy

Speakers may associate an element of the code-switched noun from the donor language – often, but not always, its ending – with a similar element in the recipient language, thereby assigning the former the gender associated with the latter. This has been referred to as the phonological strategy in much of the previous literature (e.g. Munarriz-Ibarrola et al., 2021), but we will refer to it here as a *shape-based strategy*, since examples from the literature do not always refer to spoken data or tasks. Indeed, in written code-switches, orthographic similarities (which largely reflect the phonological form of the words) determine gender assignment according to this principle (Montes-Alcalá & Lapidus Shin, 2011, p. 122).

It is striking that the shape-based strategy has been reported less frequently than the other two attested gender assignment strategies (translation equivalent and default). In a multi-task study focusing on Basque-Spanish bilinguals (n = 30), Parafita Couto et al. (2015) found that participants associated the Basque definite marker *–a* with the Spanish feminine marker *–a* in naturalistic speech and also preferred this form in a judgement task, resulting in mixed forms such as *la liburu-**a*** 'the.FEM book-DEF' (cf. *el libro* 'the.MASC book').

Badiola and Sande (2018) find a similar pattern in a judgement task with Basque-Spanish simultaneous bilinguals (n = 21), whereby Basque nouns terminating in *-a* without a definite marker are also assigned the Spanish feminine definite article *la*, such as *la **azterka*** 'the.FEM exam' (Badiola & Sande, 2018, p. 17). Adding to the Basque-Spanish debate, Munarriz-Ibarrola et al. (2021) find that L1 Basque-L2 Spanish bilinguals (n = 13) rely predominantly on the shape-based strategy in a forced-switch elicitation task (see also Section 2.3 on how these studies present conflicting results regarding the use of default strategy).

Similarly, in an online two response forced-choice acceptability judgement task, Purepecha-Spanish bilinguals (n = 12) demonstrated a preference for the Spanish feminine definite article *la* with inserted Purepecha nouns terminating in *-a*, as in *la japonda* 'the.FEM lake', where *el lago* is a masculine noun in Spanish (Bellamy, Parafita Couto, & Stadthagen-González, 2018).

In a comparative experimental and observational study of Russian-English bilinguals in Russia and the USA, Chirsheva (2009, p. 75) found that the "phonological shape of English code-switched nouns plays the most important role in their gender assignment" with Russian adults in the USA using this strategy 89% of the time, as opposed to 65.3% for Russian students in Russia with English as their L2. One of the various strategies employed by Russian-Estonian bilinguals in Estonia is to assign case and gender, or gender alone, on the basis of the phonological shape of the Estonian noun. For example, the Estonian term *kaïbemaks* 'value added tax' is treated as a second declension masculine Russian noun (i.e. with zero ending)

and therefore receives the masculine genitive marking *-a*, see (4), where Estonian is marked in italics.

(4) No eto bez *kaïbemaks*-a
But this.MASC without value.added.tax-GEN.MASC
'But this without value added tax' (Zabrodskaya, 2009, p. 369)

Likewise, Leisiő (2001, reported in Chirsheva, 2009, p. 64) found that when Finnish nouns are inserted into Russian in code-switching contexts, the prevailing gender assigned is feminine due to their phonological shape and consequent association with feminine forms in Russian.

Finally, Bellamy and Wichers Schreur (2021) also show how the similarity between the initial phonological segment of the inserted Georgian word and the form of the Tsova-Tush gender marker (B, D or J) significantly influences the gender assigned in a three response forced choice acceptability judgement task, although the effect is also observed to a lesser extent in production. Recall that Tsova-Tush marks gender word-initially, which may have made it more salient to the participants in this study, since the repetition may create an alliterative association.

2.3 Default strategy

Perhaps the most common gender assignment strategy encountered in the literature is the default strategy, where speakers use one gender for most other-language insertions, irrespective of their semantic and morpho-phonological properties, as well as the gender of their translation equivalent.[5] On the basis of naturalistic corpus data, it has been demonstrated that various Spanish-English speaker communities apply a masculine default to the vast majority of English nouns in mixed noun phrases. Otheguy and Lapidus (2003, p. 214) found that 87% of English language insertions (ELIs) in the 33 sociolinguistic interviews that they conducted with Latinos in New York City were assigned masculine gender, in contrast to the 53% of nouns taking masculine gender in the unilingual Spanish speech of three of their informants. The other 13% of inserted English nouns were assigned feminine gender on the basis of phonology and animacy: first, word-final schwa or /*-ar*/ in English was interpreted as *-a* in Spanish, attracting feminine gender, see (5).

(5) boiler > *la boila*
report card > *la repocá* (Otheguy & Lapidus, 2003, p. 215)

5. It may be, however, that the fact that this strategy is so frequently encountered in the literature is due to the over-representation of Spanish-English studies.

Second, biological sex determined the grammatical gender of a noun, such that a woman who drinks heavily, for example, was *una lush* 'a.FEM lush' (Otheguy & Lapidus, 2003, p. 215).

Aaron (2015) finds a similar preference for the masculine default in Spanish-English code-switching in the naturalistic speech of bilinguals from northern New Mexico (see Torres Cacoullos & Travis, in preparation). Of the 239 'singleton' nouns elicited in this dataset, 101 (42%) were assigned masculine gender, as opposed to just 16 (7%) feminine, although the majority (51%) were assigned no gender at all (Aaron, 2015, p. 469). The author states that this preference for masculine assignment "likely has nothing to do with code-mixing tendencies per se, but may rather simply follow from patterns and preferences that are internal to Spanish" (ibid), namely that masculine has an unmarked or default status (e.g. Beatty-Martínez & Dussias, 2019).

Clegg and Waltermire (2009) analysed the naturalistic speech of 15 Spanish-English bilinguals in northern New Mexico according to three factors: biological gender of the inserted noun, synonymic gender (i.e. the gender of the translation equivalent), and the terminal phoneme of the inserted noun. Using the VARBRUL programme (Rand & Sankoff, 1990), the authors demonstrate that nouns referring to animate referents are categorically assigned the gender of the referent, but for inanimates, the terminal phoneme is the deciding factor: English nouns with typical masculine Spanish endings are assigned masculine gender in 89% of instances, although it is unclear whether this is a phonological analogy or a default strategy. In contrast, English nouns ending in phonemes typically associated with feminine nouns in Spanish are assigned feminine gender in only 50% of cases. This asymmetry suggests that a masculine default strategy also applies.

In contrast, Balam (2016) found that biological gender was not deterministic in gender assignment. Indeed, in this extensive study of the naturalistic production of 62 speakers of Northern Belizean Spanish, Balam (2016, 420) found that 99.6% of tokens were marked with masculine gender, including both male and female human referents (e.g. *los nuns* 'the.MASC nuns', *los women* 'the.MASC women').

Similarly, Valdés Kroff (2016) reported on the gender assignment preferences of Spanish-English speakers in the Bangor Miami Corpus. Of the 304 mixed noun phrases containing a Spanish determiner and an English noun, 297 (93.7%) were assigned masculine gender and just 8 (2.5%) feminine. This means that masculine gender was being used with almost all nouns, irrespective of the gender of their translation equivalent. The small number of feminine-marked mixed noun phrases almost completely concerned female human referents, such as *la assistant* 'the.FEM assistant' (cf. *la asistente/ayudante*; Valdés Kroff, 2016, p. 290). That said, female animates who would usually be assigned feminine gender in unilingual Spanish were also assigned masculine gender as in (6).

(6) Ella es un *renaissance woman*
She be.3.SG ART.DEF.MASC renaissance woman
'She is a renaissance woman.' (adapted from Valdés Kroff, 2016, p. 291)

DuBord (2004) analysed mixed noun phrases as occurring in interviews with 18 Mexican-American Spanish-English bilinguals in Southern Arizona. Overall, 130 of the 174 tokens (74.7%) transcribed were assigned masculine gender and 43 (24.7%) assigned feminine gender. One token (0.6%) was assigned both genders. Biologically masculine (animate) referents received masculine assignment in an almost unambiguous 96.2% of cases, while feminine assignment to feminine animates was also high (75%) but based on far fewer tokens: 8 as opposed to 27. Phonologically masculine nouns were assigned masculine gender in 82.3% of cases overall, while phonologically feminine nouns were only assigned feminine gender in 38.9% of cases. Phonologically neutral tokens were also assigned masculine gender in the majority of instances (70.7%), indicating a default strategy (see DuBord, 2004, p. 34). When the translation equivalent of the inserted English noun was masculine, masculine gender was assigned in 82.3% of cases, whereas the same calculation for the feminine is just 40%. Moreover, the four nouns with no clear translation equivalent in Spanish were all assigned masculine gender (DuBord, 2004, p. 35). While it is hard to adjudicate between assignment strategies for masculine nouns, the preference for masculine gender with phonologically and semantically (as well as biologically) feminine nouns is indicative of a default strategy.

Chaston (1996, p. 201) identifies 42 code-switched noun phrases in his corpus of Spanish-English bilinguals (n = 18) of Mexican heritage in Texas, 40 of which are assigned masculine gender. He acknowledges that this tendency is not merely a coincidence, but is unsure whether it occurs because the speaker does not know the gender of the Spanish translation equivalent, or because speakers tend to select the masculine article with English words or new cognates (Chaston, 1996, p. 202). More recently, using a two response forced choice acceptability judgement task, Delgado (2018) has shown how heritage Spanish-English bilinguals in Chicago (n = 21) assign masculine gender to all non-familial words, including newly encountered words, but use the feminine (and masculine) for the translation equivalent in the familial setting.

In an acceptability judgement task, Liceras et al. (2008) found a masculine default preference for adult Spanish L2 speakers (n = 142, of whom 61 were L1 English and 74 L1 French), in contrast to the translation equivalent for Spanish L1 speakers (n = 72). Moreover, in her study of the natural speech of one L1 English-L2 Spanish speaker living in Argentina, Franceschina (2001, p. 239) found that he "only used masculine forms of articles in combination with English nouns". In a sentence

selection task,[6] heritage Spanish-English simultaneous bilinguals in Canada assigned masculine gender to English nouns with a feminine translation equivalent in 40–50% of cases, with the ending of the noun (i.e. *-a, -e* or a consonant) not playing a significant role (Valenzuela et al., 2012). Regarding simultaneous Basque-Spanish bilinguals from Gernika (n = 21), Badiola and Sande (2018) identified a masculine default strategy in response to a 7-point Likert scale acceptability judgement task. Participants provided higher acceptability ratings to Basque nouns without lexical *-a* accompanied by a Spanish masculine determiner, irrespective of the gender of the translation equivalent. In contrast, Basque nouns with lexical *-a* were given higher ratings with a feminine determiner, also regardless of the gender of the translation equivalent (Badiola & Sande, 2018, pp. 31–32; cf. Parafita Couto et al., 2015).

Moving away from language dyads including Spanish, Chirsheva (2009 and references therein) lists various language pairs in which nouns are assigned masculine gender as a default, namely Russian-English, American Norwegian-English, American Lithuanian-English, American Portuguese-English, American Italian-English and French-English, although these may be references to borrowing rather than code-switching, following the definition we provide in Section 1. Nonetheless, the clear preference for a masculine default is striking. Some studies on Dutch (which differentiates between common and neuter gender) in contact with other languages also point towards common gender as a default. Clyne (1977; see also Clyne & Pauwels, 2013) analyses a corpus of elicited production data of 200 English-Dutch bilinguals in Australia and reports a common gender default as the main strategy, driven by the phonetic similarity between the Dutch determiner [də] and the English determiner [ðə]. Boumans (1998) analysed naturalistic speech recordings of 15 Moroccan Arabic-Dutch bilinguals in the Netherlands and found that common gender was also assigned in all cases of Moroccan Arabic insertions into Dutch.

However, the default need not always be masculine. Take, for example, the findings of Parafita Couto et al. (2015), where Basque-Spanish bilinguals re-interpret the word-final definiteness marker *-a* in Basque as analogical with the feminine gender, leading to a feminine default (see also Section 2.2). Weinreich (1953, p. 45) reports a feminine default tendency for borrowed English nouns in American German and American Yiddish, but only as one of several gender assignment strategies. While not strictly a feminine default, English nouns with a masculine-like phonetic form in Ukrainian were assigned feminine gender in 42% of cases, twice as often as the

6. In this task "participants read a dialogue between two bilingual speakers that included either a code-switched DP or an agreement copula sentence. They were asked to choose the concluding statement that sounded most natural to them" (Valenzuela et al., 2012, p. 486).

masculine gender (21%) by 25 L1 Ukrainian-L2 English speakers residing in the USA. English nouns with feminine-like phonetics receive feminine gender in 88% of cases, but never masculine (Budzhak-Jones, 1998, p. 176). Note, however, that the gender of the translation equivalent was not provided for the English code-switches in Budzhak-Jones' study, therefore many of these instances of seemingly default strategy may in fact be more appropriately analysed as applying the translation equivalent strategy. having presented an overview of gender assignment in mixed NPs in adult speech; let us now consider what occurs child language.

2.4 Child code-switching patterns

Studies of gender assignment in mixed noun phrases amongst children are still uncommon in the code-switching literature; we provide a short overview of the most notable ones here. Cantone and Müller (2008) focus on the free production of four simultaneous German-Italian bilingual children aged 1;8-5, finding that the translation equivalent strategy is preferred. Radford et al. (2007) find similar (but not unambiguous) results for a single Italian-English bilingual child (Lucy), although she also uses a number of hybrid forms, whereby a terminal vowel is added to the English noun in order to render it gendered, like Italian, as in *la butterflaia ~ la farfalla* ('the.FEM butterfly'). Reminiscent of the Spanish-German data presented in Section 2.2.2, both balanced and unbalanced bilingual children of German and French, Italian or Spanish ($n = 13$) or two of these Romance languages "more frequently mark the gender of the noun actually switched on the determiner than the gender of the equivalent noun" (Eichler et al., 2012, p. 250). Notably these two studies both involve German and Spanish bilingual children. However, Ezeizabarrena (2009) finds no clear preference for a single gender assignment strategy in the small number of tokens elicited in the speech of a bilingual Basque-Spanish child, in a longitudinal sample of over 20 hours of adult-child conversation.

Balam, Lakshmanan and Parafita Couto (2021), in contrast, demonstrate that early simultaneous Spanish-English bilingual children ($n = 40$) of different grade levels (second to fifth grade in English immersion and two-way bilingual programmes) in Miami use the masculine default strategy when assigning gender to English code-switched nouns in Spanish. In both cases the children evince native-like acquisition of the respective gender systems in unilingual speech, but display differing behaviour in mixed noun phrases (see also Section 4).

Similarly, Fernández Fuertes, Liceras and Bel (2011) find that simultaneous bilingual Spanish-English children ($n = 11$) provide a higher reference score for a default option in an acceptability judgement task than L1 Spanish children. Finally it is also worth mentioning the study by Jorschick et al. (2011) on the naturalistic

speech of three German-English bilingual children aged two to four years, in which the authors provide evidence that these children prefer the translation equivalent strategy at a higher than chance level.

While data regarding the acquisition of code-switching patterns in children, including gender assignment in mixed noun phrases, is still in its infancy, the results available suggest that different strategies may occur in unilingual and code-switched speech, and that these patterns are learned in childhood.

For reference, an overview of the main findings of each study presented in this section can be found in Appendix 1. The role of type of bilingualism will be treated in the following section, along with other extralinguistic factors.

3. Extralinguistic factors

As should be clear from Section 2, there seems to be variability between bilingual speaker communities, even of the same language pair, as to which gender assignment strategy is applied most frequently in a given study. It is also rare for one speaker community to use only one strategy. However, it should be noted that the populations studied vary across a number of extralinguistic factors, namely order of acquisition of the languages in question (i.e. the type of bilingual), the type of data and tasks that gave the results, the medium of the language under investigation (written vs. oral), and community norms. We will outline here how results pattern in line with these factors.

3.1 Type of bilingual

Studies of gender assignment include a wide variety of bilingual speaker types: simultaneous (2L1), early and late sequential, heritage speakers, and L2 learners. The extent to which these speaker groups code-switch, and therefore use and/or hear mixed noun phrases on a regular basis, also varies between studies. It is important, therefore, to draw out the main findings from these varying groups.

It has been claimed that early simultaneous bilingual adults are more likely to show a preference for the (masculine) default gender assignment strategy (López, 2020; Liceras et al., 2008), a claim largely supported by the results in Badiola and Sande (2018), focusing on simultaneous Spanish-Basque bilinguals. In contrast, sequential bilinguals who learned the gendered language first (in this case, Spanish) may be more likely to prefer a translation equivalent strategy (e.g. Iriondo, 2017; see also Liceras et al., 2016; Fernández Fuertes et al., 2011 for Spanish-English). Bilinguals who had Spanish as (one of) their L1(s) and Basque as their L2 were

found to rely predominantly on the translation equivalent strategy in an elicited forced-switch production task, whereas those who had Basque (no gender) as their sole L1 mainly preferred a shape-based (here, phonological) l strategy (Munarriz-Ibarrola et al. 2021).

Furthermore, Bellamy and Wichers Schreur (2021) find that sequential Tsova-Tush-Georgian bilinguals also prefer the translation equivalent strategy, particularly in production, which seems to support López's (2020) claim since Tsova-Tush possesses a five-way gender system. However, the participants in Parafita Couto et al. (2015) were predominantly L1 Spanish, acquiring Basque (ungendered) on average at the age of 3.46 (+/− 1.55), yet they displayed a preference for a feminine default strategy in mixed NPs. We discuss possible explanations for these preferences in Section 4.

Also relevant to the choice of gender assignment strategy is a bilingual's relative dominance in the two languages although, again, this is not always reported in code-switching studies (e.g. Zabrodskaya, 2009). Moreover, individual dominance should not be conflated with language dominance at the societal level. Dominance is dynamic, therefore an individual's dominant language in childhood may not be the same as in adulthood, and fluctuation between the two is likely to occur across the lifespan (e.g. Grosjean, 2001). It has been claimed that if a speaker's dominant language is Spanish (i.e. a gendered language), then this speaker should prefer the translation equivalent strategy (López, 2020, p. 78; Klassen & Liceras, 2017, p. 82; see also Otheguy & Lapidus, 2003; Liceras et al., 2008).

Corroborative findings are reported by Fuller and Lehnert (2009) for German-dominant German-English bilinguals; recall that German possesses a three-gender system while English lacks grammatical gender. By the same token, bilinguals dominant in a non-gendered language should prefer a default strategy (e.g. Klassen & Liceras, 2017, p. 83). This claim is supported by Gómez Carrero (2015) for L1 English-L2 Spanish speakers in Spain, Valenzuela et al. (2012) for heritage Spanish speakers in Canada who are dominant in English, and Liceras et al. (2008) for L1 English-L2 Spanish bilinguals (but see Klassen & Liceras, 2017 for inconclusive results for L2 Spanish-L1 English bilinguals), and by Munarriz-Ibarrola et al. (2021) for Spanish-Basque bilinguals. However, such a preference is not found among the Tsova-Tush-Georgian participants in Bellamy and Wichers Schreur's (2021) study; all were dominant in Georgian, an ungendered language, yet they showed a marked preference for the translation equivalent strategy, with little to no evidence of a default strategy in comprehension and production respectively. Perhaps this counter-evidence highlights the need for more studies investigating language pairs beyond Spanish-English.

3.2 Task type

A variety of methods have been employed to investigate gender assignment in mixed NPs, each of which elicits a particular type of data, for which different preferences have been observed. For production, both spontaneous and semi-spontaneous, or guided speech have been collected: spontaneous speech from a particular speech community is drawn together in a corpus, which is then transcribed and annotated accordingly. Perhaps the best known and most widely used corpus in code-switching studies so far is the Bangor Miami corpus[7] of Spanish-English speakers, 35 hours of spontaneous recorded conversation, fully accessible online (Deuchar, 2013; see Blokzijl et al., 2017 and Valdés Kroff, 2016 for examples of studies analysing data from this corpus). An additional corpus of Spanish-English bilingual speech is the New Mexico Spanish-English Bilingual (NMSEB) Corpus (Torres Cacoullos & Travis, 2015; in prep.), although it is not publicly available.[8]

Semi-spontaneous production is elicited in many studies, since collecting this type of data enables the researcher to control more carefully for the features under investigation, especially those that occur infrequently in corpora. Methods utilised include (semi-)structured interviews (e.g. Otheguy & Lapidus, 2003), prompted monologues (e.g. Pearson, 2002 who used the 'frog story'), and interactive game-like tasks, such as the director-matcher or 'toy task' (e.g. Bellamy et al., 2018, see Gulberg et al., 2009 for a description), or the related map task (e.g. Beatty-Martínez & Dussias, 2017).

Acceptability judgement tasks are often used to test 'comprehension', although given the negative social attitudes often associated with code-switching, it is debatable as to what precisely judgement tasks are tapping into (e.g. Stadthagen-González et al., 2018). Nonetheless, both Likert scale (e.g. Liceras et al., 2008) and forced-choice (e.g. Bellamy et al., 2018) judgement tasks are found reported in the literature (see also Parafita Couto & Stadthagen-González (2017) on determiner assignment in mixed NPs, using both types of judgement task).

Furthermore, psycholinguistic online methods to code-switching are being increasingly applied to investigate how code-switches, including mixed NPs, are processed. Underpinning this line of research is the hypothesis that comprehension ought to reflect production patterns, for example when masculine default

7. See bangortalk.org.uk. This site also includes spontaneous data from Welsh-English bilinguals (the Siarad corpus), as well as Welsh-Spanish bilinguals in Patagonia.

8. Note that other bilingual corpora also exist, such as the extensive Ottawa-Hull corpus of French-English bilingual speech although it is also not open access and has not, to the best of our knowledge, been used to systematically study gender assignment patterns in this community (see Poplack, 2018).

agreement is applied to both masculine and feminine nouns, what Valdés Kroff (2012. p. iii) calls a "production asymmetry".

Indeed, Valdés Kroff et al. (2017) employed the visual world paradigm to demonstrate how Spanish-English bilinguals reflected their asymmetric use of gender in mixed NPs in comprehension. Using a 2-picture design with auditory input (typical to, and one of the key advantages of, this experimental paradigm), the authors tested whether gender, as marked on a Spanish determiner, facilitates the identity of the upcoming English noun (ibid, p. 6). They found that their bilingual participant group only made use of the feminine cue to facilitate noun identification, reflecting their propensity to use the masculine default in code-switching mode.

Fernández Fuertes, Gómez Carrero and Martínez (2020) also used eye-tracking to test how mixed NPs were processed by L1 Spanish-L2 English bilinguals (n = 19) in Spain. The results of a sentence-reading task indicated that participants processed gender-congruent switches faster than incongruent switches. In other words, mixed NPs formed according to the translation equivalent strategy were processed faster than those adhering to a masculine default. The authors relate these findings to those of Liceras et al. (2008), claiming that they support an adapted version of their Grammatical Features Spell-Out Hypothesis, which states that code-switching choices will favor the functional categories containing the largest array of uninterpretable features.

While it is extremely important that data be gathered using different methods, we must also bear in mind that different tasks can elicit different gender assignment strategies, even within the same test population. Some tasks may well be drawing on top-down, social or pragmatic influences that manifest themselves as particular strategies, or switch costs in the case of processing studies (Beatty-Martínez et al. 2018, p. 3; see also Beatty-Martínez & Dussias, 2019).

Take, for example, the multi-method study with early sequential Purepecha-Spanish bilinguals in Mexico reported in Bellamy et al. (2018). In the director-matcher task testing production, the participants displayed an overwhelming preference (over 90%) for a masculine default strategy, whereas in the two option forced-choice acceptability judgement task, the phonology of the inserted Purepecha noun largely determined the gender agreement chosen in Spanish.

In a study using the same methodology, Bellamy and Wichers Schreur (2021) found that Tsova-Tush–Georgian early sequential bilinguals preferred the translation equivalent strategy in a director-matcher task, but showed a more varied set of responses, with the shape-based strategy being more present, in a judgement task. Bierings, Parafita Couto, and Mateo Pedro (2019) also found a similar possible task effect for their unexpected mixed noun-adjective order in a director-matcher task with Kaqchikel-Spanish bilinguals (n = 20) in Guatemala.

Parafita Couto et al. (2015) had already identified differences in strategy as a function of data type, with the naturalistic speech of Spanish-Basque early sequential bilinguals displaying evidence of a shape-based strategy, but a feminine default emerged as the preferred strategy in a three-choice auditory judgement task (see also Gómez Carrero, 2015).

Multi-method studies are still not the norm in code-switching research, with many continuing to rely solely on judgement tasks (e.g. Badiola & Sande, 2018; Delgado, 2018; Vanden Wyngaerd, 2021). However, it is vital that naturalistic and experimental (both for production and comprehension) data are gathered for as many communities and language pairs as possible in order to be able to tease apart the results stemming from possible task effects from natural behavioural patterns (see Valdés Kroff, 2016). Such a wealth of data would enable us to move further towards the long-term goal of a predictive model of gender assignment in code-switching.

3.3 Community norms

Code-switching is not equally common in all bilingual communities. In some, such as among Spanish-English speakers in Miami and Basque-Spanish speakers in Gernika (Basque Country), it is ubiquitous, occurring both intra- and inter-clausally (Badiola & Sande, 2018, p. 22; Beatty-Martínez & Dussias, 2019). Other communities show lower levels of switching, such as Spanish-English bilinguals in El Paso, Texas and Granada (Spain) (Królikowska et al., 2019). Consequently, we should take into account whether participants are habitual code-switchers or not when designing studies.

While still in its early stages, research that explicitly takes into account the variation inherent in the bilingual experience is advancing our understanding of how individuals produce, process and comprehend code-switches (Beatty-Martínez et al. 2018). The results of three production tasks undertaken by 32 heritage Spanish-English bilinguals highlighted how regular and infrequent code-switchers favoured the Spanish masculine determiner with an English noun whose translation equivalent is feminine (i.e. a gender mismatch, generally construed as a masculine default), whereas participants who did not engage in code-switching did not exhibit this tendency (Denbaum & Prada Pérez, 2020, p. 28). Beatty-Martínez and Dussias (2017) used ERPs to show how habitual and non-habitual code-switches process congruent and non-congruent switches in the DP differently, as a function of their code-switching experience. Code-switches that adhered to patterns used in an accompanying production 'map task' resulted in no processing cost in the ERP experiment (see also Beatty-Martínez et al., 2018).

Moreover, input is crucial: the norms common to the wider speech community will be reflected in the communicative behaviour of the individual (Poplack, 1980). These norms, in our case regarding choice of gender assignment strategy, are likely to be community-specific and therefore may not be shared by speakers of the same language pair in different places (cf. Gardner-Chloros, 2009; Toribio, 2017).

Królikowska et al. (2019) investigated gender assignment in mixed NPs in four Spanish-English speaking communities, namely Granada (Spain), Pennsylvania State University (USA), Puerto Rico (USA) and El Paso, Texas (USA) in order to identify whether differences can be observed between communities. Using a map task, they found that both the translation equivalent and default strategies were present in all communities, but that in Pennsylvania and Puerto Rico participants preferred the default strategy over the translation equivalent, whereas in Granada and El Paso, both were used with roughly equal frequency. Results from a language background questionnaire highlighted that the bilingual participants in Puerto Rico code-switched the most and those in Granada the least.

Balam (2016) also found that his participants, who also frequently engaged in code-switching, preferred the masculine default strategy. It appears, therefore, that there is a connection between amount of code-switching and assignment strategy: the more speakers switch, seemingly the more likely they are to use the default. More research is needed into comparing communities sharing the same language pair (cf. Beatty-Martínez & Dussias, 2019), also including naturalistic data from the communities in question (see Section 4).

Studying patterns in language pairs that keep one language constant but vary the second language (which may or may not be the dominant language) is also highly instructive as a way of trying to tease apart the relative contribution of linguistic and extralinguistic factors in gender assignment. Take, for example, pairs where Dutch occurs with a variety of other languages, such as Spanish, Papiamento and Turkish, all spoken in The Netherlands. Van Osch et al. (2022) find that "in code-switching mode, most speakers tend to assign common gender to inserted nouns, but some speakers also apply a gender assignment strategy based on the translation equivalent of the noun in Dutch […]." While this study seems to indicate speakers converging on one preferred strategy, the extent to which this pattern is generalisable to other language dyads remains an area for much more research.

A further empirical question to be explored in more depth is whether languages sharing the same structure (which are also often genealogically related) also display the same gender assignment patterns when in contact with the same language. For example, do Spanish and Italian, which share a binary gender system descended from Latin, behave in the same way when spoken as one part of a dyad with German? González Vilbazo (2005; cited in López, 2020, p. 108) claims that code-switched German nouns are assigned the same gender in Spanish, not the

gender of the Spanish translation equivalent.[9] Therefore, German feminine nouns are marked with a Spanish feminine determiner, while masculine and neuter nouns are marked with a masculine determiner. Similarly, bilingual Spanish-German, Italian-German and French-German children prefer to mark the gender of the noun that has been switched over the translation equivalent, although this strategy does also occur (Eichler et al., 2012, p. 250). Unfortunately, there are too few studies of (closely) related languages to be able to comment further on this apparent parallel.

Moreover, situations where one language in the dyad is kept the same, both within the same country as well as in different locations, are instructive regarding gender assignment patterns. Varieties of a number of European immigrant languages in the USA, including American Norwegian, American Lithuanian, American Portuguese and American Italian, all in contact with English, have been reported as using a masculine default for all loanwords (Chirsheva, 2009, p. 64 and references therein). However, Weinreich (1953, p. 45) also reports that American German and American Yiddish assign a feminine default most frequently to English nouns in bilingual speech. This small sample already suggests that language pairs sharing one societally dominant language may not necessarily treat other-language insertions in the same way, indicating that multiple linguistic and extralinguistic factors are at work.

3.4 Child language acquisition

If input and community norms play such an important role in shaping the gender assignment strategies adopted by a particular community (amongst other features), then it is clear that we need to know how children acquire them. The code-switching patterns produced by adults in relation to those acquired by children is a seriously under-researched area in code-switching studies, but one that has the potential to greatly expand our understanding of this linguistically constrained but societally-variable speech activity (see Deuchar forthcoming for an overview of research into child code-switching).

Lanza (1997) represents probably the first attempt at such an endeavour, reporting on code-switching in both child utterances and adult responses for two simultaneous Norwegian-English bilingual children and their caregivers. There

9. Note that Spanish-German here refers to the bilingual speaker community of this language pair in Barcelona studied by González-Vilbazo (2005), rather than to Spanish-German bilinguals more broadly. As we have seen from the Spanish-English data, data are required from many speaker communities in order to identify the code-switching patterns present. We thank an anonymous reviewer for ensuring we highlight this point.

is evidence that one of the children involved in this study, Siri, not only repeats mixed noun phrases produced by her father, such as *i vindu-et* 'the window-DEF. NEUT' (where *vindu* is a phonologically adapted insertion of 'window'), but also produces novel forms on the same template. Two of these forms follow the same gender assignment strategy – the translation equivalent – while one appears to follow a masculine default strategy (ibid, p. 144).

Using naturalistic data accessible in the CHILDES database (MacWhinney, 2000), Balam et al. (2021) demonstrate that both second-grade and fifth-grade children (i.e aged around seven and ten, respectively) in English-immersion and two-way bilingual schools Miami display native-like acquisition of grammatical gender in Spanish, but in mixed noun phrases, they use masculine agreement with feminine translation equivalent nouns. The authors conclude that from the age of seven, children in these communities utilise two different gender assignment strategies, one for unilingual Spanish contexts and one for bilingual contexts. They therefore acquire the two systems, each one appropriate to its own context; they are highly proficient bilinguals, who display little to no deviance from the so-called monolingual standard (MacSwan, 2021, p. 88). Indeed, code-switching occurs in communities of stable bilingualism, such as Miami, as well as in "immigrant communities, regional minorities and native multilingual groups alike" (Gardner-Chloros, 2009, p. 20).

In this section, we have seen how various extralinguistic factors can modulate the gender assignment strategy, which can differ depending on the order of acquisition of the languages involved, the type of task used to elicit such forms, as well as which norms have developed at the community level, irrespective of the language dyad. It is also worth underlining the need for more corpora, of both unilingual and bilingual speech, to further investigate the relationship between code-switching in child and adult language, as well as to gain a more holistic picture of patterns between communities.

4. Discussion and future perspectives

Gender assignment in mixed NPs varies in non-random ways according to a number of linguistic and extralinguistic factors, notably language pair, type of bilingual task type, and community norms. Since some studies do not report all of these factors, their interaction and the relative importance of each one remains unclear. Consequently, both modeling and predicting the preferred gender assignment strategy or strategies for a particular speaker community is a complex task. Given the larger number of studies on Spanish-English code-switching, employing multiple

methodologies and conducted in various speaker communities, we are in a position to outline some initial generalisations regarding the relationship between the factors we have reviewed here and the strategies observed. Given the paucity of studies focusing on other language pairs, especially those possessing more extensive gender systems, however, generalisations in an even broader sense remain distant.

4.1 Bilingual profile

In general terms, the sequential Spanish-English bilinguals tested to date have been shown to prefer the translation equivalent strategy, whereas early simultaneous bilinguals lean towards a masculine default (see Section 3.1). Importantly, the sequential bilinguals reported in these studies have learned the gendered language, Spanish, before the genderless language, English (see also Bellamy & Wichers Schreur, 2021; Munarriz-Ibarrola et al., 2021, amongst others, for supporting evidence). However, recall that dominance also played a role in these latter studies. These findings contradict López's (2020, p. 78) claim that speakers whose dominant language is genderless will prefer the default strategy. Initial results would suggest, then, that acquiring the gendered language first is not the sole determinant of the translation equivalent strategy, but that dominance in the gendered language following simultaneous acquisition also suffices. The interaction and relative strength of factors should therefore be borne in mind when testing predictions in new (or the same) language pairs.

4.2 Mental representation of gender

But how can we account for these varying patterns based on order of acquisition? Munarriz-Ibarrola et al. (2021) relate the preference for the shape-based strategy by L1 Basque-L2 Spanish speakers to lexical representation, and more specifically that their Spanish gender representations are not so stable. Two main linguistic approaches to gender representation in the mental lexicon can be delineated: lexicalist and structural. In the lexicalist approach, a lexically-driven generative linguistic model of grammar, gender is assigned to lexical items (i.e. nouns) through rules in the lexicon, which is language-specific. Gender is therefore an inherent feature of a lexical item in the mental lexicon (see e.g. MacSwan, 2000 for an application to bilingual grammar). The structural approach, in contrast, separates the lexical items from the gender features, with the latter being inserted late in the derivation. This approach is favoured by proponents of exoskeletal models of grammar, such as Distributed Morphology (see notably López, 2020 for its application to gender

in mixed NPs). Given the evidence that a different gender can be assigned to the same inserted noun in different contexts, a model that reflects this possibility in the bilingual grammar would be explanatorily preferable.

The two contrasting linguistic approaches to gender representation find parallels in the neurolinguistic and psycholinguistic literature. Results using different methodologies applied to different language pairs provide conflicting support again for two models: (i) an integrated gender model in which both languages share a gender node (Salamoura & Williams, 2007); and (ii) an autonomous representation model, which postulates that the two gender systems are independent from each other (e.g. Costa et al., 2003). Moreover, evidence converges to show that two distinctive mechanisms underlie grammatical gender selection, not just representation, cross-linguistically. First are languages in which the morpho-phonological cues present on nouns predominantly modulate gender; speakers of Romance languages such as Spanish or Italian make use of this mechanism, as the cognate noun *la luna* 'the moon.FEM' demonstrates. Second are languages that rely more on semantics than the relatively uninformative morpho-phonological properties inherent to the language's nouns; this mechanism is observed in Germanic languages such as Dutch and German (Wang & Schiller, 2019, p. 5; Jescheniak, Schriefers, & Lemhöfer, 2014, p. 3). The L1 Basque speakers presented above may therefore be applying the morpho-phonological mechanism for gender assignment to Basque insertions into Spanish, in the same way as they would to gendered Spanish nouns. We may thus not be dealing with unstable representations as such, but representations associated to those of the L1.

In language dyads where both languages possess gender, if the gender systems are indeed integrated, which implies that the same gender node is activated when the two languages share the same gender but different ones are activated where there is a gender incongruence, then the translation equivalent strategy may only be triggered in cases where the gender is shared. In cases where the genders conflict, then other features, such as the phonology of the noun (i.e. something about its shape), may come into play. If, on the other hand, the gender representations are autonomous, then we might expect a more consistent use of the translation equivalent strategy. It should be underlined, however, that these are purely speculative hypotheses, but hypotheses that merit empirical investigation for a nascent predictive model of gender assignment.

We have also observed in this chapter that the same lexeme in one language can be assigned different genders in unilingual vs. code-switched speech (e.g. Balam et al., 2021), as well as in in response to different tasks (e.g. Bellamy et al., 2018). It is notable that Spanish-English bilinguals who code-switch frequently produce native-like gender agreement in unilingual Spanish contexts, but overwhelmingly

resort to a masculine default when inserting lone English nouns in code-switching mode. This differential behaviour suggests that bilinguals have different modes or speech styles at their disposal, akin to registers in monolingual speakers, which they engage in accordance with the social context. In this way, code-switchers "are driven to follow community-established norms", both in unilingual and mixed speech (Valdés Kroff, 2016, p. 284; see also Valdés Kroff et al., 2017; Aaron, 2015).

Bilinguals who form part of a code-switching community must therefore learn a hybrid system; they require a set of community-specific distributional patterns to be able to understand code-switched speech, as well as to be able to plan upcoming utterances. These patterns seem to be in place by around the age of seven, and follow through into adulthood: Balam et al. (2021) claim that a default gender is used in order to "prioritize a principle of economy [common to multilinguals] rather than to faithfully maintain the grammatical procedures of the gendered language" (see also Otheguy & Lapidus, 2003). The use of the masculine determiner in mixed NPs therefore becomes a useful cue in comprehending code-switched speech (Valdés Kroff, 2016, p. 293). In this way, Valdés Kroff (2016, p. 284) treats code-switching as "by and large a planned mode of bilingual speech […] built from the bilingual's constituent languages".

However, it remains unclear whether such a principle of economy is common to all habitual code-switching communities, since research to date focuses solely on Spanish-English bilinguals. Variation in gender assignment, both in unilingual and code-switched speech, between speaker communities is to be expected due to the multiplicity of linguistic and extralinguistic influences acting on any language at any given time (Valdés Kroff, 2016, p. 298). The norms that emerge are therefore modulated by these influences to different degrees; a more elaborated model of gender assignment in mixed NPs would include all of these factors both within and across language dyads.

Hybrid representation of gender, and perhaps also the principle of economy, could also be applied to the differential patterns observed in mixed NPs between tasks. We have seen how some bilinguals are able to assign a different gender to the same noun in different contexts, suggesting hybrid production, if not hybrid representation in the mental lexicon. Recall here the more unstable representation of Basque nouns in the Spanish-Basque lexicon suggested above; perhaps this flexibility allows for differential gender assignment depending on the type of trigger or data source, for example written vs. aural code-witched input. Particularly for nouns from a non-gendered language, the connection between the lexical entry and the syntactic gender node in mixed NP production requires further research (see also Klassen, 2016). Indeed, instances of bilinguals who assign a different gender to the same noun depending on situational context (or other factors) prove the most

challenging, especially for linguistic models of gender representation. One could argue that late insertion approaches, such as Distributed Morphology, may be able to deal with these cases more effectively, but a great deal more theoretical work is required before more solid generalisations can be offered.

4.3 The notion of default

We should also consider more carefully what it means to possess a default strategy, especially in languages with more extensive gender systems, such as those found in Nakh-Daghestanian languages like Tsova-Tush (on defining default, see Corbett & Fraser, 1994). In Spanish, the most studied of the gendered languages in code-switching research, masculine is considered the default gender, because it is used in the absence of gender information about a lexeme. For example a noun lacking modification or an indefinite pronoun such as *nada* 'nothing' receives masculine agreement, as on the adjective in *para ti nada es bueno* 'for you nothing is good.MASC' (Delgado, 2018, p. 44;[10] see also Harris, 1991; Roca, 1989). It is also applied to agreement relating to groups of mixed animates, e.g. *los hombres y las mujeres son listos* 'the.MASC.PL men and the.FEM.PL women are ready.MASC.PL'. Masculine gender is therefore treated as the unmarked gender; Harris (1991) applies a privative feature system, whereby [*f*] denotes 'plus feminine' but its absence denotes masculine gender (see also López, 2020 for discussion). As highlighted by Bellamy and Wichers Schreur (2021), it remains unclear how such a privative system could be applied to a more extensive, non-binary system. For example, all loanwords in Chechen (Nakh-Daghestanian) receive the J gender (named for the form of its agreement target), but it is unclear why it should be singled out as the default, given that it cannot be assigned an absence of a single feature, as is the case for Spanish. It is important that we understand how the default is assigned language-internally, amongst many other factors, not only to facilitate meaningful comparisons between groups, but also to be able to better understand the processes underpinning gender assignment.

10. Delgado (2018) also finds that English nouns belonging to the non-familial sphere are uniformly assigned masculine gender (i.e. a default) in mixed Spanish-English NPs. Frequency likely also contributes to the familiarity of a given lexeme, therefore we follow the suggestion of an anonymous reviewer to also approach the notion of default from a usage-based perspective (see e.g., Hur, López Otero, & Sánchez, 2020 on gender assignment in heritage Spanish speakers).

5. Concluding remarks

Gender assignment in mixed NPs generally proceeds on the basis of three strategies: translation equivalent, shape-based, and default. A bilingual speaker community may use a combination of these strategies, or may overwhelmingly prefer just one, which in the case of many Spanish-English communities is often the default. Currently, however, there is insufficient data from a balanced cross-linguistic sample of language dyads to be able to draw clear conclusions as to which linguistic and extralinguistic factors condition these choices. That said, studies focusing on a single language dyad in various locations using multiple methodologies are highly instructive, since they enable us to begin to tease apart the various factors involved in the processing and production of gender assignment (Beatty-Martínez & Dussias, 2019).

In order to deepen our understanding of gender assignment in mixed NPs, therefore, two key advances are required: (i) a widening of the language dyads under investigation, particularly to include languages with non-binary or ternary gender systems; and (ii) an increase of studies on the same language dyad, ideally employing a 'corpus to cognition' approach, namely building experimental materials from naturalistic data (Beatty-Martínez et al., 2018). Both avenues should also factor in frequency of code-switching in the community as well as order of acquisition of the languages in the dyad, since these factors are emerging as key factors in gender assignment patterns. A more comprehensive understanding of gender assignment would also contribute to the theoretical advancement of code-switching and bilingual grammar more generally.

Acknowledgements

The authors would like to thank three anonymous reviewers for their feedback on this chapter, as well as Dalila Ayoun for her assistance, input and patience.

Funding

Kate Bellamy was supported in the research leading to this chapter by a Marie Skłodowska-Curie Individual Fellowship, grant agreement number 845430. Any omissions or errors remain our own.

The following abbreviations are used in this chapter

D	D gender	FEM	feminine	MASC	masculine
DAT	dative	GEN	genitive	NEUT	neuter
DEF	definite	J	J gender		

References

Aaron, J. E. (2015). Lone English-origin nouns in Spanish: The precedence of community norms. *The International Journal of Bilingualism*, 19(4), 459–480. https://doi.org/10.1177/1367006913516021

Anderson, T. K., & Toribio, A. J. (2017). Attitudes towards lexical borrowing and intra-sentential code-switching among Spanish-English bilinguals. *Spanish in Context*, 4(2), 217–240. https://doi.org/10.1075/sic.4.2.05and

Arndt, W. W. (1970). Nonrandom assignment of loanwords: German noun gender. *Word*, 26(2), 244–253. https://doi.org/10.1080/00437956.1970.11435595

Backus, A. (2015). Codeswitching and language change: One thing leads to another? *International Journal of Bilingualism*, 9(3–4), 307–340. https://doi.org/10.1177/13670069050090030101

Badiola, L., & Sande, A. (2018). Gender assignment in Basque/Spanish mixed determiner phrases: A study of simultaneous bilinguals. In L. López (Ed.), *Code-switching – Experimental answers to theoretical questions: In honor of Kay González-Vilbazo* (pp 15–38). John Benjamins. https://doi.org/10.1075/ihll.19.02bad

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, 405–435. https://doi.org/10.1163/19552629-00903001

Balam, O., Lakshmanan, U., & Parafita Couto, M. C. (2021). Gender assignment strategies among simultaneous Spanish/English bilingual children from Miami, Florida. *Studies in Hispanic and Lusophone Linguistics*, 14: 2–52.

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), 1–11. 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. https://doi.org/10.3390/languages3020019

Bellamy, K., Parafita Couto, M. C., & Stadthagen-González, H. (2018). Investigating gender assignment strategies in mixed Purepecha-Spanish nominal constructions. Special Issue *Interdisciplinary perspectives on code-switching. Languages*, 3(3), 28. https://doi.org/10.3390/languages3030028

Bellamy, K., & Wichers Schreur, J. (2021). When semantics and phonology collide: Gender assignment in mixed Tsova-Tush-Georgian nominal constructions. *International Journal of Bilingualism*.

Bierings, E., Parafita Couto, M. C., & Pedro, M. P. (2019). Contrasting code-switching theories: Insights from Kaqchikel-Spanish code-switched nominal constructions. In *Conference Proceedings, UC Santa Cruz*. Retrieved on 8 June 2021 from https://escholarship.org/uc/item/6m86k80j

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), article 20. https://doi.org/10.3390/languages2040020

Boumans, L. P. C. (1998). The syntax of code-switching: Analysing Moroccan Arabic/Dutch conversation (PhD dissertation). Katholieke Universiteit van Nijmegen, Tilburg University Press.

Budzhak-Jones, S. (1998). Against word-internal codeswitching: Evidence from Ukrainian-English bilingualism. *The International Journal of Bilingualism*, 2(2): 161–182. https://doi.org/10.1177/136700699800200203

Cantone, K. F., & Müller, N. (2008). Un nase or una nase? What gender marking within switched DPs reveals about the architecture of the bilingual language faculty. *Lingua*, 118, 810–826. https://doi.org/10.1016/j.lingua.2007.05.007

Chaston, J. M. (1996). Sociolinguistic analysis of gender agreement in article/noun combinations in Mexican American Spanish in Texas. *Bilingual Review / La Revista Bilingüe*, 21(3), 195–202.

Chirsheva, G. (2009). Gender in Russian–English code-switching. *The International Journal of Bilingualism*, 13(1), 63–90. https://doi.org/10.1177/1367006909103529

Clegg, J. H., & Waltermire, M. (2009). Gender assignment to English-origin nouns in the Spanish of the Southwestern United States. *Southwest Journal of Linguistics*, 28(1), 1–18.

Clyne, M. (1977). Nieuw-Hollands or Double-Dutch. *Dutch Studies*, 3, 1–20.

Clyne, M., & Pauwels, A. (2013). The Dutch language in Australia. In F. Hinskens & J. Taeldeman (Eds.), *Language and space. An international handbook on linguistic variation, Vol. 3: Dutch* (pp. 858–879). Mouton de Gruyter. https://doi.org/10.1515/9783110261332.858

Corbett, G. G. (1991). *Gender*. Cambridge University Press. https://doi.org/10.1017/CBO9781139166119

Corbett, G. G., & Fraser, N. M. (1994). Default genders. In B. Unterbeck & M. Rissanen (Eds.), *Gender in grammar and cognition, I: Approaches to gender* (pp. 55–98). Mouton de Gruyter.

Costa, A., Kovacic, D., Frank, J., & Caramazza, A. (2003). On the autonomy of the grammatical gender systems of the two languages of a bilingual. *Bilingualism: Language and Cognition*, 6(3), 181–200. https://doi.org/10.1017/S1366728903001123

Delgado, R. (2018). The familiar and the strange: Gender assignment in Spanish/English mixed DPs. In L. López (Ed.), *Code-switching – Experimental answers to theoretical questions: In honor of Kay González-Vilbazo* (pp. 39–62). John Benjamins. https://doi.org/10.1075/ihll.19.03del

Denbaum, N., & de Prada Pérez, A. (2020). How do Spanish heritage speakers in the US assign gender to English nouns in Spanish-English code-switching? The effect of noun canonicity and codeswitcher type. *Linguistic Approaches to Bilingualism*. https://doi.org/10.1075/lab.19029.den

Deuchar, M. (forthcoming). Multilingualism in early childhood: The role of input.

Deuchar, M. (2020). Code-switching in linguistics: A position paper. *Languages*, 5(22): https://doi.org/10.3390/languages5020022

Deuchar, M. (2013). The Miami Corpus: Documentation File. Downloaded from http://bangortalk.org.uk/speakers.php?c=miami (14/08/2021).

Deuchar, M. (2012). Code switching. In C. A. Chapelle (ed.) *The encyclopedia of applied linguistics*. Wiley-Blackwell. https://doi.org/10.1002/9781405198431.wbeal0142

Deuchar, M., Davies, P., Herring, J., Parafita Couto, M. C., & Carter, D. (2014). Building bilingual corpora. In E. Thomas & I. Mennen (Eds.), *Advances in the Study of Bilingualism* (pp. 93–110). Multilingual Matters. https://doi.org/10.21832/9781783091713-008

Deuchar, M., & Quay, S. (1999). Language choice in the earliest utterances: A case study with methodological implications. *Journal of Child Language*, 26(2), 461–475. https://doi.org/10.1017/S0305000999003852

DuBord, E. M. (2004). Gender assignment to English words in the Spanish of Southern Arizona. *Divergencias: Revista de Estudios Lingüísticos y Literarios*, 2(2), 27–40.

Eichler, N., Hager, M., & Müller, N. (2012). Code-switching within determiner phrases in bilingual children: French, Italian, Spanish and German. *Zeitschrift für französische Sprache und Literatur*, 122, 227–258.

Ezeizabarrena, M. J. (2009). Development in language mixing: Early Basque-Spanish bilingualism. In J. Grinstead (Ed.), *Hispanic child languages: Typical and impaired development* (pp. 57–89). John Benjamins. https://doi.org/10.1075/lald.50.04eze

Fernández Fuertes, R., Gómez Carrero, T., & Martínez, A. (2020). Where the eye takes you: The processing of gender in codeswitching. *RAEL: Revista Electrónica de Lingüística Aplicada*, 18(1), 1–17.

Fernández Fuertes, R., Liceras, J. M., & Bel, A. (2011). Functional-lexical code-switching in simultaneous and sequential child bilingualism. Paper presented at the International Symposium of Bilingualism (ISB-8), 15–18 June, 2011 University of Oslo, Norway.

Flom, G. T. (1903). The gender of English loan-nouns in Norse dialects in America: A contribution to the study of the development of grammatical gender. *The Journal of English and Germanic Philology*, 5(1), 1–31.

Franceschina, F. (2001). Morphological or syntactic deficits in near-native speakers? An assessment of some current proposals. *Second Language Research*, 17(3), 213–247. https://doi.org/10.1177/026765830101700301

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

Gardner-Chloros, P. (2009). *Code-switching*. Cambridge University Press. https://doi.org/10.1017/CBO9780511609787

Gómez Carrero, T. (2015). Un pencil or una pencil? The importance of Spanish gender in switched DPs (Unpublished Master's thesis). University of Valladolid.

González-Vilbazo, K. 2005. Die Syntax des Code-Switching (Unpublished PhD dissertation). Universitu of Cologne.

Grosjean, F. (2001). The bilingual's language modes. In J. Nicol (Ed.), *One mind, two languages: Bilingual language processing* (pp. 1–22). Oxford: Blackwell.

Gullberg, M., Indefrey, P., & Muysken, P. (2009). Research techniques for the study of code-switching. In B. Bullock & J. Toribio (Eds.), *The Cambridge handbook of linguistic code-switching* (pp. 21–39). Cambridge University Press. https://doi.org/10.1017/CBO9780511576331.003

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

Haugen, E. (1969). *The Norwegian language in America: A study in bilingual behaviour*. Indiana University Press.

Herring, J., Deuchar, M., Parafita Couto, M. C., & Moro Quintanilla, M. (2010). "I saw the madre": Evaluating predictions about codeswitched determiner-noun sequences using Spanish-English and Welsh-English data. *International Journal of Bilingual Education and Bilingualism*, 13(5), 553–573. https://doi.org/10.1080/13670050.2010.488286

Hockett, C. F. (1958). *A course in modern linguistics*. New York: MacMillan Company.

Hur, E., López Otero, J. C., & Sánchez, L. (2020). Gender agreement and assignment in Spanish heritage speakers: Does frequency matter? *Languages*, 5(4), 48. https://doi.org/10.3390/languages5040048

Ibrahim, M. H. (1973). *Grammatical gender: Its origin and development*. Mouton. https://doi.org/10.1515/9783110905397

Iriondo, A. (2017). Kode-alternantzia elebidun gazteetan: DS nahasiak [Code-switching in young bilinguals: Mixed DPs]. *Gogoa*, 16, 25–56. https://doi.org/10.1387/gogoa.18646

Jescheniak, J. D., Schriefers, H., & Lemhöfer, K. (2014). Selection of freestanding and bound gender-marking morphemes in speech production: A review. *Language, Cognition and Neuroscience*, 29(6), 684–694. https://doi.org/10.1080/01690965.2012.654645

Jorschick, L., Endesfelder Quick, A., Glässer, D., Lieven, E., & Tomasello, M. (2011). German-English-speaking children's mixed NPs with 'correct' agreement. *Bilingualism: Language and Cognition*, 14(2), 173–183. https://doi.org/10.1017/S1366728910000131

Klassen, R. (2016). The representation of asymmetric grammatical gender systems in the bilingual mental lexicon. *Probus*, 28(1), 9–28.

Klassen, R., & Liceras, J. M. (2017). The representation of gender in the mind of Spanish-English bilinguals: Insights from code-switched adjectival predicates. *Borealis: An International Journal of Hispanic Linguistics*, 6(1), 77–96. https://doi.org/10.7557/1.6.1.4100

Królikowska, M. A., Bierings, E., Beatty-Martínez, A. L., Navarro-Torres, C., Dussias, P. E., & Parafita Couto, M. C. (2019). Gender assignment strategies within the bilingual determiner phrase: Four Spanish-English communities examined. Paper presented at the 3rd Conference on Bilingualism in the Hispanic and Lusophone World (BHL), Leiden, The Netherlands, 9–11 January.

Lanza, E. (1997). *Language mixing in infant bilingualism: A sociolinguistic perspective*. Oxford University Press.

Leisiő, L. (2001). Integration and gender assignment in Finland Russian. *SKY Journal of Linguistics*, 14, 87–119.

Liceras, J. M., Fernández Fuertes, R., & Klassen, R. (2016). Language dominance and language nativeness: The view from English-Spanish codeswitching. In R. E. Guzzardo Tamargo, C. M. Mazak, & M. C. Parafita Couto (Eds.), *Spanish-English Codeswitching in the Caribbean and the US* (pp. 107–138). John Benjamins. https://doi.org/10.1075/ihll.11.05lic

Liceras, J. M., Fernández, R., Bel, A. & Martínez, C. (2012), The mental representation of Gender and Agreement Features in Child 2L1 and Child L2 grammars: Insights from code-switching. Presentation at UIC Bilingualism Forum, 04/10/2012. Chicago: University of Illinois.

Liceras, J. M., Fernández Fuertes, R., Perales, S., Peréz-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, 827–851. https://doi.org/10.1016/j.lingua.2007.05.006

López, L. (2020). *Bilingual grammar: Toward an integrated model*. Cambridge University Press. https://doi.org/10.1017/9781108756181

MacSwan, J. (2000). The architecture of the bilingual language faculty: Evidence from code-switching. *Bilingualism: Language and Cognition*, 3(1), 37–54. https://doi.org/10.1017/S1366728900000122

MacSwan, J. (2021). Theoretical approaches to the grammar of code-switching. In E. Adamou & Y. Matras (Eds), *The Routledge handbook of language contact* (pp. 88–109). Routledge.

MacWhinney, B. (2000). The CHILDES project: Tools for analyzing talk. *Child Language Teaching and Therapy*, 8(2), 217–218. https://doi.org/10.1177/026565909200800211

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

Moro Quintanilla, M. (2014). The semantic interpretation and syntactic distribution of determiner phrases in Spanish-English codeswitching. In J. MacSwan (Ed.), *Grammatical theory and bilingual codeswitching* (pp. 213–226). The MIT Press.

Munarriz-Ibarrola, A., Ezeizabarrena, M. J., de Castro-Arrazola, V., & Parafita Couto, M. C. (2021). Gender assignment strategies and L1 effects in the elicited production of mixed Spanish-Basque DPs. *Linguistic Approaches to Bilingualism*. Online first. https://doi.org/10.1075/lab.20016.mun.

Muysken, P. C. (2000). *Bilingual speech: A typology of code-mixing*. Cambridge University Press.

Myers-Scotton, C. (2002). *Contact linguistics: Bilingual encounters and grammatical outcomes*. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780198299530.001.0001

Myers-Scotton, C., & Jake, J. L. (2000). Testing the 4-M model: An introduction. *International Journal of Bilingualism*, 4(1), 1–8. https://doi.org/10.1177/13670069000040010101

Otheguy, R., & Lapidus Shin, N. (2003). An adaptive approach to noun gender in New York contact Spanish. In R. Núñez-Cedeño, L. López, & R. Cameron (Eds.), *A Romance perspective on language knowledge and use: Selected papers from the 31st Linguistic Symposium on Romance Languages (LSRL), Chicago, 19–22 April 2001* (pp. 209–229). John Benjamins. https://doi.org/10.1075/cilt.238.17oth

Pablos, L., Parafita Couto, M. C., Boutonnet, B., De Jong, A., Perquin, M., De Haan, A., & N. O. Schiller. (2019). Adjective-noun order in Papiamento-Dutch code-switching. *Linguistic Approaches to Bilingualism*, 9(4), 710–735. https://doi.org/10.1075/lab.17036.pab

Parafita Couto, M. C., Davies, P., Carter, D., & Deuchar, M. (2014). Factors influencing code-switching. In E. Thomas & I. Mennen (Eds.), *Advances in the study of bilingualism* (pp. 111–140). Multilingual Matters. https://doi.org/10.21832/9781783091713-009

Parafita Couto, M. C., Deuchar, M., & Fusser, M. (2015). How do Welsh-English bilinguals deal with conflict? Adjective-noun order resolution. In G. Stell & K. Yakpo (Eds.), *Code-switching at the crossroads between structural and sociolinguistic perspectives* (pp. 65–84). Mouton de Gruyter. https://doi.org/10.1515/9783110346879.65

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. https://doi.org/10.1177/1367006917729543

Parafita Couto, M. C., Munarriz, A., Epelde, I., Deuchar, M., & Oyharçabal, B. (2015). Gender conflict resolution in Spanish-Basque mixed DPs. *Bilingualism: Language and Cognition*, 18(2), 304–323. https://doi.org/10.1017/S136672891400011X

Parafita Couto, M. C., & Stadthagen-González, H. (2017). El book or the libro? Insights from acceptability judgements into determiner/noun code-switches. *The International Journal of Bilingualism*, 23(1), 349–360. https://doi.org/10.1177/1367006917728392

Pearson, B. Z. (2002). Narrative competence among monolingual and bilingual school children in Miami. In D. K. Oller & R. E. Eilers (Eds), *Language and literacy in bilingual children* (pp. 135–174). Multilingual Matters. https://doi.org/10.21832/9781853595721-008

Pensalfini, R., & Meakins, F. (2019). Gender lender: Noun borrowings between Jingulu and Mudburra in Northern Australia. *Journal of Language Contact*, 12, 440–478. https://doi.org/10.1163/19552629-01202007

Poplack, S. (1980). Sometimes I'll start a sentence in Spanish, y termino en español. Towards a typology of code-switching. *Linguistics*, 18(7/8), 581–618. https://doi.org/10.1515/ling.1980.18.7-8.581

Poplack, S. (2018). *Borrowing: Loanwords in the speech community and in the grammar*. Oxford University Press.

Poplack, S., & Meechan, M. (1998). How languages fit together in codemixing. *International Journal of Bilingualism*, 2(2), 127–138. https://doi.org/10.1177/136700699800200201

Poplack, S., Pousada, A., & Sankoff, D. (1982). Competing influences on gender assignment: Variable process, stable outcome. *Lingua*, 57, 1–28. https://doi.org/10.1016/0024-3841(82)90068-7

Radford, A., Kupisch, T., Köppe, R., & Azzaro, G. (2007). Concord, convergence and accommodation in bilingual children. *Bilingualism: Language and Cognition*, 10(3), 239–256. https://doi.org/10.1017/S1366728907003057

Rand, D., & Sankoff, D. (1990). *GoldVarb. A variable rule application for the Macintosh*. Université de Montréal, Centre de recherches mathématiques.

Reed, C. E. (1942). The gender of English loan words in Pennsylvania German. *American Speech*, 17(1), 25–29. https://doi.org/10.2307/486853

Roca, I. M. (1989). The organisation of grammatical gender. *Transactions of the Philological Society*, 87(1), 1–32. https://doi.org/10.1111/j.1467-968X.1989.tb00617.x

Salamoura, A., & Williams, J. N. (2007). The representation of grammatical gender in the bilingual lexicon: Evidence from Greek and German. *Bilingualism: Language and Cognition*, 10(3), 257–275. https://doi.org/10.1017/S1366728907003069

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 judgements for code-switching research. *Linguistic Approaches to Bilingualism*, 8, 67–97. https://doi.org/10.1075/lab.16030.sta

Stolz, C. (2008). Loan word gender: A case of romancisation in Standard German and related enclave varieties. In T. Stolz, D. Bakker, & R. S. Palomo (Eds.), *Aspects of language contact: New theoretical, methodological and empirical findings with special focus on Romancisation processes* (pp. 399–440). Mouton de Gruyter.

Stolz, C. (2009). A different kind of gender problem: Maltese loanword gender from a typological perspective. In B. Comrie, R. Fabri, E. Hume, M. Mifsud, T. Stolz & M. Vanhove (Eds.), *Introducing Maltese linguistics: Selected papers from the 1st International Conference on Maltese Linguistics, Bremen, 18–20 October, 2007* (pp. 321–353). John Benjamins. https://doi.org/10.1075/slcs.113.22sto

Toribio, A. J. (2017). Structural approaches to code-switching: Research then and now. In R. E. V. Lopes, J. Ornelas de Avelar, & S. M. L. Cyrino (Eds.), *Romance languages and linguistic theory 12. Selected papers from the 45th Linguistic Symposium on Romance Languages (LSRL), Campinas, Brazil* (pp. 213–233). John Benjamins. https://doi.org/10.1075/rllt.12.14tor

Torres Cacoullos, R., & Travis, C. E. (2015). Gauging convergence on the ground: Code-switching in the community. *International Journal of Bilingualism*, 19(4): 365–386. https://doi.org/10.1177/1367006913516046

Torres Cacoullos, R., & Travis, C. E. (in preparation). New Mexico Spanish-English bilingual (NMSEB) corpus, National Science Foundation 1019112/1019122.

Treffers-Daller, J. (1993). *Mixing two languages: French-Dutch contact in a comparative perspective*. Mouton de Gruyter.

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.), *Spanish-English code-switching in the Caribbean and the US* (pp. 281–300). John Benjamins. https://doi.org/10.1075/ihll.11.12val

Valdés Kroff, J. R. (2012). Using eye-tracking to study auditory comprehension in codeswitching: Evidence for the link between comprehension and production (Unpublished PhD dissertation). The Pennsylvania State University.

Valdés Kroff, J. R., Dussias, P. E., Gerfen, C., Perrotti, L., & Bajo, M. T. (2017). Experience with code-switching modulates the use of grammatical gender during sentence processing. *Linguistic Approaches to Bilingualism*, 7(2), 163–198. https://doi.org/10.1075/lab.15010.val

Valenzuela, E., Faure, A., Ramírez-Trujillo, A. P., Barski, E., Pangtay, Y., & Diez, A. (2012). Gender and heritage Spanish bilingual grammars: A study of code-mixed determiner phrases and copula constructions. *Hispania*, 95(3), 481–494.

Van Osch, B., Boers, I., Grijzenhout, J., Parafita Couto, M. C., Sterken, B., & Tat, D. (2022). Cross-linguistic influence in bilingual grammars: Evidence from gender assignment in unilingual Dutch and mixed speech. In D. Ayoun (Ed.), *The acquisition of Gender. Cross-linguistic perspectives*. John Benjamins. (this volume) https://doi.org/10.1075/sibil.63.09van

Vanden Wyngaerd, E. (2021). Bilingual implications: Using code-switching data to inform linguistic theory (Unpublished PhD dissertation). Université Libre de Bruxelles.

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. https://doi.org/10.3389/fpsyg.2020.549762

Violin-Wigent, A. (2006). Gender assignment to nouns codeswitched into French: Observations and explanations. *International Journal of Bilingualism*, 10(3), 253–276. https://doi.org/10.1177/13670069060100030101

Wang, M., & Schiller, N. O. (2019). A review on grammatical gender agreement in speech production. *Frontiers in Psychology*, 9(2754), 1–7. https://doi.org/10.3389/fpsyg.2018.02754

Weinreich, U. (1953). *Languages in contact*. Mouton.

Zabrodskaya, A. (2009). Evaluating the Matrix Language Frame model on the basis of a Russian–Estonian codeswitching corpus. *The International Journal of Bilingualism*, 13(3), 357–377. https://doi.org/10.1177/1367006908346629

Appendix 1. Overview of the studies reported in Section 2, their main gender assignment strategy, language pair(s) and their respective number of genders (in brackets), data collection type and age group studied

Gender assignment strategy	Reported in	Language pair (no. of genders)	Data collection method	Adults or children
	– Bellamy & Wichers Schreur (2021)	– Tsova-Tush (5)-Georgian (0)	– Guided production	– Adults
	– Munarriz Ibarrola et al. (2021)	– Basque (0)- Spanish (2)	– Guided production	– Adults
	– Iriondo (2017)	– Basque (0)- Spanish (2)	– AJT (written)	– Adults
	– Klassen & Liceras (2017)	– Spanish (2)-English (0)	– AJT (written)	– Adults
	– Jorschick et al. (2011)	– German (3)-English (0)	– Naturalistic production	– Children
	– Gómez Carrero (2015)	– English (0)-Spanish (2)	– AJT (written), guided production	– Adults
Translation equivalent	– Zabrodskaya (2009)	– Russian (3)-Estonian (0)	– Naturalistic production	– Adults
	– Cantone & Müller (2008)	– German (3)-Italian/French/ Spanish (2)	– Naturalistic production	– Children
	– Liceras et al. (2008)	– Spanish (2)-English (0)	– AJT	– Both
	– Radford et al. (2007)	– Italian (2)-English (0)	– Naturalistic production	– Children
	– González-Vilbazo (2005)	– Spanish (2)-German (3)	– Naturalistic production	– Adults
	– Füller & Lehnert (2000)	– German (3)-English (0)	– Naturalistic production	– Adults
	– Treffers-Daller (1993)	– French (2)-Belgian Dutch (3)	– Naturalistic and guided production	– Adults
Phonological analogy	– Bellamy & Wichers Schreur (u.r.)	– Tsova-Tush (5)-Georgian (0)	– AJT (written and spoken), guided production	– Adults
	– Munarriz-Ibarrola et al. (in press)	– Basque (0)-Spanish (2)	– Guided production	– Adults
	– Badiola & Sande (2018)	– Basque (0)-Spanish (2)	– AJT (written)	– Adults
	– Parafita Couto et al. (2015)	– Basque (0)-Spanish (2)	– Naturalistic and guided production, AJT	– Adults
	– Chirsheva (2009)	– Russian (3)-English (0)	– Naturalistic production and observation	– Adults
	– Zabrodskaya (2009)	– Russian (3)-English (0)	– Naturalistic production	– Adults

Gender assignment strategy	Reported in	Language pair (no. of genders)	Data collection method	Adults or children
	– Balam, Lakshmanan & Parafita Couto (2021)	– Spanish (2)-English (0)	– Naturalistic production	– Children
	– Bellamy, Parafita Couto & Stadthagen-González (2018)	– Spanish (2)-Purepecha (0)	– Directed production	– Adults
	– Delgado (2018)	– Spanish (2)-English (0)	– AJT	– Adults
	– Badiola & Sande (2018)	– Basque (0)-Spanish (2)	– AJT	– Adults
	– Balam (2016)	– Spanish (2)-English (0)	– Naturalistic production	– Adults
	– Valdés Kroff (2016)	– Spanish (2)-English (0)	– Naturalistic production	– Adults
	– Aaron (2015)	– Spanish (2)-English (0)	– Naturalistic production	– Adults
	– Parafita Couto et al. (2015)	– Basque (0)-Spanish (2)	– Naturalistic and guided production, AJT	– Adults
	– Fernández Fuertes et al. (2011)	– Spanish (2)-English (0)	– AJT	– Adults and children
	– Chirsheva (2009)	– American Norwegian (3), American Lithuanian (2), American Portuguese (2), American Italian(2)-English(0)	– Naturalistic production (?)	– Adults
Default				
	– Clegg & Watermire (2009)	– Spanish (2)-English (0)	– Naturalistic production	– Adults
	– Zabrodskaya (2009)	– Russian (3)-English (0)	– Naturalistic production	– Adults
	– Liceras et al. (2008)	– Spanish (2)-English (0)	– Naturalistic production, AJT	– Adults and children
	– DuBord (2004)	– Spanish (2)-English (0)	– Naturalistic production	– Adults
	– Otheguy & Lapidus (2003)	– Spanish (2)-English (0)	– Naturalistic production	– Adults
	– Franceschina (2001)	– English (0)-Spanish (2)	– Naturalistic production	– Adult
	– Leisiö (2001)	– Russian (3)-Finnish (0)	– Unclear	– Adults
	– Budzhak-Jones (1998)	– Ukrainian (3)-English (0)	– Naturalistic production, observation	– Adults
	– Chaston (1996)	– Spanish (2)-English (0)	– Naturalistic production	– Adults
	– Clyne (1977)	– Dutch (2)-English (0)	– Naturalistic production	– Adults
	– Weinreich (1953)	– American German/American Yiddish (3)-English (0)	– Production, observation	– Adults

Note that the same study can appear in more than one strategy. The reader is advised to consult the text for an interpretation of the results summarised here.

CHAPTER 3

Empirical evidence for subtle gender biases in language

Yulia Esaulova and Lisa von Stockhausen
University of Cologne / University of Duisburg-Essen

This chapter presents empirical evidence from social psychological and psycholinguistic studies to provide an insight on mechanisms underlying gender inequality represented and maintained through language. The reviewed phenomena cover linguistic manifestations of gender and gender stereotypes that range from relatively superficial word semantics to gender asymmetries ingrained deeply in grammatical language systems, such as the category of verbs and the case system. The existing studies on biases in language suggest that even linguistic phenomena seemingly unrelated to gender (e.g. negations or thematic structure) may function as mechanisms supporting gender stereotypical beliefs. Such biases present a special challenge when it comes to identifying them; however, revealing their mechanisms has important implications for initiatives aiming at neutralizing their negative social consequences.

Keywords: gender, bias, gender-fair language, stereotype

1. Empirical evidence for gender biases in language

Language does not only **allow** for communication – it is a formal system that imposes its grammar rules on the users, often leaving them no choice but to comply such as the grammatical category of gender that may force one to use masculine and feminine forms (e.g. masculine or feminine pronouns) when referring to people. Such mandatory linguistic choices not only reflect, but also shape cognitive representations of sexes. This idea is reflected, for instance, in Slobin's (2003) theoretical approach, which assumes that the encoding options a language provides oblige its speakers to attend to certain referent characteristics as significant for communication. To use Slobin's (2003) terms, speakers' "thinking for speaking" and "listening for thinking" is thus tuned to gender via the grammatical encoding of gender in their language. Stereotypical thinking and expectations about genders

https://doi.org/10.1075/sibil.63.03esa

maintained through language often contributes to gender disparity, where women are disadvantaged in terms of educational, economic, political, and health-related opportunities (Hellinger & Bußmann, 2001, 2002, 2003; Hodel et al., 2017; Maass & Arcuri, 1996; Sczesny, Formanowicz, & Moser, 2016). Gender-fair language and other equal opportunity policies introduced in recent years aim to reduce the existing gender gap in Europe and around the world (Hellinger & Bierbach, 1993; Bundesministerium für Frauen und Öffentlichen Dienst, 2009; Valdrová, 2013).

Despite substantial changes, continued debating and re-evaluating of strategies (e.g. Gabriel, Gygax, & Kuhn, 2018), however, the issue is far from solved (Schwab et al., 2017). The versatility of mechanisms involved in maintaining gender hierarchies and stereotypes on the linguistic level may be part of the reason why this is the case. One factor contributing to this versatility consists in the variability of languages with respect to how the gender category is realized (Stahlberg et al., 2007). In so-called grammatical gender languages (e.g. German, French), gender markings may be present not only on nouns but also on other parts of speech. In contrast, in natural gender languages (e.g. English), it is mostly personal pronouns that are affected, and in genderless languages (e.g. Finnish) gender is marked only lexically, using specifying words such as female/male or lexical gender nouns (e.g. mother) (see *Language Index of Grammatical Gender Dimensions* by Gygax et al., 2019 for more details). Another factor comes from linguistic preferences or tendencies that language users have that may or may not seem to be directly related to gender but nevertheless support gender biases. Lexical preferences, such as the use of a particular word form (e.g. the generic use of masculine forms, as described below), are quite explicit, as they directly relate to gender on the level of grammar via gender markings or agreement. On the other hand, structural choices on the sentence level, such as the use of a non-canonical word order, are rather implicit, as they do not necessarily relate to gender on the grammatical level. As a result, the fact that they may relate to gender hierarchies often goes unnoticed. This chapter brings together psycholinguistic evidence about both explicit and implicit tendencies in an attempt to contribute to the understanding of mechanisms underlying gender biases in language.

First, we present well-documented evidence that specific word choices (e.g. the use of abstract vs. concrete words, the generic vs. specific use of masculine forms) and common preferences in phrasings (e.g. the use of negation and irony) may be indicative of stereotypical beliefs about gender. Next, we turn to aspects that have received less attention and discuss evidence that gender biases may be supported by rather fundamental properties of language, such as the grammatical category of verbs and the case system. Finally, we examine recent findings that language users make predictions regarding thematic structures of sentences based on the available gender cues in order to resolve ambiguous linguistic input.

The goal of this chapter is to provide an overview (rather than an exhaustive review) of research on gender biases in language that identifies well-studied areas, as well as those that need to receive more attention in the future. Our objective is to demonstrate the versatility of forms in which gender biases exist in language. We therefore focus on empirical observations of gender biases in language without a further analysis of causal relationships between language and gender biases. While some of the gender biases described below cannot be easily avoided, bringing awareness to their – sometimes rather subtle – nature may be the first step towards making gender equality initiatives more efficient.

2. Linguistic phenomena representing biases

2.1 Linguistic intergroup and expectancy bias

The diversity of linguistic means available in a natural language allows its users not only to express an infinite number of ideas, but also to formulate the same idea in a number of ways by making different structural and word choices. While these choices may seem effortless and random, there is evidence for certain regularities behind them. The extent to which a given formulation is abstract or concrete, for instance, corresponds to whether the described event is perceived as either stereotypical – including but not restricted to gender-related stereotypes – or rather unconventional. Semin and Fiedler's (1988, 1992) linguistic category model identifies four levels of abstraction, with adjectives being the most abstract, followed by state verbs (e.g. to hate) and interpretive action verbs (e.g. to hurt), to the most concrete descriptive action verbs (e.g. to hit). The description of an event may be accurate irrespective of its abstraction level; however, the implications of the chosen abstract or concrete categories in its formulation may be critically different. An abstract formulation would imply that the event represents an expected and typical case (e.g. 'Mary loves cooking'), whereas a concrete one would suggest that it is unexpected and unique (e.g. 'John is cooking lunch'). This phenomenon was first established in a series of experiments by Maass et al. (1989) as the linguistic intergroup bias. Members of competing groups in a horse-riding competition were asked to participate in a language comprehension (Experiment 1) and a language production (Experiment 2) task. In Experiment 1, participants saw cartoons representing members of their own or the opposite group in socially desirable or undesirable situations. After each cartoon, they were provided with four descriptions that were constructed according to the four abstraction categories (Semin & Fiedler, 1988) and had to choose the best-fitting one. In Experiment 2, Maas et al. asked participants to describe the cartoons in their own words and then coded

the responses for language abstraction. The results of both experiments revealed that participants preferred abstract descriptions of positive events and concrete descriptions of negative events involving in-group members, whereas the opposite preferences were displayed for scenes involving out-group members. The authors describe the observed bias as a linguistic mechanism used to maintain and transmit positive and negative perceptions of social groups.

Fiedler, Semin, and Finkenauer (1993) adopted the perspective of language-based biases to address the discriminatory behavior with respect to gender groups. The authors analyzed free descriptions on various topics elicited from female and male respondents and found that derogative statements about the out-group were more abstract than derogative statements about the in-group. Describing the behavior in abstract as opposed to concrete terms promotes its perception as a typical characteristic of the actor and can thus be used strategically to encourage positive or negative beliefs about a group. Similar experimental paradigms were used to determine the linguistic intergroup bias in other social groups as well, such as sports teams, nations, organizations, and various political and interest groups (Arcuri, Maass, & Portelli, 1993; Bencivenni, 1990; Rubini & Semin, 1994; for an overview see Giles & Maass, 2016).

This bias also manifests itself in the content analyses of media reports (Maass, Corvino, & Arcuri, 1994), as well as subsequent experiments on the perception of such reports (Geschke et al. 2010), showing that the way information is transmitted may add implicit semantic values and contribute to stereotyping and prejudice. In this way, the bias can be seen as an instantiation of generic statements (e.g. 'women are submissive') expressing generalizations that in turn foster essentialist beliefs. In our example, the submissiveness of women is phrased and interpreted as attributable to the very nature of women rather than to changeable extrinsic circumstances (for the analysis of generics and a related discussion, see Haslanger, 2011; Leslie, 2014; McConnell-Ginet, 2012). Gender-related social norms expressed via generic statements (e.g. 'men don't cry') have been shown to communicate and sustain stereotypes in a concise and efficient manner affecting both adults and children (Cimpian & Markman, 2009, 2011; Gelman, Ware, & Kleinberg, 2010; Leslie, 2015; Rhodes, Leslie, & Tworek, 2012).

Similarly, the linguistic expectancy bias posited by Wigboldus, Semin, and Spears (2000) claims that expectancy-consistent information tends to be communicated at a higher level of abstraction than expectancy-inconsistent information. When the behavior of a group member is described in rather abstract terms (e.g. 'Mary is emotional'), it is likely to be perceived as (stereotype-)consistent and recurrent. When, on the other hand, more concrete terms are used (e.g. 'John is crying'), the behavior is perceived as an exception to what is generally expected. Both the linguistic intergroup bias and the expectancy bias, which can be seen as

the generalization and extension of the former, indicate that language users are sensitive to different levels of abstraction as a cue reflecting stereotypical expectations and beliefs.

2.2 Negation bias

Linguistic means allowing for implicit biases in language are certainly not confined to abstractions or generic statements. Another mechanism that functions in a similar way is negation that may as well be used to signal inconsistent behaviors. It is widely reported that processing costs are higher for negations compared to affirming sentences (e.g. Carpenter & Just, 1975; Deutsch, Gawronski, & Strack, 2006), but this association appears to be contextually dependent (e.g. Glenberg et al., 1999). In fact, the difference in processing times disappears when the negation denies an assumed expectation or communicates information that is unusual or inconsistent with common beliefs (Villiers & Tager Flusberg, 1975; Wason, 1965). This is in line with Jordan's (1998) view of negations as not merely semantic opposites of the equivalent assertive statements, but rather as markers signaling and denying the inferred expectations.

In one of the experiments on negation bias carried out by Beukeboom, Finkenauer, and Wigboldus (2010), participants read descriptions of a person's behavior (e.g. 'The girl comforts the sales woman') and were asked to rate on a scale the applicability of the two statements that followed: a negation (e.g. 'She is not aggressive') and an affirmation (e.g. 'She is sweet'). The statements containing negations were rated as more applicable when they referred to stereotype-inconsistent descriptions (e.g. 'The girl curses at the sales woman') than stereotype-consistent descriptions.

In another study (Beukeboom et al., 2010), the authors addressed gender-stereotypical expectations concerning sports. Participants were presented with photos of people performing stereotypically female (e.g. jazz ballet) or male (e.g. rugby) sports, as well as photos of a man or a woman who were said to practice that sport. First, participants described what they thought was characteristic of a sport and people who practiced it, which activated the existing stereotypes. Then, participants described their impressions and expectations regarding the performance of the man or the woman in the photo. The number of used negation markers, such as *not* and *no*, was higher in descriptions of men and women practicing a sport that contradicted rather than corresponded to gender stereotypes about it.

Beukeboom et al.'s (2010) findings provide evidence that negations are used and perceived as indicators of information that is inconsistent with existing expectations. Similarly to the linguistic intergroup and expectancy bias, the negation bias is a phenomenon that is not restricted to gender-related expectations, but can

be observed in a variety of contexts that are prone to stereotyping (e.g. behaviors associated with certain professions or social groups). As such, the negation bias reflects and induces inferences about the described behaviors or people and thus may contribute to their stereotypical representations. Whether the negation bias affects language comprehension to a similar extent as it affects language production still remains to be clarified by future research.

2.3 Irony bias

The use of irony is yet another linguistic mechanism that systematically communicates whether information is consistent or inconsistent with what is generally expected. The functions of irony vary from conveying humor to being polite to criticizing, sometimes in quite an aggressive manner known as sarcasm (e.g. Gibbs, 2000). A crucial characteristic of ironic statements is that they have a double meaning: a literal one and an intended one, which is contrary to the literal meaning. Ironic comments can only be interpreted in context and they should be both literally inappropriate and relevant to the given context (Attardo, 2000).

The literal interpretation of 'thank you', for instance, is inappropriate when someone was rude to you and at the same time it is relevant as a response, since its reproachful intention is related to the violated politeness, which is perceived as a norm by both the recipient and the speaker. Implicit reference to such a shared knowledge in ironic statements makes irony a tool that potentially reinforces existing norms, expectations, stereotypes, and the like.

Because of the similarities in the processing of irony and negations (Giora et al., 2007; Mayo, Schul, & Burnstein, 2004), some consider irony an "indirect negation", since the literal evaluation in irony does not need markers such as "not" or "no" to be negated (e.g. Giora, Fein, & Schwartz, 1998). Burgers and Beukeboom (2016) studied whether irony – similarly to negation – may function as a bias promoting stereotypical expectations in language production and perception. In their production experiments, participants were presented with the situation of either a positive or a negative behavior (e.g. failing vs. passing an exam). Depending on the condition, either a stereotype-consistent actor (e.g. a blonde in case of a failed exam – conforming to the existing gender stereotypes) or a stereotype-inconsistent one (e.g. a geek) was involved. Then participants had to choose whether comments about the situation were literal or ironic (Experiment 1) and evaluate the appropriateness of ironic comments (Experiments 2 and 3). The results showed that participants were more likely to categorize comments as ironic in stereotype-inconsistent situations, and they also rated irony as more appropriate in those situations than in stereotype-consistent ones.

In the language perception experiment, participants had to rate ironic and literal comments based on three criteria: repetition likelihood (i.e. how likely the actor is to repeat the behavior in the future), dispositionality (to what extent the behavior is typical for the actor), and generalizability (the stability of behavior across contexts). These criteria reflected essentialist beliefs indicating the extent to which actors are perceived as fundamentally similar to other members of their social category and across situations (Carnaghi et al., 2008; Rhodes, Leslie, & Tworek, 2012).

Compared to literal comments, ironic ones were rated lower on all three criteria, indicating that irony reduced the perceived essentialism. Irony thus can be seen as a linguistic marker of the uniqueness of an event or its actor and therefore may function like other linguistic biases, indicating situations and behaviors that are inconsistent with stereotypes about them.

Just like the aforementioned biases, the preference of ironic comments to literal ones can be observed in a variety of situations associated with expectations and beliefs that are not exclusively related to but certainly include gender stereotypes. Since irony has been regarded as a bias relatively recently, contexts related to gender in which the preference for irony surfaces need to be further differentiated. Future research may want to determine, for instance, which factors may predict the strength of the preference in gender-related contexts and how likely it is to be directed at oneself rather than others (e.g. speakers' ability to dissociate themselves from a stereotype when it corresponds to their own sex, age, profession, etc.).

2.4 Male bias in "generic" masculine and "gender-fair" alternatives

Personal denotations used in language to refer to men and women have been considered as another linguistic source conveying expectations about gender groups. In grammatical gender languages, grammatically masculine word forms (e.g. *Musiker*.MASC vs. *Musikerin*.FEM 'musician' in German) often correspond to the biological sex of referents. However, this correspondence is not always perfect, since grammatically masculine role nouns are sometimes used as generic to refer to both men and women (i.e. *Musiker*.MASC to designate both female and male musicians), while grammatically feminine forms are always used in a gender-specific way to refer to women (i.e. *Musikerin*.FEM can only be used to designate a female musician).

This asymmetry has received attention from both social psychological and psycholinguistic research, which showed that masculine forms – even if meant as generic – may still lead to a gender-specific interpretation and thus bias recipients' perception and behavior. Following earlier findings by Bem and Bem (1973) that stereotypically phrased job advertisements prevented applicants from considering jobs with an opposite-sex bias despite their qualifications, later studies revealed that masculine wordings in job advertisements not only reduced female students'

identification with those jobs and their motivation to pursue them as careers (Stout & Dasgupta, 2011), but also caused others to perceive those jobs as less suitable for women than for men (Horvath & Sczesny, 2016). Furthermore, sensitivity to such gender biases surfaces as early as primary school, so that occupations presented in masculine forms (e.g. *Sportler* 'athletes (masculine)') are perceived by children as less accessible, less interesting, and promising lower success to women compared to respective paired forms (e.g. *Sportlerinnen*.FEM *und Sportler*.MASC 'athletes') (Vervecken, Hannover, & Wolter, 2013).

Using masculine, but not feminine, forms as generic is considered to reflect an androcentric perspective, defined as the conflation of maleness with the norm, taking males as the default and attributing gender differences to females (Bem, 1993). This perspective can also be found in subject-complement structures, such as 'Girls are as good as boys at math'. The item in the complement position ('boys') is considered a reference point expressing the norm, or what is typical, and the subject ('girls') implies the difference to the norm. This example shows how even expressions of gender equality may defeat their purpose (Chestnut & Markman, 2018).

Another case in point are paired forms which themselves have been shown to express and shape gender stereotypes depending on word order. The first mentioned gender is perceived as having higher relevance in a context (Kesebir, 2017). These findings suggest that even linguistic forms that are considered gender-fair may promote gender biases. The use of masculine forms as generic (and its gender-specific interpretation) is one of the most well-studied among gender biases in language and has been well documented. In the following section we will cover some of the extensive literature related to the topic, keeping it concise to bring more focus on gender biases in language that have so far received considerably less attention.

2.5 Gender mismatch effects as indicators of biases

Gender mismatch effects reported in psycholinguistic studies signal and confirm the existence of gender biases on the cognitive level of language processing. These effects display processing costs that manifest in behavioral and electro-physiological measures, such as slowed reaction and reading times or enhanced event-related potentials, when the given gender cues are inconsistent with gender stereotypes.

Role nouns that represent groups of people often contain knowledge, beliefs, and expectancies about these groups – i.e., cognitive structures known as stereotypes (Hamilton & Trolier, 1986). Stereotypical gender of role nouns is often assessed using questionnaires (e.g. Kennison & Trofe, 2003; Misersky et al., 2014; Siyanova-Chanturia, Pesciarelli, & Cacciari, 2012), which identify role nouns as being more often assigned to women (e.g. beautician), to men (e.g. footballer),

or being neutral in this respect (e.g. student). When the stereotypical gender of a role noun does not correspond to the pronoun referring to it, as in 'The footballer wanted to play in the match. She had been training very hard during the week' from Carreiras et al.'s (1996) self-paced reading experiment, reading times are inflated, indicating gender mismatch effects.

These effects are explained in terms of a mental model (Garnham, 1987), which includes the mental representation of a character based on the role noun information and requires an update when the conflicting pronoun information is encountered. Kreiner, Sturt, and Garrod (2008) demonstrated that stereotypical gender may not be part of the lexical representation of the word – in contrast to definitional or lexical gender (as in 'mother' or 'king') – but is rather inferred from world knowledge.

In grammatical gender languages, gender mismatch effects occur even when grammatical gender markings of the role noun are congruent with an anaphor referring to it, that is, when grammatical gender agreement is not violated and only stereotypical gender is incongruent, as in *Oft hatte die Elektrikerin gute Einfälle, regelmässig plante sie neue Projekte*. 'The electrician (feminine) often had good ideas, she regularly planned new projects.' (e.g. Esaulova, Reali, & von Stockhausen, 2014). Interestingly, the processing of stereotypical gender is not limited to role nouns but can also be elicited by gender-stereotypical descriptions, such as 'K. L. installs power lines and cables, checks electricity voltage. In this field she has a lot of experience' (Reali, Esaulova, & von Stockhausen, 2015).

It is worth noting that the interaction between stereotypical and grammatical gender is sensitive to the stimulus material, the morphological expression of gender in a given language and the type of measurement, so that the mismatch effect may not always manifest (for a comparison of mismatch effects in English, French, and German see e.g. Garnham et al., 2012; for electro-physiological data from Italian see Siyanova-Chanturia, Pesciarelli, & Cacciari, 2012).

Although the processing of stereotypical and grammatical information may differ in terms of time course (e.g. Irmen, 2007), mismatch effects elicited by stereotypical gender cues are similar to the violations of gender agreement caused by incongruities in grammatical gender markings. As a result, grammatically accurate linguistic descriptions of persons (e.g. *die Elektrikerin* 'the electrician (feminine)') that are incongruent with gender stereotypes induce reactions comparable to those produced by language errors.

This perception of role nouns and their descriptions in language as only adequate when they comply with corresponding stereotypes reinforces existing expectations about gender roles even though they are no longer applicable in modern society.

3. Social agency and its linguistic representations

The biases described above may be argued to both stem from and be partly responsible for cognitive gender schemas that guide our behavior. Sex-role socialization also appears to largely rely on the characteristics of gender schemas that define masculine and feminine behavior, so that men are socialized to be achievement oriented and independent, whereas women are socialized to be relationship oriented and nurturing (Cross & Madson, 1997).

Interestingly, these characteristics are in line with agency and communion, the two fundamental dimensions guiding the social judgment of human behavior (Bakan, 1966; Abele et al., 2008). Agency is related to striving and experiences of achievement, power, and competence, whereas communion reflects connection and cooperation with others.

The two dimensions are typically operationalized using adjectives that describe personality traits (e.g. Meadows, 2003; Tausch, Kenworthy, & Hewstone, 2007; Wojciszke, 2005). When rated for agency and communion, trait words appear to provide consistent results across languages (Abele et al., 2008). Moreover, agency characteristics have been related to gender hierarchies that attribute higher power and status rankings to masculinity and show that men are perceived as more agentic compared to women (e.g. Koenig et al., 2011; Cuddy, Fiske, & Glick, 2008; Eagly & Karau, 2002). However, the way agency is represented in language goes far beyond adjective semantics and word level to support the described gender hierarchies, or biases, in various manners.

3.1 Verbs as markers of agency

In fact, the very linguistic categories (i.e. verbs, nouns, etc.) may be inherently related to agency and introduce gender biases in language when they are used. Formanowicz et al. (2017) propose, for instance, that verbs as a grammatical category convey agency and may support the perceptions of individuals or groups as more or less agentic. Crucially, they suggest that the association between verbs and agency goes beyond mere semantics of particular words and should rather be seen as a meta-semantic effect (Fiedler, 2008). This suggestion is made based on the dynamic properties of verbs that imply activity and is supported by corpora analyses and further experimental evidence.

In their 2017 study, Formanowicz et al. addressed the question of whether verbs are perceived as more agentic compared to other grammatical categories. In a series of three experiments, participants were asked to evaluate whether pseudo-words matched agency or communion characteristics and to what extent. The presented

pseudo-words could be identified as verbs, nouns, or adjectives through suffixes typical for the corresponding grammatical categories. In line with corpora analyses, the experimental data indicate that language users ascribe agency to verbs but not to nouns or adjectives.

In another study (Formanowicz et al., 2017), the authors assessed the number of co-occurrences of verbs following stereotypically agentic ("men") and non-agentic ("women") targets in the corpora of Polish and German languages. In these languages, the subject most commonly occurs before the verb and is related to the thematic role of an agent performing an action (rather than a patient role that has an action performed on it). The results show that agentic targets were more likely to be followed by verbs than non-agentic ones in both languages. This finding is interpreted as evidence in favor of a general link between agency and verbs irrespective of particular semantics of a given verb.

Taken together, these findings suggest that verbs as a category can be used as a vehicle to encode agency, which speakers successfully decode. As such, verbs can be seen as a linguistic tool that subtly supports the biased perception of gender in terms of agency, so that men are viewed as more agentic than women.

3.2 Case syncretism

Another linguistic phenomenon indicating implicit mechanisms involved in the maintenance of gender biases is the case system. Krifka (2009), for instance, considers the historical development of the gender and case system in Germanic languages to explain gender asymmetries present in modern German. These asymmetries are reflected in case syncretism (i.e. identical forms are used for different cases), which affects feminine (and neuter) but not masculine determiners.

Masculine determiners in German have an overt case marking in nominative (*der*) and accusative (*den*), so that distinct forms are available for each case. In contrast, neuter and feminine determiners have identical morphological forms in both cases (*das* and *die* respectively). As to neuter, the explanation provided for case syncretism is based on animacy, since neuter is originally a class of inanimates. Animates occur as both thematic agents (in nominative case) and patients (in cases other than nominative, e.g. accusative) and thus need to be marked to avoid ambiguity concerning their thematic role. Inanimates, on the other hand, occur in agent roles relatively rarely and therefore the distinction between nominative and accusative does not prove to be indispensable.

Evidently, this functional approach cannot be adopted to explain case syncretism for feminine, as it should not be any less animate than masculine. Krifka (2009) suggests that the asymmetry must result from the rare use of feminine noun phrases

in agent roles within the speech community, eventually leading to the illustrated syncretism.

Formanowicz et al. (2017) provides some corpora-based support for such an explanation, showing that "women" are indeed used less often as agents preceding verbs than "men". Nominative/accusative case syncretism for feminine determiners in German is a clear example of bias-supporting mechanisms found in a language's grammatical systems, that is, its very structure, as opposed to linguistic means that occur more at the surface, such as preferential phrasings (e.g. as in the use of negations). Even though this bias relies on gender on the grammatical level, we consider it implicit, since its relation to gender biases has so far been largely overlooked.

The lack of empirical studies on how such case syncretism may impact behavior with respect to gender biases makes it a research area that yet needs to be advanced.

3.3 Thematic agents

Thematic structure appears to be the most intuitive manifestation of how agency is instantiated in language and often seems to be the linguistic foundation for agency-related biases. Verbs determine the thematic structure of their arguments, which results in a hierarchical organization where agents have the highest ranking, followed by other roles, such as patients. Several semantic characteristics of arguments, such as animacy, person, and definiteness, are regarded as hierarchically organized information types – or prominence factors – and have been shown to influence how easily an entity is assigned an agent role (e.g. Siewierska, 2004; Wang et al., 2012). These characteristics interact with thematic structure such that arguments with higher-ranked characteristics (e.g. animate) map onto higher-ranked thematic roles (agents), a phenomenon known as harmonic alignment (Aissen, 2003).

A number of studies on language comprehension and production have used various experimental paradigms and methods to demonstrate that animate entities are perceived as better agents than inanimate ones (e.g. Bornkessel-Schlesewsky & Schlesewsky, 2009; Gennari & MacDonald, 2008; Trueswell, Tanenhaus, & Kello, 1993). Several recent studies aimed to verify whether similar preferences may arise based on gender characteristics of arguments, in which case they would also have to be considered as prominence factors, along with animacy.

In a series of eye-tracking experiments, Esaulova (2015) examined whether native French readers create gender-based expectations when they encounter the gender-ambiguous indirect object pronoun *lui*. In her experiments, the author used sentences like *En vérité, la diététicienne lui a recommandé, donc à ce/cette pharmacien/pharmacienne, un plan rigoureux* 'In fact, the dietician.FEM (typically female) recommended to him/her (gender-ambiguous), so to this.MASC/FEM pharmacist.MASC/FEM (neutral stereotypical gender), a strict plan'.

In these sentences the ambiguous pronoun referred to a role noun that was a patient and an object and appeared later in the sentence. Gender characteristics of role nouns in the sentence were manipulated. The patient/object nouns were neutral in terms of stereotypical gender and were either grammatically masculine or feminine. The agent/subject nouns were grammatically feminine and differed in their stereotypical gender – female or neutral – in Experiment 1, while they were grammatically masculine and stereotypically male or neutral in Experiment 2. Such an experimental setup had as its purpose to reveal whether readers have a preference for nouns with particular gender profiles to appear as agents or patients.

Both experiments demonstrated that grammatically feminine patients were processed faster than masculine ones, suggesting that feminine nouns fit the patient role better. The results of the stereotypical gender manipulation were concordant with grammatical gender biases: stereotypically female nouns took longer to be processed as agents compared to neutral ones, which in turn were more difficult than stereotypically male agents. Overall, the findings suggest a hierarchy, with male/masculine arguments fitting the agent role better than female/feminine ones, revealing another linguistic mechanism conveying implicit gender biases.

Moreover, this pattern was confirmed in another study in German (Esaulova, Reali, & von Stockhausen, 2017), which provides evidence that gender characteristics of nouns may be perceived hierarchically and as indicative of thematic roles across languages. In this study, readers were presented with sentences like *Die Flugbegleiterin, die viele Tourist-en/-innen beobachtet hat/haben, ist aufmerksam*. 'The flight attendant.FEM (typically female), who has observed many tourists.FEM/MASC (neutral stereotypical gender)/whom many tourists.FEM/MASC (neutral stereotypical gender) have observed, is attentive.'

In German, the thematic structure of these sentences remains ambiguous until readers reach the final word of the relative clause – the auxiliary verb *hat/haben* 'has/have' – which then disambiguates thematic agents/patients via number agreement. This ambiguity served to identify whether gender cues of role nouns would be used as cues for their thematic roles.

The findings indeed indicate that gender tends to be taken as an informative cue for the assignment of thematic roles, so that masculine/neutral nouns are perceived as better agents than feminine/female nouns (see Esaulova & von Stockhausen, 2015 for a detailed review of the design and results).

In a similar vein, Esaulova, Puebla Antunes and Felser (n.d.) examined whether lexical gender cues in nouns like 'mother' or 'king', where gender is part of the definition of the word, may also be used by readers to imply thematic roles of arguments. The design of the study included the manipulation of noun-phrase gender cues in constructions with subject- and object-extracted relative clauses, such as *Allerdings war es schön, dass die Stiefväter den Tanten/den Stiefvätern die Tanten*

gedankt haben. 'Certainly it was nice that the stepfathers (nominative/dative) the aunts (dative/nominative) thanked.' If gender cues were relevant for thematic role assignment, as shown in previous studies, readers would process masculine noun phrases in agent roles easier than feminine ones and the opposite would be true regarding patient roles.

While native German readers were sensitive to the non-canonical word order and found object-extracted relative clauses more difficult to process than subject-extracted ones, they did not show any gender preferences for agent or patient roles. Interestingly, when non-native German speakers were tested using the same materials, they showed sensitivity to both the word order and the gender characteristics of nouns.

However, their preferences for gender were not related to thematic roles but rather attributable to the processing of plural case markings. In particular, the mismatch effect was observed on noun phrases when the plural determiner *die*, which is identical for singular feminine nouns, appeared with masculine nouns, so that *die Stiefväter* 'the stepfathers' required more time to process than *die Tanten* 'the aunts'.

Future research should evaluate whether the presence of ambiguous context is a necessary condition for readers to form gender-based expectations regarding thematic roles. At this point, however, we would like to argue that there is evidence that the processing of thematic structure closely interacts with the processing of gender in such a way that gender may be considered a prominence factor. Thematic gender characteristics of nouns appear to influence the extent to which they are perceived as fitting agent roles well, even though findings so far indicate that this seems to occur only when an ambiguous context is provided.

4. Conclusions

This chapter provides an overview of linguistic phenomena that can be seen as tools maintaining the existing gender stereotypes in ways that are often overlooked or underestimated. It certainly does not give a full account of such linguistic means, but rather illustrates versatile mechanisms that may implicitly bias the expectations and perceptions of language users regarding male and female gender roles. On the one hand, some of these mechanisms rely on linguistic choices, such as word choices (e.g. abstract vs. concrete) and structural choices (e.g. thematic role assignment). On the other hand, some mechanisms are ingrained in the very core of the language, its grammatical system, as we observed in the example of case syncretism in German. Each of these mechanisms in its own way systematically contributes to biased perceptions of men and women, supporting existing gender stereotypes on semantic, morphological, syntactic, and other language levels.

As pervasive as gender biases appear, uncovering them in language is not a straightforward task. Part of the difficulty comes from the interdisciplinary methodological diversity that is required in order to identify biases, comprising both social psychological and psycholinguistic expertise. Some tendencies (e.g. preferences for gender characteristics of thematic agents) seem to be identifiable only in contexts where language users need to rely on gender-based expectations, such as in experimental paradigms involving ambiguity resolution, whereas others (e.g. expectancy bias) may be displayed in spontaneous language production.

Another understandable difficulty is that most of the available empirical studies on gender biases have been conducted in the West and are Eurocentric. Even though many of the linguistic mechanisms described in this chapter can be extended to other languages, as long as these languages allow for the phenomena in question (i.e. negation, irony, thematic agents/patients, etc.), the validity of such generalizations must be confirmed by proper research evidence in the future.

Choosing adequate methods is critical to both expand our knowledge about gender biases and trace the dynamics with which they may change in response to corresponding political and social interventions, such as the implementation of gender-fair language strategies. Word choice and phrasing preferences may be altered given an adequate intervention, which explains the efficiency of some gender-fair language interventions, such as those targeting the use of alternatives for masculine forms used generically. While the revision of the grammatical systems of a language in this sense may be problematic, revealing gender asymmetries encoded in formal language systems should bring awareness to gender biases that would otherwise go unnoticed.

Understanding language-supported gender biases as a complex multi-level phenomenon has several important implications. An obvious one concerns measures directed to achieve gender equality, which should involve a coordinated approach targeting both explicit and implicit manifestations of gender asymmetries. Moreover, showing examples of gender asymmetries and their effects on language processing should encourage better control for gender when it comes to experimental materials used in psycholinguistic studies. Given the examples of how cognitive processes are affected by gender-related expectations, better-controlled linguistic materials should be relevant even for research that does not directly address gender differences.

While this chapter outlined several phenomena in language where gender biases are at play, the extent to which they are present across languages and the relative impact of a separately taken linguistic mechanism for the construction, maintenance, and transmission of gender information should be clarified by future research.

References

Abele, A. E., Cuddy, A. J. C., Judd, C. M., & Yzerbyt, V. Y. (2008). Fundamental dimensions of social judgment. *European Journal of Social Psychology*, 38(7), 1063–65. https://doi.org/10.1002/ejsp.574

Abele, A. E., Uchronski, M., Suitner, C., & Wojciszke, B. (2008). Towards an operationalization of the fundamental dimensions of agency and communion: Trait content ratings in five countries considering valence and frequency of word occurrence. *European Journal of Social Psychology*, 38(7), 1202–1217. https://doi.org/10.1002/ejsp.575

Aissen, J. (2003). Differential object marking: Iconicity vs. economy. *Natural Language & Linguistic Theory*, 21(3), 435–483. https://doi.org/10.1023/A:1024109008573

Arcuri, L., Maass, A., & Portelli, G. (1993). Linguistic intergroup bias and implicit attributions. *British Journal of Social Psychology*, 32(3), 277–285. https://doi.org/10.1111/j.2044-8309.1993.tb01001.x

Attardo, S. (2000). Irony as relevant inappropriateness. *Journal of Pragmatics*, 32(6), 793–826. https://doi.org/10.1016/S0378-2166(99)00070-3

Bakan, D. (1966). *The duality of human existence: Isolation and communion in western man*. Beacon Press.

Bem, S. (1993). *The lenses of gender: Transforming the debate on sexual inequality*. Yale University Press.

Bem, S. L., & Bem, D. J. (1973). Does sex-biased job advertising aid and abet sex discrimination. *Journal of Applied Social Psychology*, 3(1), 6–18. https://doi.org/10.1111/j.1559-1816.1973.tb01290.x

Bencivenni, N. (1990). L'uso di categorie linguistiche in situazioni intergruppo e low impatto sulle informazioni successive [The use of linguistic categories in intergroup situations and their impact on subsequent information processing] (Unpublished Master's thesis). University of Padua.

Beukeboom, C. J., Finkenauer, C., & Wigboldus, D. H. J. (2010). The negation bias: When negations signal stereotypic expectancies. *Journal of Personality and Social Psychology*, 99(6), 978–992. https://doi.org/10.1037/a0020861

Bornkessel-Schlesewsky, I., & Schlesewsky, M. (2009). The role of prominence information in the real-time comprehension of transitive constructions: A cross-linguistic approach. *Language and Linguistics Compass*, 3(1), 19–58. https://doi.org/10.1111/j.1749-818X.2008.00099.x

Bundesministerium für Frauen und Öffentlichen Dienst. (2009). *Geschlechtergerechte Stellenausschreibungen* [Gender-fair job advertisements] (Unabhängiger Bericht der Gleichbehandlungsanwaltschaft iS §3 Abs 5 GBK/GAW-Gesetz). Bundesministerium für Frauen und Öffentlichen Dienst.

Burgers, C. & Beukeboom, C. J. (2016). Stereotype transmission and maintenance through interpersonal communication: The irony bias. *Communication Research*, 43(3): 414–441. https://doi.org/10.1177/0093650214534975

Carnaghi, A., Maass, A., Gresta, S., Bianchi, M., Cadinu, M., & L. Arcuri. (2008). Nomina sunt omina: On the inductive potential of nouns and adjectives in person perception. *Journal of Personality and Social Psychology*, 94(5), 839–859. https://doi.org/10.1037/0022-3514.94.5.839

Carpenter, P. A., & Just, M. A. (1975). Sentence comprehension: A psycholinguistic processing model of verification. *Psychological Review*, 82(1): 45–73. https://doi.org/10.1037/h0076248

Carreiras, M., Garnham, A., Oakhill, J., & Cain, K. (1996). The use of stereotypical gender information in constructing a mental model: Evidence from English and Spanish. *The Quarterly Journal of Experimental Psychology, A* 49(3), 639–663. https://doi.org/10.1080/713755647

Chestnut, E. K., & Markman, E. M. (2018). "Girls are as good as boys at math" implies that boys are probably better: A study of expressions of gender equality. *Cognitive Science*, 42, 2229–2249. https://doi.org/10.1111/cogs.12637

Cimpian, A., & Markman, E. M. (2009). Information learned from generic langauge becomes central to children's biological concepts: Evidence from their open-ended explanations. *Cognition*, 113(1), 14–25. https://doi.org/10.1016/j.cognition.2009.07.004

Cimpian, A., & Markman, E. M. (2011). The generic/nongeneric distinction influences how children interpret new information about social others. *Child Development*, 82(2), 471–492. https://doi.org/10.1111/j.1467-8624.2010.01525.x

Cross, S. E., & Madson, L. (1997). Models of the self: Self-construals and gender. *Psychological Bulletin*, 122(1), 5–37. https://doi.org/10.1037/0033-2909.122.1.5

Cuddy, A. J. C., Fiske, S. T., & Glick, P. (2008). Warmth and competence as universal dimensions of social perception: The stereotype content model and the BIAS map. In M. P. Zanna (Ed.), *Advances in experimental social psychology* (Vol. 40, pp. 61–149). Elsevier. https://doi.org/10.1016/S0065-2601(07)00002-0

Deutsch, R., Gawronski, B., & Strack, F. (2006). At the boundaries of automaticity: Negation as reflective operation. *Journal of Personality and Social Psychology*, 91(3), 385–405. https://doi.org/10.1037/0022-3514.91.3.385

Eagly, A. H., & Karau, S. J. (2002). Role congruity theory of prejudice toward female leaders. *Psychological Review*, 109(3), 573–598. https://doi.org/10.1037/0033-295X.109.3.573

Esaulova, Y. (2015). The prominence of gender information in on-line language processing: Cross-linguistic evidence of implicit gender hierarchies (Unpublished PhD dissertation). University of Heidelberg.

Esaulova, Y., Puebla Antunes, C., & Felser, C. (n.d.). The processing of case, word order and prominence features in German. Manuscript in preparation.

Esaulova, Y., Reali, C., & von Stockhausen, L. (2014). Influences of grammatical and stereotypical gender during reading: Eye movements in pronominal and noun phrase anaphor resolution. *Language, Cognition and Neuroscience*, 29(7), 781–803. https://doi.org/10.1080/01690965.2013.794295

Esaulova, Y., Reali, C., & von Stockhausen, L. (2017). Prominence of gender cues in the assignment of thematic roles in German. *Applied Psycholinguistics*, 38(5), 1133–1172. https://doi.org/10.1017/S014271641700008X

Esaulova, Y., & von Stockhausen, L. (2015). Cross-linguistic evidence for gender as a prominence feature. *Frontiers in Psychology*, 6(September), 1356. https://doi.org/10.3389/fpsyg.2015.01356

Fiedler, K. (2008). Language: A toolbox for sharing and influencing social reality. *Perspectives on Psychological Science*, 3(1), 38–47. https://doi.org/10.1111/j.1745-6916.2008.00060.x

Fiedler, K., Semin, G. R., & Finkenauer, C. (1993). The battle of words between gender groups a language-based approach to intergroup processes. *Human Communication Research*, 19(3), 409–41. https://doi.org/10.1111/j.1468-2958.1993.tb00308.x

Formanowicz, M., Roessel, J., Suitner, C. & Maass, A. (2017). Verbs as linguistic markers of agency: The social side of grammar. *European Journal of Social Psychology* 47 (5): 566–579. https://doi.org/10.1002/ejsp.2231

Gabriel, U., Gygax, P. M., & Kuhn, E. A. (2018). Neutralising linguistic sexism: Promising but cumbersome? *Group Processes & Intergroup Relations*, 21(5), 844–858. https://doi.org/10.1177/1368430218771742

Garnham, A. (1987). *Mental models as representations of discourse and text*. Halsted Press.

Garnham, A., Gabriel, U., Sarrasin, O., Gygax, P., & Oakhill, J. (2012). Gender representation in different languages and grammatical marking on pronouns: When beauticians, musicians, and mechanics remain men. *Discourse Processes*, 49(6), 481–500. https://doi.org/10.1080/0163853X.2012.688184

Gelman, S. A., Ware, E. A., & Kleinberg, F. (2010). Effects of generic language on category content and structure. *Cognitive Psychology*, 61(3), 273–301. https://doi.org/10.1016/j.cogpsych.2010.06.001

Gennari, S. P., & MacDonald, M. C. (2008). Semantic indeterminacy in object relative clauses. *Journal of Memory and Language*, 58(2), 161–187. https://doi.org/10.1016/j.jml.2007.07.004

Geschke, D., Sassenberg, K., Ruhrmann, G., & Sommer, D. (2010). Effects of linguistic abstractness in the mass media: How newspaper articles shape readers' attitudes toward migrants. *Journal of Media Psychology*, 22(3), 99–104. https://doi.org/10.1027/1864-1105/a000014

Gibbs, R. W. (2000). Irony in talk among friends. *Metaphor and Symbol*, 15(1–2), 5–27. https://doi.org/10.1080/10926488.2000.9678862

Giles, H., & Maass, A. (2016). *Advances in intergroup communication*. Peter Lang. https://doi.org/10.3726/b10467

Giora, R., Fein, O., Aschkenazi, K., & Alkabets-Zlozover, I. (2007). Negation in context: A functional approach to suppression. *Discourse Processes*, 43(2), 153–172. https://doi.org/10.1080/01638530709336896

Giora, R., Fein, O., & Schwartz, T. (1998). Irony: Grade salience and indirect negation. *Metaphor and Symbol*, 13(2), 83–101. https://doi.org/10.1207/s15327868ms1302_1

Glenberg, A. M., Robertson, D. A., Jansen, J. L., & Johnson-Glenberg, M. C. (1999). Not propositions. *Cognitive Systems Research*, 1(1), 19–33. https://doi.org/10.1016/S1389-0417(99)00004-2

Gygax, P. M., Elmiger, D., Zufferey, S., Garnham, A., Sczesny, S., von Stockhausen, L., Braun, F., & Oakhill, J. (2019). A language index of grammatical gender dimensions to study the impact of grammatical gender on the way we perceive women and men. *Frontiers in Psychology*, 10(July), 1604. https://doi.org/10.3389/fpsyg.2019.01604

Hamilton, D. L., & Trolier, T. K. (1986). Stereotypes and stereotyping: An overview of the cognitive approach. In J. F. Dovidio & S. L. Gaertner (Eds.), *Prejudice, discrimination, and racism* (pp. 127–163). Academic Press.

Haslanger, S. (2011). Ideology, generics, and common ground. In C. Witt (Ed.), *Feminine metaphysics: Explorations in the ontology of sex, gender, and the self* (pp. 179–209). Springer. https://doi.org/10.1007/978-90-481-3783-1_11

Hellinger, M., & Bierbach, C. (1993). *Eine Sprache für beide Geschlechter: Richtlinien für einen nicht-sexistischen Sprachgebrauch* [A language for both genders – Guidelines for non-sexist language use]. Dt. UNESCO-Komm.

Hellinger, M., & Bußmann, H. (Eds). (2001–2002–2003). *Gender across languages: The linguistic representation of women and men* (Vols. 1, 2, 3). John Benjamins. https://doi.org/10.1075/impact.9

Hodel, L., Formanowicz, M., Sczesny, S., Valdrová, J., & von Stockhausen, L. (2017). Gender-fair language in job advertisements: A cross-linguistic and cross-cultural analysis. *Journal of Cross-Cultural Psychology*, 48(3), 384–401. https://doi.org/10.1177/0022022116688085

Horvath, L. K., & Sczesny, S. (2016). Reducing women's lack of fit with leadership positions? Effects of the wording of job advertisements. *European Journal of Work and Organizational Psychology*, 25(2), 316–328. https://doi.org/10.1080/1359432X.2015.1067611

Irmen, L. (2007). What's in a (role) name? Formal and conceptual aspects of comprehending personal nouns. *Journal of Psycholinguistic Research*, 36(6), 431–456. https://doi.org/10.1007/s10936-007-9053-z

Jordan, M. P. (1998). The power of negation in English: Text, context and relevance. *Journal of Pragmatics*, 29(6), 705–752. https://doi.org/10.1016/S0378-2166(97)00086-6

Kennison, S. M., & Trofe, J. L. (2003). Comprehending pronouns: A role for word-specific gender stereotype information. *Journal of Psycholinguistic Research*, 32, 355–378. https://doi.org/10.1023/A:1023599719948

Kesebir, S. (2017). Word order denotes relevance differences: The case of conjoined phrases with lexical gender. *Journal of Personality and Social Psychology*, 113(2), 262–279. https://doi.org/10.1037/pspi0000094

Koenig, A. M., Eagly, A. H., Mitchell, A. A., & Ristikari, T. (2011). Are leader stereotypes masculine? A meta-analysis of three research paradigms. *Psychological Bulletin*, 137(4), 616–642. https://doi.org/10.1037/a0023557

Kreiner, H., Sturt, P., & Garrod, S. (2008). Processing definitional and stereotypical gender in reference resolution: Evidence from eye-movements. *Journal of Memory and Language*, 58(2), 239–261. https://doi.org/10.1016/j.jml.2007.09.003

Krifka, M. (2009). Case syncretism in german feminines: Typological, functional and structural aspects. In P. O. Steinkrüger & M. Krifka (Eds.), *On inflection* (pp. 141–172). De Gruyter Mouton.

Leslie, S. J. (2014). Carving up the social world with generics. *Oxford Studies in Experimental Philosophy*, 1, 208–232. https://doi.org/10.1093/acprof:oso/9780198718765.003.0009

Leslie, S. J. (2015). 'Hillary Clinton is the only man in the Obama administration': Dual character concepts, generics, and gender. *Analytic Philosophy*, 56(2), 111–141. https://doi.org/10.1111/phib.12063

Maass, A., & Arcuri, L. (1996). Language and stereotyping. In C. N. Macrae, C. Stangor, M. Hewstone (Eds.), *Stereotypes and stereotyping* (pp. 193–226). The Guilford Press.

Maass, A., Corvino, P., & Arcuri, L. (1994). Linguistic intergroup bias and the mass media. *Revue de Psychologie Sociale*, 1, 31–43.

Maass, A., Salvi, D., Arcuri, L., & Semin, G. R. (1989). Language use in intergroup contexts: The linguistic intergroup bias. *Journal of Personality and Social Psychology*, 57(6), 981–993. https://doi.org/10.1037/0022-3514.57.6.981

Mayo, R., Schul, Y., & Burnstein, E. (2004). 'I am not guilty' vs 'I am innocent': Successful negation may depend on the schema used for its encoding. *Journal of Experimental Social Psychology*, 40(4), 433–449. https://doi.org/10.1016/j.jesp.2003.07.008

McConnell-Ginet, S. (2012). Generic predicates and interest relativity. *Canadian Journal of Linguistics*, 57(2), 289–302. https://doi.org/10.1017/S0008413100004771

Meadows, N. K. (2003). Agency and communion: Multifaceted or unitary constructs? (Unpublished PhD dissertation). Western University, London, Ontario.

Misersky, J., Gygax, P. M., Canal, P., Gabriel, U., Garnham, A., Braun, F., Chiarini, T. et al. (2014). Norms on the gender perception of role nouns in Czech, English, French, German, Italian, Norwegian, and Slovak. *Behavior Research Methods*, October.

Reali, C., Esaulova, Y., & von Stockhausen, L. (2015). Isolating stereotypical gender in a grammatical gender language: Evidence from eye movements. *Applied Psycholinguistics*, 36(4), 977–1006. https://doi.org/10.1017/S0142716414000010

Rhodes, M., Leslie, S.-J., & Tworek, C. M. (2012). Cultural transmission of social essentialism. *Proceedings of the National Academy of Sciences of the United States of America*, 109(34): 13526–13531. https://doi.org/10.1073/pnas.1208951109

Rubini, M., & Semin, G. R. (1994). Language use in the context of congruent and incongruent in-group behaviours. *British Journal of Social Psychology*, 33(3), 355–362. https://doi.org/10.1111/j.2044-8309.1994.tb01031.x

Schwab, K., Samans, R., Zahidi, S., Leopold, T. A., & Ratcheva, V. (2017). *The global gender gap report: 2017*. World Economic Forum.

Sczesny, S., Formanowicz, M., & Moser, F. (2016). Can gender-fair language reduce gender stereotyping and discrimination? *Frontiers in Psychology*, 7(February). https://doi.org/10.3389/fpsyg.2016.00025

Semin, G. R., & Fiedler, K. (1988). The cognitive functions of linguistic categories in describing persons: Social cognition and language. *Journal of Personality and Social Psychology*, 54(4), 558–568. https://doi.org/10.1037/0022-3514.54.4.558

Semin, G. R., & Fiedler, K. (1992). The inferential properties of interpersonal verbs. In G. R. Semin & K. Fiedler (Eds.), *Language, interaction and social cognition* (pp. 58–78). Sage.

Siewierska, A. (2004). *Person*. Cambridge University Press. https://doi.org/10.1017/CBO9780511812729

Siyanova-Chanturia, A., Pesciarelli, F., & Cacciari, C. (2012). The electrophysiological underpinnings of processing gender stereotypes in language. *PLoS ONE*, 7(12), e48712. https://doi.org/10.1371/journal.pone.0048712

Slobin, D. I. (2003). Language and thought online: Cognitive consequences of linguistic relativity. In D. Gentner & S. Goldin-Meadow (Eds.), *Language in mind: Advances in the study of language and thought* (pp. 157–192). The MIT Press.

Stahlberg, D., Braun, F., Irmen, L., & Sczesny, S. (2007). Representation of the sexes in language. In K. Fiedler (Ed.), *Social communication* (pp. 163–187). Psychology Press.

Stout, J. G., & Dasgupta, N. (2011). When he doesn't mean you: Gender-exclusive language as ostracism. *Personality and Social Psychology Bulletin*, 37(6), 757–769. https://doi.org/10.1177/0146167211406434

Tausch, N., Kenworthy, J. B., & Hewstone, M. (2007). The confirmability and disconfirmability of trait concepts revisited: Does content matter? *Journal of Personality and Social Psychology*, 92(3), 542–556. https://doi.org/10.1037/0022-3514.92.3.542

Trueswell, J. C., Tanenhaus, M. K., & Kello, C. (1993). Verb-specific constraints in sentence processing: Separating effects of lexical preference from garden-paths. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 19(3), 528–553.

Valdrová, J. (2013). Typologische Unterschiede zwischen Sprachen als Argument gegen geschlechtergerechte Sprachkritik und Sprachpflege? [Typological differences between languages as an argument against language criticism and language cultivation?]. In M. Nekula, K. Šichová, & J. Valdrová (Eds.), *Bilingualer Sprachvergleich und Typologie: Deutsch – Tschechisch* [Bilingual language use and typology: German – Czech] (pp. 141–158). Julius Groos.

Vervecken, D., Hannover, B., & Wolter, I. (2013). Changing (s)expectations: How gender fair job descriptions impact children's perceptions and interest regarding traditionally male occupations. *Journal of Vocational Behavior*, 82(3), 208–220. https://doi.org/10.1016/j.jvb.2013.01.008

Villiers, J. G. de, & Tager Flusberg, H. B. (1975). Some facts one simply cannot deny. *Journal of Child Language*, 2(2), 279–286. https://doi.org/10.1017/S0305000900001100

Wang, L., Schlesewsky, M., Philipp, M., & Bornkessel-Schlesewsky, I. (2012). The role of animacy in online argument interpretation in Mandarin Chinese. In M. Lamers & P. de Swart (Eds.), *Case, word order, and prominence. Psycholinguistic and theoretical approaches to argument structure* (pp. 91–119). Springer. https://doi.org/10.1007/978-94-007-1463-2_5

Wason, P. C. (1965). The contexts of plausible denial. *Journal of Verbal Learning and Verbal Behavior*, 4(1), 7–11. https://doi.org/10.1016/S0022-5371(65)80060-3

Wigboldus, D. H. J., Semin, G. R., & Spears, R. (2000). How do we communicate stereotypes? Linguistic bases and inferential consequences. *Journal of Personality and Social Psychology*, 78(1), 5–18. https://doi.org/10.1037/0022-3514.78.1.5

Wojciszke, B. (2005). Morality and competence in person- and self-perception. *European Review of Social Psychology*, 16(1), 155–88. https://doi.org/10.1080/10463280500229619

CHAPTER 4

The acquisition of grammatical gender in child and adult heritage speakers of Spanish

Beyond the source of differences

Alejandro Cuza and Liliana Sánchez
Purdue University / University of Illinois at Chicago

Previous work on the acquisition of grammatical gender in child and adult heritage speakers of Spanish has found significant mismatches in gender agreement stemming from overgeneralization of the masculine form to contexts in which the feminine is required. It has been argued that these divergences stem from various sources including incomplete acquisition during childhood (e.g. Montrul & Potowski, 2007; Montrul, Foote, & Perpiñan, 2008), form/meaning mapping issues (e.g. Alarcón, 2011) or reconfiguration of gender features (e.g. Cuza & Pérez-Tattam, 2016; Scontras, Polinsky, & Fuchs, 2018). The goal of the present chapter is to examine this previous work on grammatical gender and the extent to which it can shed light on more recent proposals on heritage language theorizing. We follow the Bilingual Alignment Approach (Sánchez, 2019) to argue that the divergences heritage speakers show can be better accounted for in relation to crosslinguistic influence at the level of bilingual alignments, linguistic proficiency and specific patterns of language exposure and usage.

Keywords: grammatical gender, heritage speakers, bilingual alignments, language experience

1. Introduction

The acquisition of gender in Spanish monolingual children often occurs without much difficulty by the age of three, although some studies have documented gender mismatches with non-canonical nouns (e.g. López-Ornat, Fernández, Gallo, & Mariscal, 1994; Mariscal, 2009; Gathercole, 2002; Pérez-Pereira, 1991). In second language (L2) learners as well as in child and adult heritage speakers, however, gender assignment and agreement in Spanish has been found particularly challenging to acquire (e.g. Alarcón, 2009, 2011; Cuza & Pérez-Tattam, 2016; Franceschina, 2001; Hur, López-Otero, & Sánchez, 2020; Montrul, Foote, & Perpiñán, 2008;

https://doi.org/10.1075/sibil.63.04cuz

Montrul & Potowski, 2007; Sagarra & Herschensohn, 2011; Scontras et al., 2018; Valenzuela et al., 2012).[1]

As we discuss shortly, heritage speakers often show significant gender mismatches in the nominal domain compared to the monolingual speakers acting as baseline, evidenced in overgeneralization of the masculine form with determiners and adjectives (e.g.*_un nube rojo_ vs. _una nube roja_ for 'a red cloud'). Despite extensive research, it is still unclear what the sources of the divergences are in relation to grammatical gender (e.g. Hur et al., 2020; Montrul, 2002, 2008; Polinsky, 2018; Polinsky & Scontras, 2019, 2020). Do heritage speakers undergo incomplete acquisition of [+strong] gender features during early childhood? Are the mismatches in gender morphology evidence of form/meaning mapping issues, or are they the result of a process of feature reassembly (Lardiere, 2008, 2009) and grammatical restructuring as recently argued by Sánchez and colleagues? (e.g. Pérez Cortes, Putnam, & Sánchez, 2019; Putnam & Sánchez, 2013).

The goal of the present chapter is to provide a review of previous work on the acquisition of grammatical gender in child and adult heritage speakers of Spanish in contact with English as the majority language. In doing so, we discuss the extent to which current proposals in heritage language theorizing can best account for the divergences we often find in heritage speakers of Spanish in regard to gender assignment and agreement. Crosslinguistic influence from the majority language, language dominance, age of onset of bilingualism and patterns of language activation and use can all affect the extent of morphosyntactic shifts in heritage language grammars. However, a crucial question in the literature is whether these factors bring about representational issues in the form of incomplete development (e.g. Montrul, 2002, 2008; Montrul et al., 2008; Silva-Corvalán, 2014, 2018a; b; Montrul & Silva-Corvalán, 2019) and L1 attrition (e.g. Polinsky, 2011, 2013, 2008, 2018), or whether the divergences are the result of feature reassembly and reconfiguration stemming from low patterns of language activation and use for production and comprehension purposes (e.g. Giancaspro & Sánchez, 2021; Pérez-Cortes et al., 2019; Putnam & Sánchez, 2013; Sánchez, 2019). In the next section, we discuss gender expression in English and Spanish followed by a discussion on the nature of differences in heritage language bilingualism in Section 3. In Section 4, we provide a summary of previous research in the acquisition of gender in Spanish. Section 5 discusses Sánchez's (2019) Bilingual Alignment Approach as an optimal account for the divergences heritage speakers show. The conclusions of the chapter are presented in Section 6.

1. Heritage speakers are second or third-generation immigrants who have acquired their minority language at home during childhood while a majority language is also spoken (e.g. Kupisch & Rothman, 2018; Polinsky, 2018; Valdes, 2001).

2. Gender expression in Spanish and English

English and Spanish diverge in the way the two languages instantiate gender assignment and agreement within the Determiner Phrase (DP) (e.g. Demonte, 1999; Zagona, 2002). Spanish nouns are lexically marked for feminine or masculine gender. In most cases of inanimate nouns, gender agreement is marked morphologically with the suffix *–a* usually for feminine and the suffix *–o* for masculine which makes canonical or transparent nouns easier to acquire given the morphophonemic cues (e.g. Bruhn de Garavito & White, 2002).[2] However, this does not mean that canonical nouns are more frequent than nouns with other endings or that most gender mismatches occur with noncanonical ending nouns. Gender assignment on animate nouns is semantically motivated (natural gender) and in most nouns denoting humans or animals their gender corresponds with their biological sex (e.g. *hombre*-MASC 'man', *mujer*-FEM 'women'). Most Spanish determiners and adjectives are also inflected for gender, and they must agree with the noun they modify in both gender and number, as in (1a–b) below:

(1) a. [$_{DET}$ una-FEM] [$_{N}$ casa-FEM [$_{ADJ}$ roja-FEM]]
'A red house'
b. [$_{DET}$ un-MASC] [$_{N}$ anillo-MASC [$_{ADJ}$ pequeño-MASC]]
"A small ring"

Gender assignment is a lexical phenomenon that tends to be determined by native speakers on the basis of agreement between the DET and the N, and to a lesser extent by agreement between the N and the ADJ (Grüter et al., 2012). Gender agreement is a syntactic operation between the N and the DET and/or the ADJ (noun raising) (e.g. Bernstein, 1993; Carstens, 1991; Cinque, 1994).

Given the lack of a one-to-one correlation between overt morphology and grammatical gender, gender agreement with non-canonical nouns is often more difficult to master among monolingual and bilingual children (e.g. Cuza & Pérez-Tattam, 2016; Gathercole, 2002; López-Ornat et al., 1994; Mariscal, 2009; Montrul & Potowski, 2007) and adult heritage speakers of Spanish (e.g. Alarcón, 2011; Hur et al., 2020; Montrul et al., 2008).

In contrast to Spanish, English is devoid of rich morphological inflection on determiners, nouns, and adjectives. There is no gender marking on nouns, except for some lexicalized animate nouns (actor/actress). Therefore, no agreement between determiner, noun and adjective is possible. This represents a learnability problem

2. Harris (1991) argues that these endings are not true gender markers in Spanish but word type markers. However, most masculine and feminine nouns follow these prototypical gender forms (e.g. Teschner & Russell, 1984).

for English-dominant heritage speakers of Spanish as they have to develop these form/meaning connections that are nonexistent in English.

Regarding word order, the Spanish DP instantiates the functional category Number Phrase (NumP), the locus for gender and number agreement features. The [+strong] gender and number features of NumP are argued to trigger overt movement of the noun to a preadjectival position for feature checking of uninterpretable features. This operation then yields the canonical N+ADJ word order in Spanish (*nube blanca* "white cloud").[3] Alternative analyses propose NP movement rather than noun movement (e.g. Cinque, 2010; Lamarche, 1991; Sánchez, 1996). Gender agreement under such analyses would be the result of the Agree operation between [+strong] gender features in D as a probe and their uninterpretable counterparts in the head of NP and AdjP. In English, the NumP instantiates [−strong] number features and therefore the N remains in situ (postadjectival position) (ADJ+N) (*red car*).

In sum, English and Spanish diverge in the way they instantiate gender. Spanish determiners, nouns and adjectives are lexically marked for gender, and in most cases, this is reflected morphologically with the endings *a* for feminine or *o* for masculine. English, however, is devoid of gender morphology on determiners, nouns or adjectives except for some lexicalized nouns. The presence of [+strong] gender and number features in Spanish have been argued to trigger noun raising to a preadjectival position; in English, in contrast, the noun remains *in situ* given English [−strong] gender and number features. These differences between the two languages often lead to crosslinguistic influence from English as dominant language into Spanish, as we discuss in the following section.

3. The nature of differences in heritage language bilingualism

Gender agreement in Spanish is problematic for Spanish heritage speakers. Various proposals have been put forward to account specifically for these divergences and for the divergences that heritage speakers show in general compared to the monolingual baseline. For example, previous work has argued that the divergences heritage speakers show might stem from incomplete acquisition due to limited exposure to and use of the minority language during childhood (e.g. Montrul, 2002, 2008, 2009; Montrul et al., 2008; Montrul & Sánchez-Walker, 2013; Polinsky, 2006; Silva-Corvalán, 2014). In other words, heritage speakers do not benefit from

3. Attributive adjectives normally appear in postnominal position in Spanish. However, non-restrictive adjectives can appear before or after the noun with subtle differences in meaning (e.g. *famoso lingüista* vs. *lingüista famoso* "famous linguist").

enough language input during early childhood to fully activate target forms in the minority language, leading to incomplete development of specific aspects of grammar or morphosyntactic forms.

Theoretically speaking, an incomplete acquisition approach is controversial in the sense that it is unclear what *incomplete* means in terms of the internal grammatical representation that children develop upon exposure to input. Rather than incomplete development, it might be the case that children develop a different albeit complete grammatical representation on the basis of the input to which they are exposed (e.g. Polinsky & Scontras, 2020; Kupisch & Rothman, 2018). Furthermore, as discussed by Kupisch and Rothman (2018), the term might have unintended negative connotations as some might interpret it in the sense that heritage speakers' grammars are deficient (e.g. Otheguy, 2016; Pascual y Cabo & Rothman, 2012).[4]

In contrast to an incomplete acquisition approach, other researchers have argued that the divergences heritage speakers exhibit stem from child L1 attrition during childhood (e.g. Bylund, 2009; Flores, 2014; Polinsky, 2008, 2011). L1 attrition refers to the loss or restructuring of previously developed grammatical structures due to crosslinguistic influence from the dominant language as well as reduced input and use of the first language (e.g. Gürel, 2004; Montrul, 2002; Schmid, 2002; Schmid & Köpke, 2007; Silva-Corvalán, 1991).

Attrition has been found primarily among bilingual children who lose aspects of their L1 due to intense contact with a majority language in their lifespan. It resembles incomplete acquisition in that both processes are argued to be selective, affecting specific areas of the grammar where the syntax interfaces with other domains including pragmatics and semantics (e.g. Sorace, 2000, 2005, 2011; Tsimpli et al., 2004).

The term attrition has also been perceived negatively in the literature, as in the case of incomplete acquisition. Although some researchers treat attrition and incomplete acquisition as separate outcomes, they are not necessarily mutually exclusive. Previous work documents cases in which both processes occur simultaneously or sequentially depending on the type of grammatical property and bilingual population (e.g. Cuza, Pérez-Tattam, Barajas, Miller, & Sadowski, 2013; Jia & Paradis, 2015; Montrul, 2002, 2008; Silva-Corvalán, 2014). For example, while examining preterit vs. imperfect morphology use cross-sectionally, Cuza et al. (2013) found overgeneralization of the preterit among the older bilingual children

4. The term *incomplete acquisition* is controversial in the field and it should be used with caution to avoid any potential misinterpretation. However, the terminological debate is beyond the scope of the current chapter. Readers are referred to various commentaries on this issue (e.g. Domínguez, Hicks, & Slabakova, 2019; Kupisch & Rothman, 2018; Montrul & Silva-Corvalán, 2019; Otheguy, 2016; Pascual y Cabo & Rothman, 2012).

only, suggesting child L1 attrition effects. The imperfect form, however, remained underdeveloped across young and older bilingual children, suggesting incomplete development.

Regarding the acquisition of Spanish gender specifically, Alarcón (2011) argues that the divergences heritage speakers have are the result of processing issues in the target form/meaning mapping. The author adopts the Missing Surface Inflection Hypothesis to conclude that both heritage and L2 learners have acquired gender features in their underlying grammars despite their difficulties in the target mapping of the abstract gender feature to its appropriate form, especially with non-canonical nouns as documented in previous research in L1 and L2 acquisition (e.g. Franceschina, 2001; Montrul et al., 2008, 2013, 2014).

Although it could be true that in general non-canonical nouns are less frequent in the input, this is not always the case, and previous work shows gender mismatches even with high frequent canonical ending nouns (i.e., **el*-MASC *niña*-FEM 'the girl') (see Montrul & Potowski, 2007 for discussion). Lexical frequency measures are often taken from monolingual corpora rather than from the actual bilingual communities the speakers come from. For example, the noun *guante* 'glove' in Spanish might be highly activated among bilinguals in the Midwest but so much in the southeast. Therefore, future research would benefit from self-reported measures of lexical frequency by the heritage speakers under examination, as conducted by Hur et al. (2020).

More recently, Cuza and Pérez-Tattam (2016) followed Putnam and Sánchez (2013) in arguing that the difficulties child heritage speakers experience with gender stem from a reassembly of the gender features due to low patterns of language activation and use. Putnam and Sánchez's (2013) activation and reassembly approach adopts Lardiere's Feature Reassembly Hypothesis (e.g. Lardiere, 1998, 2008, 2009) and maintains that the divergences heritage speakers show are the result of a process of grammatical reconfiguration caused by low patterns of language activation and use for production and comprehension purposes. Thus, it is not an issue of lack or loss of grammatical representations due to reduced input, but rather a gradual and dynamic process of language reconfiguration stemming from low lexical activation.

Following this line of thought, Sánchez (2019) has proposed the Bilingual Alignment Approach in order to account for heritage speakers' difficulties. This approach argues that the syntactic and morphological features from the two languages undergo a process of reorganization in the bilingual mind by which new associations between syntactic and/or morphological features in Language A and Language B may emerge. While the Feature Reassembly Hypothesis is focused only on how this process takes place at the representational level, the Bilingual

Alignments Approach encompasses the temporary storage of these new associations (which may be fleeting), and the longer-term associations which may become part of the grammatical representation. Bilingual alignments start out as temporary units stored in memory with information from different grammatical modules, including syntax, phonology, semantics and discourse that overtime may become stable representations.

One important characteristic of bilingual alignments is that they are permeable. That is, they allow for features from one of the languages to bleed into an association of features in the other language. An example of how bilingual alignments work is provided in Giancaspro and Sánchez (2021) based on work by López Otero (2019). López Otero (2019) found evidence of bidirectional effects in the production and intuition of Differential Object Marking (DOM) with demonstrative pronouns in Romanian-Spanish bilinguals. DOM is the use of a special marker for an internal argument. While it has been linked to multiple factors (e.g. Aissen, 2003; Leonetti, 2008; Rodríguez-Mondoñedo, 2008), animacy has been one of the most widely recognized ones (Leonetti, 2008). This is the case in Romanian and Spanish. However, in Romanian it is possible to make an inanimate demonstrative with DOM (e.g. *Mihai a vizitat-o **pe** aceasta* 'Mihai has visited DOM this one') (e.g. Mardale, 2008; López-Otero, 2019; Ticio & Avram, 2015).[5]

This has been accounted for as a case in which the feature referential stability is involved and ranked higher than animacy. Lopez Otero's Romanian-Spanish bilingual data show cases in which DOM is marked with inanimate demonstratives in Spanish following the Romanian pattern (e.g. **David ha visitado **a** esta* 'David has visited DOM this one') and cases in which DOM is not marked with inanimate demonstrative in the Romanian of bilingual speakers (**Mihai a vizitat **Ø** aceasta* 'Mihai has visited DOM this one'). The overgeneralization of DOM marking with inanimate demonstratives in Spanish suggests transfer from Romanian into Spanish due to alignment with the referential stability feature in Romanian regardless of the Spanish animacy features; this is an instance of the emergence of a new Spanish Demonstrative Alignment permeable to the Romanian Demonstrative Alignment. On the other hand, the lack of DOM marking in Romanian with inanimate demonstratives suggests transfer from Spanish into Romanian due to alignment with the animacy feature in Spanish; this is an instance of the emergence of a new feature in the Romanian Alignment based on the Spanish Demonstrative Alignment. These two types of bilingual alignments illustrated in (1) and (2) below shows how the notion of alignments works. In (1), the referential stability feature of Romanian *aceasta*

5. Examples taken from Lopez-Otero (2019). In these particular examples, the demonstrative pronoun 'this' refers to 'gallery'.

'this' has permeated the set of features that are associated with the phonological form (PF) *este* 'this' in Spanish and is ranked higher than animacy. In (2), the animacy feature of Spanish is ranked higher in the alignment that corresponds to *aceasta:*

(2) Bilingual Spanish Demonstrative Alignment permeable to the Romanian Alignment
PF *este* ('this')
Features: [+ Referential Stability] …. [+ animate]

(3) Bilingual Romanian Demonstrative Alignment permeable to the Spanish Alignment
PF *aceasta* ('this')
Features: [+ animate] …. [+ Referential Stability]

Unlike other proposals, these alignments are not assumed to be necessarily part of the grammatical representation of the bilingual. If we find low levels of acceptance and production of these alignments, we can assume that they are transient forms that emerge as the bilinguals process the language. However, if high levels of production and acceptance are found, then it would be reasonable to assume they have become part of the bilingual's grammatical representation. It is also possible to find grammatical asymmetries between production and comprehension which suggests that the divergences are still not part of the internal representation. The extent to which heritage speakers are going to access these alignments varies depending on their specific patterns of language activation and use.

The alignments are thus flexible and permeable in nature, allowing for the coexistence of multiple alignments depending on the level of lexical activation and the characteristics of the bilingual community. Bilingual speakers who activate their native grammars more frequently as part of their day-to-day interaction with bilinguals and monolingual speakers in their communities are going to show alignments similar to those of native speakers. However, heritage speakers who activate their dominant language more often than their heritage language may exhibit linguistic patterns in production and comprehension of both innovative and non-innovating alignments *vis-à-vis* the baseline. This, however, is not an indication of either multiple grammars (Amaral & Roeper, 2014) or incomplete/unpaired representation. Contrary to an incomplete acquisition approach, the crucial argument behind a bilingual alignment approach is that the linguistic divergences that heritage speakers often show are flexible depending on the specific patterns of language activation for production and comprehension purposes. The divergences do not represent necessarily an irreversible outcome stemming from reduced input during childhood, as proposed by input based approaches.

In relation to grammatical gender specifically, intense exposure to English in the lifespan of the bilingual speaker might cause Spanish heritage speakers to

produce utterances that are more aligned with the [−strong] gender features of the English type. The more these alignments are activated in online production and comprehension, the stronger the possibility of grammatical reassembly might be, but this is not necessarily the outcome (Putnam & Sánchez, 2013). Thus, frequency of activation plays a pivotal role in the acquisition of gender as well as in other aspects of the heritage grammar including mood selection or Differential Object Marking (e.g. Hur, 2020; Hur et al. 2020; Giancaspro, 2017; López-Otero, 2019).

4. Previous research in the acquisition of grammatical gender

Although grammatical gender might appear easy to master, as it is acquired relatively early by monolingual speakers, heritage speakers of Spanish often struggle with it in production and comprehension (e.g. Alarcón, 2011; Cuza & Pérez-Tattam, 2016; Montrul et al., 2008). This is thought to be primarily due to crosslinguistic influence from English as well as reduced exposure and use of Spanish as heritage language in a language contact situation.

Previous work has found overgeneralization of the masculine form to contexts in which the feminine form was required. Researchers have argued that these divergences might stem from various sources, including incomplete acquisition during childhood due to reduced input and L1 attrition (e.g. Montrul & Potowski, 2007; Montrul et al., 2008), difficulties with form meaning mappings (e.g. Alarcón, 2011) or issues related to feature reanalysis and reconfiguration (e.g. Cuza & Pérez-Tattam, 2016; Scontras et al., 2018; Polinsky & Scontras, 2019; Putnam & Sánchez, 2013).

Montrul et al. (2008) was one of the first studies to examine the knowledge that adult heritage speakers of Spanish have of grammatical gender and how it compares to L2 learners' internal representation. Specifically, the authors investigated the production and comprehension of grammatical gender in 69 adult heritage speakers and 72 L2 learners as well as the role of age of onset of bilingualism. Results from an oral picture description task, a written recognition task, and a written comprehension task showed overgeneralization of the masculine gender among both groups, confirming previous work in L2 acquisition (e.g. Franceshina, 2001; White et al., 2004). The results also showed advantages by group depending on the task type. The heritage speakers outperformed the L2 learners in oral production, while the L2 learners outperformed the heritage speakers in written comprehension. The authors' account for the task effects is based on the age of onset of bilingualism and the context of learning: the heritage speakers normally showed better oral skills than L2 learners since they have been exposed to the language from early childhood. In contrast, the L2 learners received explicit instruction and developed better metalinguistic skills leading to higher performance in written production

or interpretation. The authors argued that the heritage speakers have undergone incomplete acquisition due to variable and insufficient input.

In a subsequent study comparing L2 learners and heritage speakers, Montrul, de la Fuente, and Foote (2014) also found an advantage by the heritage speakers over the L2 learners in their correct marking of gender agreement with diminutives in non-canonical nouns. The authors tested the potential role of language experience via an elicited production task. Results showed an advantage for the heritage speakers over the L2 learners. However, both groups were more accurate on gender with masculine nouns and showed more mismatch errors with feminine nouns compared to the native speakers acting as baseline. Furthermore, the gender mismatches with feminine nouns were more significant with non-canonical ending nouns. The authors argued that the differences between the two groups can be attributed to their input experience since early age, rather than to lack of abstract gender features before or after the critical period. In contrast to the L2 learners exposed to visual and aural input in an instructed setting, the heritage speakers had been more exposed to aural input since early childhood through interactions with family members and caregivers which affects their processing strategies.

The comprehension and production of gender agreement in heritage and L2 Spanish was further examined by Alarcón (2011) via a written recognition task and an oral description task. As in the case of Montrul et al. (2008), Alarcón found overgeneralization of the masculine form to feminine nouns, and more mismatches with gender assignment than with agreement. She also found an advantage for the heritage speakers over the L2 learners in oral production, but no differences between the two groups in written comprehension, in contrast to Montrul et al.'s (2008) results. Rather than arguing for incomplete acquisition leading to a lack of internal representation, Alarcón (2011) contended that both L2 learners and heritage speakers have abstract knowledge of gender features in Spanish despite impaired performance due to difficulties in the target form/meaning mapping (processing problem) and age of onset of acquisition effects.

Regarding gender agreement in code-switched utterances, Valenzuela et al. (2012) examined the extent to which heritage speakers diverged from L2 learners via a sentence selection task. The task involved gender assignment within code-mixed utterances in determiners phrases (DP condition) (*el book* 'the book') and gender agreement in copula constructions (*The party fue fantástica* 'the party was fantastic'). Results showed no significant differences among the two experimental groups and the controls with copula constructions (Noun/Adjective Agreement). However, with the DP condition, the heritage speakers chose either masculine or feminine determiners for nouns with feminine Spanish equivalents (*el*-MASC/*la*-FEM *chair* "the chair"). In these cases, they overextended the masculine form of the determiner with feminine Spanish equivalents significantly more than the controls. The

authors argue that the heritage speakers might be treating the codeswitched items as borrowings, and thus overgeneralize the default masculine gender marking. The two experimental groups showed target-like behavior with the use of the masculine determiner with nouns that have a masculine translation (*el book* 'the book').

In the particular case of child heritage speakers, Montrul and Potowski (2007) found that Spanish heritage children (simultaneous and sequential bilinguals) aged 6–11 produced more gender mismatches than monolingual children matched by age (5% error with the gender of determiners in the oral narrative task and 30% error rate on gender agreement with adjectives in the picture description task). The divergences were related to overgeneralization of the masculine form on the determiner and adjectives. Regarding age of onset of bilingualism, the authors found different results depending on task type. In the oral narrative, the simultaneous and sequential bilingual children did not diverge from each other, and both groups outperformed the child L2 learners. In the picture description task, the sequential bilinguals were more target-like than the simultaneous bilingual children, but they did not diverge from the child L2 learners, suggesting incomplete development (Montrul, 2002, 2004).

More recently, Cuza and Pérez-Tattam (2016) examined the acquisition of grammatical gender among 32 Spanish/English bilingual children born and raised in the U.S. Results from a picture naming task designed to elicit DET+N+ADJ strings with non-canonical nouns showed gender mismatches with the determiner (**un*-MASC *llave*-FEM *chiquita*-FEM 'a small key'), the adjective (*una*-FEM *nube*-FEM **rojo*-MASC 'a red cloud') or both (**un*-MASC *nube*-FEM *rojo*-MASC 'a red cloud'). Most of the mismatches stemmed from overgeneralization of the masculine representation on the determiner or the adjective with feminine nouns. Furthermore, the bilingual children showed significant N/ADJ word order errors (*un *chiquito tren* 'a small tren') due to transfer from English word order patterns. There was no interaction between performance and age since younger children behaved similarly to older children. Following Putnam and Sánchez's (2013) feature reassembly approach, the authors argue that the morphosyntactic asymmetries observed stem from restructuring of the L1 gender features and the development of a new featural matrix that instantiates both L1 and L2 properties.

Scontras et al. (2018) examined the comprehension of gender and number agreement in 75 English-dominant adult heritage speakers of Spanish to investigate whether the learners bundle or split gender and number features in gender attraction contexts. Results from an acceptability judgment task (AJT) showed grammatical divergences among the heritage speakers compared to the native Spanish baseline. Regarding number agreement, the heritage speakers accepted ungrammatical sentences without number agreement between a plural head noun and a singular adjective, and thus appear to be losing sensitivity to the singular. The

authors argued that the heritage speakers are moving away from the multi-value native norm and the singular marking emerges as a kind of default (restructuring of the Number category to a single value featural representation). Regarding gender, the heritage speakers showed significant divergences with feminine head nouns. The masculine form of adjectives served as the default gender specification (no feminine marking). The authors contended that heritage speakers of Spanish have reanalyzed their gender and number systems and have bundled them together in a single projection in contrast with the multi-value opposition (feature splitting) of the native baseline. They conclude that the results support the hypothesis of heritage grammar restructuring due to pressures from representational economy.

Similar grammatical restructuring was observed by Mitrofanova et al. (2018) among 54 Russian heritage children (age range 4;0–10;2) living in Norway. The authors examined the role of dominance, proficiency, language exposure and family type in target gender agreement use with existing nouns and nonce words and tested both transparent and opaque gender cues. Results from two picture-based elicitation tasks showed that Russian heritage children are sensitive to morphological cues for gender assignment, but still overextended the masculine form to cases where the feminine was required in oral production, crucially with neuter and opaque nouns. Results also showed a correlation with family type: children from Russian-speaking parents overused feminine forms with vowel-final nouns, especially among the older siblings. The authors concluded that the divergences stem primarily from cumulative length of exposure to Russian, consistency of Russian input in the family, the presence of older siblings, and lexical diversity as defined by the number of different words in the Russian narrative. There was no effect of bilingual dominance, corroborating previous work with Spanish children (e.g. Cuza & Pérez-Tattam, 2016).

The results of Mitrofanova et al.'s (2018) study support previous findings in other studies testing the acquisition of gender agreement in Russian. For example, Polinsky (2008) investigated the production and comprehension of gender assignment and noun categorization in a group of 12 Russian heritage speakers born in the US. In Experiment 1, the participants had to produce an unambiguous adjective or possessive pronoun after hearing a particular noun. The results showed overgeneralization of the masculine form with neuter nouns (end-stressed) and with feminine nouns ending in a palatalized consonant. This was primarily the case among the heritage speakers with significantly lower speech rate and among the less proficient speakers with clear difficulty in their lexical access. The speakers with higher lexical proficiency behaved closer to the native baseline. The author concluded that the less proficient Russian heritage speakers have reanalyzed their tripartite-gender system in Russian into a two-gender system: the neuter nouns were absorbed into the feminine class, the nouns ending in a consonant were treated as masculine, and

the nouns ending in a vowel were treated as feminine. Experiment 2 (comprehension task) included ADJ/N pairs presented auditorily; the participants were asked to push a button if the pair they heard was an acceptable combination in Russian. The results showed overgeneralization of the feminine form to neuter nouns (about 60%) among the less proficient learners, corroborating the findings of Experiment 1 (neuter-to-feminine reanalysis). Polinsky (2008: 3) suggested that Russian heritage speakers have reanalyzed their grammatical gender system in the absence of declensional information: "the system is different from its correspondent system in the baseline, but it is still coherent, albeit in its own way".

Regarding the effects of lexical frequency, Hur et al (2020) provided substantial evidence showing the impact of lexical frequency and vocabulary knowledge on the representation of gender assignment and agreement in heritage Spanish. They investigated a group of 39 Spanish heritage speakers via an elicited production task and a forced preference task examining gender assignment and agreement respectively. The participants also completed a self-rating lexical frequency task and the MiNT productive vocabulary task. Results showed more target performance with high-frequency lexical items in both production and comprehension. However, the frequency and vocabulary effects were mostly present in the production task, which tested gender assignment. This was evidenced in an overgeneralization of the masculine form with feminine nouns. The comprehension tasks showed an effect of lexical frequency but not vocabulary knowledge in the overgeneralization of the feminine form. The authors concluded that both lexical frequency and vocabulary knowledge affect target gender assignment in production; however, comprehension of gender agreement appears to be affected only by lexical frequency.

To summarize, previous acquisition research has indicated significant difficulties with the target production and interpretation of gender assignment and agreement in child and adult heritage speakers of Spanish. Results have shown overgeneralization of the masculine form to contexts where the feminine should be used with determiners and adjectives. Regarding the source(s) of the difficulties, some researchers have argued in favor of incomplete acquisition during childhood due to reduced input and use (e.g. Montrul et al., 2008; Montrul & Potowski, 2007; Montrul et al., 2008). However, others depart from the notion of incompleteness and suggest processing issues, L1 attrition or grammatical reanalysis in the life span of the speaker (e.g. Cuza & Pérez-Tattam, 2016; Scontras et al., 2018; Polinsky & Scontras, 2019). More recently, researchers have also argued for a Bilingual Alignment Approach (e.g. Sánchez, 2019). As we discuss below, this proposal goes beyond the source of differences by focusing on the actual process, and the role of language activation and proficiency in the extent of morphosyntactic shifts (e.g. Hur, 2020; Hur et al. 2020; Lopez-Otero, 2019; Pérez-Cortes, Putnam, & Sánchez, 2019).

5. Beyond the source of differences: The bilingual alignment approach

Most of the work discussed so far on the acquisition of grammatical gender has generally centered on the sources of heritage speakers' difficulties, including incomplete acquisition, grammatical reanalysis or L1 attrition in the lifespan (e.g. Montrul et al., 2008; Polinsky, 2008). Although this previous work has made significant contributions in relation to the acquisition of gender in heritage language grammars, and heritage language acquisition theorizing in general, it has left unanswered important questions that need to be dealt with in order to have a better understanding of both the processes and the outcomes of heritage language development. If we accept the argument that heritage language grammars are incomplete or attritted due to various external and internal factors, then it is unclear what remains beyond that state of incompleteness or attrition. Is the acquisition process irreversibly halted? This argument becomes unclear when dealing with young heritage children who are still undergoing language development of both their majority and minority languages (e.g. Anglin, 1993; Berman, 2007; Hoff & Core, 2015). Furthermore, as discussed by Polinsky and Scontras (2019), child language learners are still developing their cognitive abilities while learning both languages; thus, they might not yet be completely sensitive to grammatical areas involving pragmatic or semantic knowledge (e.g. Hulk & Müller, 2000; Serratrice, Sorace, & Paoli, 2004; Syrett et al. 2017).

Assuming that the heritage grammar is incomplete or attritted hinders the heritage speaker's ability to move forward along in the bilingual continuum. If divergences are present, what can parents and teachers do to resolve them? What should parents do to foster heritage language transmission and arrest the process of language attrition or incompleteness whichever the case might be? What can teachers do in the classroom to facilitate the development of the minority language? These questions cannot be easily answered if the starting point is that heritage language representations are incomplete or attritted. If attrition is assumed, the process of heritage language development by parents and practitioners would have to either start from the bottom up or focus their interventions on those areas of the grammar that are argued to undergo incompleteness or attrition. However, almost the entire morphosyntactic system of Spanish heritage speakers has been argued to have undergone instances of incomplete acquisition or attrition (e.g. in tense and aspect, DOM, argument structure, gender agreement, clitic use and position, subject pronoun expression, subjunctive mood, interrogative formation) (e.g. Cuza, 2013; Montrul, 2002; Silva-Corvalán, 2014). Thus, two key questions would be: when and where should the intervention process start? and when is the child heritage speakers' grammatical system no longer incomplete or attritted?

More recently, however, researchers have investigated what lies beyond the source of difficulties and have focused more on processing issues and the role of language activation for production and comprehension (Putnam & Sánchez, 2013; Sánchez, 2019). One example of this is the work by Sánchez (2019) on bilingual language activation, known in the literature as the Bilingual Alignment Approach. The author argues that the linguistic divergences heritage speakers often have are dependent on the way that the features in the two languages are aligned and stored for lexical access. Linguistic divergences are then not viewed as the result of impaired representation, multiple grammars, or incomplete acquisition due to reduced input during childhood. Instead, they are considered to be part of an on-going process of *on* and *off* language activation for production and comprehension purposes. Multiple alignments can coexist depending on the type of activation for comprehension or production. Heritage speakers may have access to monolingual-like alignments in comprehension, but they may access other alignments stored in temporary memory that integrate components from the non-heritage language in production.

Rather than proposing that heritage speakers have access to different grammars for comprehension and production, the alignments proposal allows for the coexistence of stable units and temporary units stored in memory. If the temporary units are accessed frequently, some grammatical features in the heritage language might get re-assembled after significant alignment with the features of the stronger language. The linguistic outcome is therefore the byproduct of real-time linguistic processing, and the result is fluid and dynamic (e.g. Pérez Cortes et al., 2009). In other words, what determines the linguistic outcome of heritage speakers' grammars at a specific point in time is precisely their differential access to storage units in the heritage language and their specific levels of language activation and use at different levels of proficiency. This can be evidenced in the heritage speaker's variability with various morphosyntactic structures, (e.g. tense and aspect, subjunctive mood, DOM, object clitic use, gender agreement) as has been recently demonstrated.

Giancaspro and Sánchez (2021) used Sánchez's (2019) approach while investigating the variability Spanish heritage speakers have with inalienable possession (e.g. *Rosa rompió su brazo* "Rosa broke her arm"). Results from an elicited production task administered to 30 adult heritage speakers at different proficiency levels showed significant clitic omission and heavy reliance on possessive pronouns. However, the heritage speakers showed target-like behavior in their grammatical intuition measured via an AJT. The authors account for these conflicting results in terms of transfer from English at the level of bilingual alignments, language proficiency and task modality. Only when bilingual speakers show significant divergences in both production and comprehension, is it logical to consider that their

linguistic representation has changed (e.g. Putnam & Sánchez, 2013). This proposal accounts for the tasks effects we often find between production and interpretation (e.g. Cuza & Frank, 2014; Hur et al. 2020; López Otero, 2019) and presents a more precise linguistic account of performance vs. competence errors.

A bilingual alignment approach could shed light on the data previous work has found in relation to gender agreement in Spanish. For example, Montrul et al. (2008) found more target-like production of gender agreement in oral production (around 80%) among the heritage speakers. However, the L2 learners outperformed the heritage speakers in written comprehension. These divergences can be accounted for in terms of language proficiency and, among the heritage speakers, difficulties accessing alignments in a written comprehension task. This is a linguistic skill that heritage speakers of Spanish do not appear to have developed much in their native language as their education has been almost entirely in English (e.g. Cuza et al., 2017; Rothman, 2007). Thus, it is not unusual for them to show more transfer effects from English in this particular modality depending on their linguistic proficiency and literacy skills. The L2 learners, in contrast, have undergone years of classroom written comprehension.

As discussed earlier, other studies on the acquisition of gender have shown similar patterns. Alarcón (2011) found more target performance among the heritage speakers in oral production, but no difficulties between the two groups with written comprehension. As argued by Alarcón (2011), the divergences in oral production are not necessarily from lack of grammatical representation but rather processing issues. Processing difficulties can also account for the divergences found by Valenzuela et al. (2012) in which the heritage speakers overextended the masculine form with switched DP constructions but they showed no divergences with copula constructions. Following Sánchez's (2019) account, processing difficulties of this sort could be accounted for in terms of permeability to English features at the level of bilingual alignments at the DP level. This would explain not only the asymmetries we often find across tasks (production vs. interpretation), but also the asymmetries we find across various contexts and conditions. For example, research examining the development of copula distinctions in child and adult heritage Spanish has found that adult heritage speakers exhibit more target-like behavior than child bilinguals in the target use of copula *ser* in *estar*-required adjectival contexts. However, with event locatives, both child and adult heritage speakers of Spanish significantly overextend copula *estar* in a context in which *ser* is required (e.g. *La boda *está en la iglesia* 'The wedding is at the church') (e.g. Cuza, Reyes, & Lustres, 2021; Sera, 1992; Silva-Corvalán & Montanari, 2008). Thus, it is not the case that copula selection is underspecified in the heritage grammar; rather, copula selection in certain contexts undergoes more variability than in others depending on various external and internal factors, including patterns of language activation and

experience (e.g. Daskalaki, Blom, Chondrigianni, & Paradis, 2020; Thordardotirr, 2017; Unsworth, 2016), literacy development (e.g. Oller & Eilers, 2002) as well as cognitive resources (e.g. Arunachalam & He, 2018; Bylund, 2009; Kuhl et al 2005).

In addition to what we have seen in Spanish, the results found by Polinsky (2008) with Russian also shed light on a bilingual alignment approach. Polinsky found divergences in both production and interpretation among the least proficient heritage speakers with clear difficulty in their lexical access. These speakers have undergone intense exposure to English in their lifespan, and they have therefore developed bilingual alignments of gender features compatible with [–strong] English type features. The constant activation of these alignments for production and comprehension have led to a grammatical restructuring of the Russian gender features or to what Polinsky calls grammatical reanalysis.

In sum, the Bilingual Alignment Approach brings together previous research in language acquisition and bilingualism to provide a more holistic account for heritage speakers' differences. The approach is optimal in that it accounts for heritage speakers' differences in terms of real-time language processing issues stemming from (a) specific patterns of language activation and use; (b) lexical frequency and (c) linguistic proficiency. These three triggers are argued to modulate the extent to which aspects of the Spanish heritage grammar are going to align with English features and eventually get restructured or reassembled. Within this view, parents and teachers can play a significant role in the process of heritage language maintenance and transmission not in the form of arresting the attrition process or teaching unacquired/fossilized rules, but rather in aligning the learning process toward the minority language via specific tasks that tackle the above-mentioned triggers (language activation, lexical frequency and proficiency). Together with the development of writing and reading skills, parents, instructors and the community at large should emphasize oral practice and communicative abilities in the minority language to activate existing grammatical knowledge.

Furthermore, a Bilingual Alignment Approach is advantageous in that it provides a linguistic framework for the asymmetries often found between production and comprehension. As recently documented, heritage speakers are more likely to show more divergences with agreement in production than in comprehension given the processing cost of online production and consequently more alignment with English or the dominant language in question. At the same time, the Bilingual Alignment Approach has the limitation of having to empirically determine when a transient alignment becomes a stable representation. In that respect, unlike the Feature Reassembly Hypothesis, the Bilingual Alignments Approach relies on a gradient and to some extent probabilistic perspective to determine when a feature becomes part of a representation. Future testing of the BAA will have to determine the extent to which correlations hold between stable alignments and the factors

mentioned above such as patterns of language activation and use, lexical frequency, and linguistic proficiency.

6. Conclusion

The goal of the present chapter was to provide a review of current work on the acquisition of gender agreement in heritage Spanish focusing on work that has found evidence of outcomes that differ from baseline patterns. Accounts for this difference include explanations such as incomplete acquisition, attrition, and restructuring. We have argued that the divergences heritage speakers exhibit with gender agreement in Spanish can be better explained in terms of Sánchez's (2019) proposal on bilingual alignments. Sánchez's model provides a linguistically grounded explanation for the divergences heritage speakers show depending on task modality (production vs. comprehension). Furthermore, this approach accounts for the variability we often find at the individual level among heritage speakers with different patterns of language experience. The approach is optimal in that it allows for subsequent morphosyntactic shifts in the bilingual continuum depending on specific patterns of language activation and use. Most importantly, Sánchez's approach predicts when the difficulties or divergences that heritage speakers experience due to various factors become representational changes in the heritage language grammar, leading to restructuring or feature reassembly. Therefore, this approach provides insights not only on the source of the divergences but also on the process of heritage language development in general. It also has limitations such as the need to empirically determine when a feature becomes part of a stable representation. Furthermore, more studies testing the effects of patterns of language activation and use, lexical frequency, and linguistic proficiency on alignments is needed to provide support for this approach.

References

Aissen, J. (2003). Differential object marking: Iconicity vs. economy. *Natural Language & Linguistic Theory*, 21, 435–483. https://doi.org/10.1023/A:1024109008573

Alarcón, I. (2009). The processing of gender agreement in L1 and L2 Spanish: Evidence from reaction time data. *Hispania*, 92(4), 814–828.

Alarcón, I. (2011). Spanish gender agreement under complete and incomplete acquisition: Early and late bilinguals' linguistic behavior within the noun phrase. *Bilingualism: Language and Cognition*, 14(3), 332–350. https://doi.org/10.1017/S1366728910000222

Amaral, L., & Roeper, T. (2014). Multiple grammars and second language representation. *Second Language Research*, 30, 3–36. https://doi.org/10.1177/0267658313519017

Anglin, J. M. (1993). Vocabulary development: A morphological analysis. *Monographs of the Society for Research in Child Development*, 58(10, Serial No. 238). https://doi.org/10.2307/1166112

Arunachalam, S., & He, A. X. (2018). Children's acquisition of nouns that denote events. In A. B. Bertolini & M. J. Kaplan (Eds.), *Proceedings of the 42nd Annual Boston University Conference on Language Development* (pp. 29–44). Cascadilla Press.

Berman, R. A. (2007). Developing linguistic knowledge and language use during adolescence. In E. Hoff & M. Shatz (Eds.), *Blackwell handbook of language development* (pp. 347–367). Blackwell. https://doi.org/10.1002/9780470757833.ch17

Bernstein, J. (1993). Topics in the syntax of nominal structures across Romance (Unpublished PhD dissertation). City University of New York.

Bosque, I., & Picallo, C. (1996). Postnominal adjectives in Spanish DPs. *Journal of Linguistics*, 32, 349–385. https://doi.org/10.1017/S0022226700015929

Bruhn de Garavito, J., & White, L. (2002). The second language acquisition of Spanish DPs: The status of grammatical features. In A. T. Pérez-Leroux & J. Liceras (Eds.), *The acquisition of Spanish morphosyntax: The L1/L2 connection* (pp. 153–178). Kluwer. https://doi.org/10.1007/978-94-010-0291-2_6

Bylund, E. (2009). Maturational constraints and first language attrition. *Language Learning*, 59, 687–715. https://doi.org/10.1111/j.1467-9922.2009.00521.x

Carstens, V. (1991). The morphology and syntax of determiner phrases in Kiswahili (Unpublished PhD dissertation). University of California.

Cinque, G. (1994). On the evidence for partial N-movement in the Romance DP. In G. Cinque, J. Koster, J. Y. Pollock, L. Rizzi, & R. Zanuttini (Eds.), *Paths towards Universal Grammar: Studies in Honor of Richard S. Kayne* (pp. 85–110). Georgetown University Press.

Cinque, G. (2010). *The syntax of adjectives*. The MIT Press. https://doi.org/10.7551/mitpress/9780262014168.001.0001

Cuza, A. (2013). Cross-linguistic influence at the syntax proper: Interrogative subject-verb inversion in heritage Spanish. *The International Journal of Bilingualism*, 17(1), 71–96. https://doi.org/10.1177/1367006911432619

Cuza, A. (2016). The status of interrogative subject-verb inversion in Spanish–English bilingual children. *Lingua*, 180, 124–138. https://doi.org/10.1016/j.lingua.2016.04.007

Cuza, A., & Frank, J. (2014). On the role of experience and age-related effects: Evidence from the Spanish CP. *Second Language Research*, 31, 3–28. https://doi.org/10.1177/0267658314532939

Cuza, A., Miller, L., Pasquarella, A., & Chen, X. (2017). The role of literacy instruction in the development of Spanish as a heritage language during childhood. *The Heritage Language Journal*, 14(2), 100–123. https://doi.org/10.46538/hlj.14.2.1

Cuza, A., & Perez-Tattam, R. (2016). Grammatical gender selection and phrasal word order in child heritage Spanish: A feature re-assembly approach. *Bilingualism: Language and Cognition*, 19(3), 332–350.

Cuza, A., Pérez-Tattam, R., Barajas, E., Miller, L., & Sadowski, C. (2013). The development of tense and aspect morphology in child and adult heritage speakers. In J. W. Schwieter (Ed.), *Innovative research and practices in second language acquisition and bilingualism* (pp. 193–219). John Benjamins. https://doi.org/10.1075/lllt.38.12cuz

Cuza, A., Reyes, N., & Lustres, E. (2021). The production of ser and estar in child and adult heritage speakers of Spanish. *Lingua*, 249. https://doi.org/10.1016/j.lingua.2020.102978

Daskalaki, E., Blom, E., Chondrogianni, V., & Paradis, J. (2020). Effects of parental input quality in child heritage language acquisition. *Journal of Child Language*, 47(4), 709–736. https://doi.org/10.1017/S0305000919000850

Demonte, V. (1999). *El adjetivo: Classes y usos. La posición del adjetivo en el sintagma nominal*. In I. Bosque & V. Demonte (Eds.), *Gramática descriptiva de la lengua español* (pp.129–215). Espasa Calpe.

Domínguez, L., Hicks, G., & Slabakova, R. (2019). Terminology choice in generative acquisition research: The case of "incomplete acquisition" in heritage language grammars. *Studies in Second Language Acquisition*, 41(2), 241–255. https://doi.org/10.1017/S0272263119000160

Flores, C. (2014). Losing a language in childhood: A longitudinal case study on language attrition. *Journal of Child Language*, 42(3), 562–590. https://doi.org/10.1017/S0305000914000233

Franceschina F.(2001). Morphological or syntactic deficits in near-native speakers? An assessment of some current proposals. *Second Language Research*. 2001;17(3):213–247. https://doi.org/10.1177/026765830101700301

Gathercole, V. C. (2002). Grammatical gender in bilingual and monolingual children: A Spanish morphosyntactic distinction. In K. Oller & R. Eilers (Eds.), *Language and literacy in bilingual children* (pp. 207–219). De Gruyter. https://doi.org/10.21832/9781853595721-010

Giancaspro, D. J. (2017). Heritage speakers' production and comprehension of lexically-and contextually-selected subjunctive mood morphology (Unpublished PhD dissertation). Rutgers University.

Giancaspro D. & Sánchez L. (2021). Me, mi, my: Innovation and variability in heritage speakers' knowledge of inalienable possession. *Glossa: a journal of general linguistics*, 6(1): 31. 1–28. https://doi.org/10.5334/gjgl.1240

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

Gürel, A. (2004). Selectivity in L2-induced L1 attrition: A psycholinguistic account. *Journal of Neurolinguistics*, 17, 53–78. https://doi.org/10.1016/S0911-6044(03)00054-X

Harris, J. (1991). The exponence of gender in Spanish. *Linguistic Inquiry*, 22, 65–88.

Hoff, E., & Core, C. (2015). Input and language development in bilingually developing children. *Seminars in Speech and Language*, 34(2), 215–226. https://doi.org/10.1055/s-0033-1353448

Hulk, A., & Müller, N. (2000). Bilingual first language acquisition at the interface between syntax and pragmatics. *Bilingualism: Language and Cognition*, 3, 227–244. https://doi.org/10.1017/S1366728900000353

Hur, E. (2020). Verbal lexical frequency and DOM in heritage speakers of Spanish. In A. Mardale & S. Montrul (Eds.), *The acquisition of differential object marking* (pp. 207–235). John Benjamins. https://doi.org/10.1075/tilar.26.08hur

Hur, E., Lopez Otero, J. C., & Sanchez, L. (2020). Gender agreement and assignment in Spanish heritage speakers: Does frequency matter? *Languages*, 5, 48. https://doi.org/10.3390/languages5040048

Jia, R., & Paradis, J. (2015). The use of referring expressions in narratives by Mandarin heritage language children and the role of language environment factors in predicting individual differences. *Bilingualism: Language and Cognition*, 18(4), 737–752. https://doi.org/10.1017/S1366728914000728

Kuhl, P., Conboy, B., Padden, D., Nelson, T., & Pruitt, J. (2005). Early speech perception and later language development: Implications for the "critical period". *Language Learning and Development*, 1, 237–264. https://doi.org/10.1207/s15473341lld0103&4_2

Kupisch, T., & Rothman, J. (2018). Terminology matters! Why difference is not incompleteness and how early child bilinguals are heritage speakers. *International Journal of Bilingualism*, 22(5), 564–582. https://doi.org/10.1177/1367006916654355

Lamarche, J. (1991). Problems for N°-movement to Num-P. *Probus,* 3(2), 215–236. https://doi.org/10.1515/prbs.1991.3.2.215

Lardiere, D. (1998). Dissociating syntax from morphology in a divergent L2 end-state grammar. *Second Language Research*, 14(4), 359–375. https://doi.org/10.1191/026765898672500216

Lardiere, D. (2008). Feature-assembly in second language acquisition. In J. Liceras, H. Zobl, & H. Goodluck (Eds.), *The role of formal features in second language acquisition* (pp. 106–140). Lawrence Erlbaum Associates.

Lardiere, D. (2009). Some thoughts on the contrastive analysis of features in second language acquisition. *Second Language Research*, 25(2), 173–227. https://doi.org/10.1177/0267658308100283

Leonetti, M. (2008). Specificity in clitic doubling and in differential object marking. *Probus*, 20, 35–69. https://doi.org/10.1515/PROBUS.2008.002

López-Ornat, S., Fernández, A., Gallo, P., & Mariscal, S. (1994). *La adquisición de la lengua española*. Siglo XXI.

López-Otero, J. C. (2019). On the acceptability of the Spanish DOM among Romanian-Spanish bilinguals. In A. Mardale & S. Montrul (Eds.), *The acquisition of differential object marking* (pp.161–181). John Benjamins. https://doi.org/10.1075/tilar.26.lop06

López-Otero, J. C. (2019). On the acceptability of the Spanish DOM among Romanian-Spanish bilinguals (Unpublished manuscript). Rutgers University,

Mardale, A. (2008). Microvariation within differential object marking: Data from Romance. *Revue Roumaine de Linguistique*, LIII(4), 449–467.

Mariscal, S. (2009). Early acquisition of gender agreement in the Spanish noun phrase: starting small. *Journal of Child Language* 36: 143–171. https://doi.org/10.1017/S0305000908008908

Mitrofanova, N., Rodina, Y., Urek, O., & Westergaard, M. (2018). Bilinguals' sensitivity to grammatical gender cues in Russian: The role of cumulative input, proficiency, and dominance. *Frontiers in Psychology*, 9, 18–94. https://doi.org/10.3389/fpsyg.2018.01894

Montrul, S. (2002). Incomplete acquisition and attrition of Spanish tense/aspect distinctions in adult bilinguals. *Bilingualism: Language and Cognition*, 5, 39–68. https://doi.org/10.1017/S1366728902000135

Montrul, S. (2004). Subject and object expression in Spanish heritage speakers: A case of morphosyntactic convergence. *Bilingualism: Language and Cognition*, 7(2), 125–142. https://doi.org/10.1017/S1366728904001464

Montrul, S. (2008). *Incomplete acquisition in bilingualism. Re-examining the age factor*. John Benjamins. https://doi.org/10.1075/sibil.39

Montrul, S. (2009). Incomplete acquisition of tense-aspect and mood in Spanish heritage speakers. *The International Journal of Bilingualism*, 13(3), 239–269. https://doi.org/10.1177/1367006909339816

Montrul, S. (2016). *Heritage language acquisition*. Cambridge University Press. https://doi.org/10.1017/CBO9781139030502

Montrul, S., Davidson, J., De La Fuente, I., & Foote, R. (2014). Early language experience facilitates the processing of gender agreement in Spanish heritage speakers. *Bilingualism: Language and Cognition*, 17, 118–138. https://doi.org/10.1017/S1366728913000114

Montrul, S., de la Fuente, I., Davidson, J., & Foote, R. (2013). The role of experience in the acquisition and production of diminutives and gender in Spanish: Evidence from L2 learners and heritage speakers. *Second Language Research*, 29, 87–118. https://doi.org/10.1177/0267658312458268

Montrul, S., Foote, R., & Perpiñan, S. (2008). Gender agreement in adult second language learners and Spanish heritage speakers: The effects of age and context of acquisition. *Language Learning*, 58(3), 503–553. https://doi.org/10.1111/j.1467-9922.2008.00449.x

Montrul, S., & Potowski, K. (2007). Command of gender agreement in school age Spanish English bilingual children. *International Journal of Bilingualism*, 11(3), 301–330. https://doi.org/10.1177/13670069070110030301

Montrul, S., & Sánchez-Walker, N. (2013). Differential object marking in child and adult Spanish heritage speakers. *Language Acquisition*, 20(2), 109–132. https://doi.org/10.1080/10489223.2013.766741

Oller, D. K., & Eilers, R. E. (2002). *Language and literacy in bilingual children*. Multilingual Matters. https://doi.org/10.21832/9781853595721

Otheguy, R. (2016). The linguistic competence of second-generation bilinguals: A critique of incomplete acquisition. In C. Tortora, M. den Dikken, I. L. Montoya, & T. O'Neill (Eds.), *Romance linguistics 2013* (pp. 301–320). John Benjamins. https://doi.org/10.1075/rllt.9.16oth

Pascual y Cabo, D., & Rothman, J. (2012). The (il)logical problem of heritage speaker bilingualism and incomplete acquisition. *Applied Linguistics*, 33, 450–455. https://doi.org/10.1093/applin/ams037

Pérez-Cortes, S., Putnam, M., & Sánchez, L. (2019). Differential access: Asymmetries in accessing features and building representations in heritage language grammars. *Languages*, 4(4), 81–134. https://doi.org/10.3390/languages4040081

Pérez-Leroux, A. T., Cuza, A., & Thomas, D. (2011). From parental attitudes to input condition Spanish-English bilingual development in Toronto. In K. Potowski & J. Rothman (Eds.), *Bilingual youth: Spanish in English-speaking societies*, (pp. 49–176). John Benjamins. https://doi.org/10.1075/sibil.42.10per

Pérez-Pereira, M. (1991). The acquisition of gender: What Spanish children tell us. *Journal of Child Language*, 18(3), 571–590. https://doi.org/10.1017/S0305000900011259

Polinsky, M. (2006). Incomplete acquisition: American Russian. *Journal of Slavic Linguistics*, 14, 191–262.

Polinsky, M. (2008). Gender under incomplete acquisition: Heritage speakers' knowledge of noun categorization. *Heritage Language Journal*, 6(1), 40–71. https://doi.org/10.46538/hlj.6.1.3

Polinsky, M. (2011). Reanalysis in adult heritage language: A case for attrition. *Studies in Second Language Acquisition*, 3(4), 305–328. https://doi.org/10.1017/S027226311000077X

Polinsky, M. (2018). *Heritage languages and their speakers*. Cambridge University Press. https://doi.org/10.1017/9781107252349

Polinsky, M., & Scontras, G. (2019). Understanding heritage speakers. *Bilingualism: Language and Cognition*, 23(1), 4–20. https://doi.org/10.1017/S1366728919000245

Polinsky, M., & Scontras, G. (2020). A road map for heritage language research. *BLC*, 23(1), 50–55.

Polinsky, M., & Scontras, G. (2020). A road map for heritage language research. *BLC*, 23(1), 50–55.

Putnam, M., & Sánchez, L. (2013). What's so incomplete about incomplete acquisition? *Linguistic Approaches to Bilingualism*, 3(4), 478–508. https://doi.org/10.1075/lab.3.4.04put

Rodríguez-Mondoñedo, M. (2008). The acquisition of differential object marking in Spanish. *Probus*, 20(1), 111–145. https://doi.org/10.1515/PROBUS.2008.004

Rothman, J. (2007). Heritage speaker competence differences, language change, and input type: Inflected infinitives in heritage Brazilian Portuguese. *International Journal of Bilingualism*, 11(4), 359–389. https://doi.org/10.1177/13670069070110040201

Sagarra, N., & Herschensohn, J. (2011). Proficiency and animacy effects on L2 gender agreement processes during comprehension. *Language Learning*, 61(1), 80–116. https://doi.org/10.1111/j.1467-9922.2010.00588.x

Sanchez, L. (1996). Syntactic structures in nominals: A comparative study of southern Quechua and Spanish (Unpublished PhD dissertation). University of Southern California.

Sánchez, L. (2019). Bilingual alignments. *Languages*, 4(82), 1–24. https://doi.org/10.3390/languages4040082

Schmid, M. (2002). *First language attrition, use and maintenance. The case of German jews in anglophone countries*. John Benjamins. https://doi.org/10.1075/sibil.24

Schmid, M. S., Köpke, B. (2007). Bilingualism and attrition. In *Language attrition: Theoretical perspectives*, Köpke, B., Schmid, M. S., Keijzer, M., Dostert, S. (eds.), pp. 1–8. Amsterdam/ Philadelphia: John Benjamins. https://doi.org/10.1075/sibil.33.02sch

Scontras, G., Polinsky, M., & Fuchs, Z. (2018). In support of representational economy: Agreement in heritage Spanish. *Glossa: A Journal of General Linguistics*, 3(1), https://doi.org/10.5334/gjgl.164

Sera, M.D. (1992). To be or not to be: use and acquisition of the Spanish copula. *Journal of Memory and Language*, 31, 408–427.

Serratrice, L., Sorace, A., & Paoli, S. (2004). Crosslinguistic influence at the syntax-pragmatics interface: Subjects and objects in English-Italian bilingual and monolingual acquisition. *Bilingualism: Language and Cognition*, 7, 183–205. https://doi.org/10.1017/S1366728904001610

Silva-Corvalán, C. (1991). Spanish language attrition in a contact situation with English. In H. Seliger & R. Vago (Eds.), *First language attrition* (pp.151–171). Cambridge University Press. https://doi.org/10.1017/CBO9780511620720.011

Silva-Corvalán, C. (2014). *Bilingual Language Acquisition: Spanish and English in the First Six Years*. Cambridge University Press.

Silva-Corvalán, C. (2018a). Simultaneous bilingualism: Early developments, incomplete later outcomes? *International Journal of Bilingualism*, 22, 497–512. https://doi.org/10.1177/1367006916652061

Silva-Corvalán, C. (2018b). Bilingual acquisition: Difference or incompleteness? In N. Lapidus Shin & D. Erker (Eds.), *Questioning theoretical primitives in linguistic inquiry: Papers in honor of Ricardo Otheguy* (pp. 245–268). John Benjamins. https://doi.org/10.1075/sfsl.76.11sil

Silva-Corvalán, C., & Montanari, S. (2008). The acquisition of *ser, estar* (and be) by a Spanish–English bilingual child: The early stages. *Bilingualism: Language and Cognition*, 11(3), 341–360. https://doi.org/10.1017/S136672890800357X

Sorace, A. (2000). Differential effects of attrition in the L1 syntax of near-native L2 speakers. In C. Howell, S. Fish, & T. Keith-Lucas (Eds), *Proceedings of the 24th Boston University Conference on Language Development* (pp.719–725). Cascadilla Press.

Sorace, A. (2005). Selective optionality in language development. In L. Cornips & K. Corrigan (Eds.), *Syntax and variation: reconciling the biological and the social* (pp. 55–80). John Benjamins. https://doi.org/10.1075/cilt.265.04sor

Sorace, A. (2011). Pinning down the concept of "interface" in bilingualism. *Linguistic Approaches to Bilingualism*, 1, 1–33. https://doi.org/10.1075/lab.1.1.01sor

Syrett, K., Lingwall, A., Pérez-Cortes, S., Austin, J., Sánchez, L., Baker, H., Germak, C., & Arias-Amaya, A. (2017). Differences between Spanish monolingual and Spanish-English bilingual children in their calculation of entailment-based scalar implicatures. *Glossa: A Journal of General Linguistics*, 2(1): 31. https://doi.org/10.5334/gjgl.76

Teschner, R., & Russell, W. (1984). The gender patterns of Spanish nouns: An invers dictionary-based analysis. *Hispanic Linguistics*, 1, 115–132.

Ticio, E., & Avram, L. (2015). The acquisition of differential object marking in Spanish and Romanian: Semantic scales or semantic features? *Revue Roumaine de Linguistique*, 60(4), 383–401.

Tsimpli, I., Sorace, A., Heycock, C., & Filiaci, F. (2004). First language attrition and syntactic subjects: A study of Greek and Italian near-native speakers of English. *International Journal of Bilingualism*, 8, 257–277. https://doi.org/10.1177/13670069040080030601

Unsworth, S. (2016). Quantity and quality of language input in bilingual language development. In E. Nicoladis & S. Montanari (Eds.), *Lifespan perspectives on bilingualism* (pp. 136–196). De Gruyter. https://doi.org/10.1515/9783110341249-008

Valenzuela, E., Faure, A., Ramírez-Trujillo, A. P., Barski, E., Pangtay, Y., & Díez, A. (2012). Gender and heritage Spanish bilingual grammars: A study of code-mixed determiner phrases and copula constructions. *Hispania*, 95(3), 481–494.

White, L., Valenzuela, E., Kozlowska-MacGregor, M., & Leung, Y. K. (2004). Gender and number agreement in nonnative Spanish. *Applied Psycholinguistics*, 25, 105–133. https://doi.org/10.1017/S0142716404001067

Zagona, K. (2002). The syntax of Spanish. Cambridge University Press.

PART II

Child, adult and heritage speakers

CHAPTER 5

The second language acquisition of grammatical gender and number in Italian

Dalila Ayoun and Stefano Maranzana
University of Arizona / Southern Methodist University

This cross-sectional empirical study tested the ability of Anglophone L2 learners of Italian (n = 87) to assign grammatical gender and number to a subset of isolated nouns drawn from a written corpus of current magazine and newspaper articles. Italian native speakers served as controls (n = 109). We first present a descriptive account of grammatical gender that outlines several idiosyncrasies and complexities than may lead to difficulties for L2 learners, particularly English native speakers, from a Feature Reassembly approach and a Minimalist perspective. We then discuss L2 learnability implications in light of the results of a written, computerized gender assignment task (GAT) showing significant effects for gender and number, ambiguous vs. transparent nouns, but not for suffixed vs non-suffixed nouns.

Keywords: grammatical gender, formal features, Italian, Anglophone instructed learners

1. Grammatical gender and number assignment in Italian

How do second language (L2) learners acquire grammatical gender particularly when it is not instantiated in their first language (L1) as it is the case for Anglophone learners of Italian? The L2 acquisition of grammatical gender is notoriously difficult in both instructed and immersion settings (for French see e.g. Ayoun, 2007; Dewaele, 2015; Lyster, 2014). The fact that Italian uses portmanteau morphemes, that is, vocalic endings conflating gender and number, two formal features expressed by two distinct morphemes in other Romance languages such as Spanish (Foote, 2015) or Portuguese (Dominguez, Cuetos, & Segui, 1999), creates an additional challenge (Ayoun & Maranzana, 2020).

Gender in Italian permeates the lexicon (i.e. all nouns are assigned a gender), morphology (i.e. vocalic endings, noun classes) and syntax (i.e., concord within the

https://doi.org/10.1075/sibil.63.05ayo

determiner phrase as well as the verb phrase for past participles with the auxiliaries *essere* 'be' and *avere* 'have' and object clitic pronouns). It is thus an important and unavoidable part of the L2 acquisition of instructed learners although anecdotal evidence from instructors suggests it is rarely emphasized.

We will first describe how grammatical gender is morphologically expressed in Italian, then how it is framed from a Minimalist perspective before reviewing the L2 acquisition literature and explaining the methodology we used to test its acquisition by Anglophone instructed learners.

1.1 Noun classes and vocalic endings

Italian exhibits both inherent lexical (or semantic) gender as well as grammatical gender. Lexical gender is semantically motivated because it corresponds to the biological gender of the animate noun; for instance, *zio* 'uncle' is masculine, while *zia* 'aunt' is feminine. In contrast, grammatical gender is said to be arbitrary because it lacks semantic motivation: there is no reason for *tabella* 'table' to be feminine rather than masculine, for example.[1]

It is usually claimed that grammatical gender in Italian is transparent because about 70% of inanimate nouns follow the predictable, transparent pattern of *-o*/*-i* for single/plural masculine and *-a*/*-e* for singular/plural feminine nouns (e.g. Thornton, Iacobini, & Burani, 1997).[2] The nouns that fall outside this pattern are said to be phonologically opaque (i.e. the vocalic ending does not clearly indicate the noun's gender) and are thus considered as exceptions. They represent about 20% of nouns in Thornton et al.'s (1997) count such as *mano* 'hand-FEM-SG' or *problema* 'problem-MSC-SG' while nouns of Greek origin ending in *-ma* are usually masculine (e.g. *idioma* 'language-MSC-SG', *fantasma* 'ghost-MSC-SG').[3] Moreover, nominal endings in *-e* can be either feminine, (e.g. *arte* 'art', *fede* 'faith') or masculine (e.g. *cuore* 'heart', *latte* 'milk', *mare* 'sea'), while those ending in *-ù* or *-i* in the singular are usually classified as irregular and feminine (e.g. *virtù* 'virtue', *tribù* 'tribe', *crisi* 'crisis') (Chini, 1995).

---

1. It is also arbitrary in that it differs across languages. Foundalis (2002) found that 84 nouns (e.g. ideas, objects) across 14 languages rarely shared the same gender, although there were greater similarities within the same language groups such as Romance or Germanic.

2. Modern Italian displays a binary gender system, but it used to have neuter as a third gender (e.g. Loporcaro et al., 2014; Stark, 2008).

3. Thornton et al. (1997) used the VdB (*Vocabolario di Base della lingua italiana*) lexical database compiled by Tullio de Mauro. It contains a little over 7,000 basic words.

This ambiguity is why vocalic endings are best considered as word markers rather than gender markers, as Acquaviva (2009, pp. 50–41) puts it: "the vowel endings do not spell out gender-number features, but class; gender and number may condition this realization, through an abstract fused morpheme added to the morphosyntactic word". A general classification includes 7 noun classes (Chini, 1998 p. 42; cited in Küpisch, Müller, & Cantone, 2002), while a more detailed classification lists 18 (Gudmundson, 2012, p. 207; see Appendix A). Each noun class displays a unique combination of noun ending, gender and number. Although most of the tokens fall within the transparent combinations *-a/-e* for feminine and *-o/-i* for masculine, the ambiguous endings *-e/-i* for both feminine and masculine singular/plural represent the second most important category in terms of number of tokens and frequency. Moreover, as Acquaviva (2009, pp. 50–51) points out, only the first two classes are productive, and the invariable nouns "are an increasingly important open class, enriched by non-assimilated loanwords like *sport* 'sport', clipped forms like *foto* 'photo' (from *fotografia* 'photography') and by formations like *la Topolino* 'the Topolino [car]', *la ministro* 'the [woman] minister' (D'Achille & Thornton, 2003, 2008)".

We argue here and elsewhere (Ayoun & Maranzana, 2020) that the exact number of noun classes is less important than the fact that a few vowels express both gender and number. Italian has a system of stressed – [i], [e], [ɛ], [a], [o], [ɔ], [u] – and unstressed – [i], [e], [a], [o], [u] – vowels (Maiden & Robustelli, 2000). Only five vowels express the combinations of two genders (feminine, masculine) and numbers (singular, plural), and there is no clearly transparent vowel as shown in Table 1 (Sgroi, 2008; Montermini, 2003).

Table 1. Phones and vocalic endings

Phone	Vowel	FS	FP	MS	MP
[a]	*-a*	*casa* 'house'	*pudenda* 'genitals'		
	-à	*città* 'city'		*baccalà* 'cod'	
[e]	*-e*	*luce* 'light'	*penne* 'pens'	*amore* 'love'	*cacciavite* 'screwdrivers'
[ɛ]	*-è*			*caffè* 'coffee'	
[i]	*-i*	*crisi* 'crisis'	*carni* 'meats'	*brindisi* 'toast'	*fiori* 'flowers'
	-ì			*lunedì* 'Monday'	
[o]	*-o*	*mano* 'hand'	*foto* 'photos'	*sogno* 'dream'	*euro* 'euros'
[ɔ]	*-ò*	*dò* 'C'		*falò* 'bonfire'	
[u]	*-ù*	*gioventù* 'youth'		*menù* 'menu'	

Table 1 shows all the vowels can represent both genders and numbers, with *-e* and *-i* appearing to be the most ambiguous. Finally, the vocalic endings of derivational suffixes are not a reliable indicator of gender either; for instance, *-tore* is masculine

singular, while *-trice* is feminine singular. Similarly, the suffixes *-sione* and *-zione* are always feminine singular (with plural forms *-sioni/-zioni*), but *-one* is masculine singular (with *-oni* for the plural). Unless L2 learners are aware of nominal morphology and are able to distinguish between a vocalic ending and a suffix, the numerous suffixes (i.e. several hundreds) are likely to be confusing as well.

1.2 Determiners as gender markers

Contra others (e.g. Chini, 1998), we argue that determiners are not reliable gender markers in Italian, and thus should not be considered as "gender morphemes" in the sense of gender markers (Chini, 1998, p. 43). Table 2 lists definite, indefinite and partitive articles.

Table 2. Definite and indefinite determiners

	Masculine		**Feminine**	
	Singular	**Plural**	**Singular**	**Plural**
definite	*il* *lo, l'*	*i* *gli*	*la* *l'*	*le*
indefinite	*un* *uno*	*dei* *degli*	*una* *un'*	*delle*
partitive	*del* *dello* *dell'*	*dei* *degli*	*della* *dell'*	*delle*

The form of the determiner depends on whether the noun begins with a vowel or a consonant: the definite article *l'* and the partitive article *dell'* precede both singular masculine and feminine nouns beginning with a vowel, and are thus not gender-marked. There is also a feminine singular partitive article *dell'* that precedes any singular noun with a vocalic initial. *Dell'* is thus ambiguous since it is both masculine and feminine singular. The definite articles *la* and *il* precede respectively feminine singular and masculine singular nouns that begin with a consonant, while the definite article *lo* and the partitive article *dello* are used with masculine nouns which begin with some consonants or consonant clusters such as *-z* or *-s* followed by another consonant (e.g. *il libro* 'the book'; *dello stadio* 'of the stadium'). It is worth noting that four of these definite and indefinite determiners (i.e. *l', un, un', dell'*) lack the vocalic endings that are said to reliably indicate gender, and assumed to match the nominal vocalic endings. Moreover, the only difference between the indefinite articles *un*-MSC and *un'*-FEM is the apostrophe, a very small diacritic arguably easily overlooked in the written mode and without any counterpart in speech. Finally, there are fewer feminine determiners ($n = 9$) than masculine determiners ($n = 14$),

possibly making the latter more salient and thus introducing a potential bias toward the masculine. We contend that none of these determiners provide clear and transparent indications of gender marking for L2 learners.

1.3 Compounds, grammatical homonyms and other idiosyncrasies

The gender of compound nouns may be the same as either the first or the second noun, and nominal compounds do not always follow the transparent vocalic endings *-o/-a* since some compounds ending in *-a* are masculine as shown in (1):[4]

(1) *terra*-FEM 'land', *entroterra*-MSC 'inland'
fine-FEM 'end', *finesettimana*-MSC 'weekend'

Another peculiarity specific to compounds is that when a compound contains a feminine noun ending in *-e* such as *case* 'houses', the plural morphology is feminine, but the compound agreement with the determiner and adjectives is in the masculine as shown in (2) (adapted from Gracanin-Yuksek, 2006, pp. 114–115):

(2) *testa*-SG 'head'–> *teste*-PL 'heads'
poggiatesta-MSC 'headrest'
dei-MSC-PL /**delle*- FEM-PL *poggiateste/*poggiatesti*

The plural formation of compounds in general tends to be ambiguous as it depends on the gender of their components. For instance, N+N compounds use the plural of the first N when both nouns are feminine or masculine, otherwise, they take on the plural of the second noun, with a few exceptions such as *banconota*-FEM-SG → *banconote*-FEM-PL. After analyzing several corpora, Micheli (2016) concludes that "the formation of the plural in compound nouns constitutes an element of instability within the Italian morphological system: it does not obey systematic rules, but varies according to the typology of the compound, and thus to the syntactic category of the constituents, and within the same typology" (ibid, p. 254; translation ours).

Other idiosyncrasies include grammatical homonyms whose meaning fluctuates with gender as in (3):

(3) *fronte* 'war front-MSC', 'forehead-FEM'
morale 'humor/spirit-MSC', 'ethic-FEM'
retroscena 'behind the scenes-MSC', 'backstage-FEM'
capitale-MSC 'capital, wealth-MSC', 'capital city-FEM'

4. Although there are several types of compounds in Italian, only three types (N+N, A+A, V+N) are productive according to Scalise (1994). See also Masine & Scalise (2012).

The gender of a few nouns fluctuates with number as exemplified in (4):

(4) *centinaio*-MSC-SG –> *centinaia*-FEM-PL 'hundred/hundrEds.'
uovo-MSC-SG 'egg'–> *uova*-FEM-PL 'eggs'
lenzuolo-MSC-SG 'sheet' –> *lenzuola*-FEM-PL 'sheets'
eco-FEM-SG 'echo' –> *echi*-MSC-PL 'echoes'
riso-MSC-SG 'laugh' –> *risa*-FEM-PL 'laughter'

When nouns are truncated, they retain the same gender regardless of the new vocalic ending and are invariant in the plural as in (5):

(5) *fotografia* –> *foto*-FEM-SG-PL 'photo'
cinematografo –> *cinema*-MSC-SG-PL 'cinema'
motocicletta –> *moto*-FEM-SG-PL 'motorcycle'
bicicletta –> *bici*-FEM-SG-PL 'bike'
televisione –> *tele*-FEM-SG-PL 'television'

Truncation thus creates opaque vocalic endings: *-o* and *-i* for feminine singular (instead of *-a*) and vice-versa, *-a* for masculine singular (instead of *-o*)

Quite a few masculine nouns have two plural forms: one in *-i* and one in *-a* as exemplified in (6):

(6) *gridо* –> *gridi, grida*-MSC, FEM 'screams'
labbro –> *labbri*-MSC 'edges', *labbra*-FEM 'lips'
gesto –> *gesti*-MSC-PL 'gestures', *gesta*-FEM-PL 'deeds.'
dito-MSC-SG 'finger'–> *dita*-FEM-PL 'fingers'

Finally, borrowings are generally masculine by default (e.g. Dressler & Thornton, 1996; Riente, 2003) (e.g. *sport, bar, gas, film*), but some are feminine (e.g. *holding, gang*).

To sum up, Italian nouns bear vocalic endings that are morphological suffixes expressing both gender and number as portmanteau morphemes (e.g. De Martino et al. 2011). In other words, the gender and number features are conflated, and there is no discrete morpheme for number (see e.g. Acquaviva, 2009), contrary to other Romance languages such as French whose gender information is opaque (e.g. Ayoun, 2010, 2018), but which indicates number separately from gender with a final *-s* or *-x* for most nouns (e.g. *la rose*-FEM-SG, *les roses*-FEM-PL 'rose'*; le chou*-MSC-SG, *les choux*-MSC-PL 'cabbage').

Let us now see how Minimalism accounts for grammatical gender and the way it is expressed in Italian.

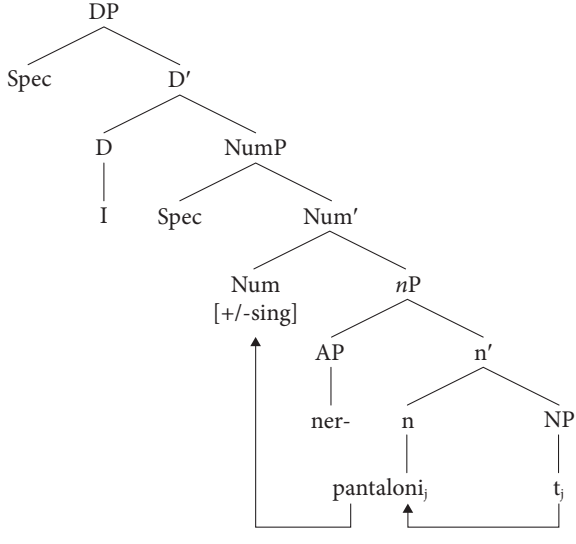
2. Grammatical gender and minimalism: Theory and L2 acquisition

In languages with a nominal classification system, the lexical entry of nouns includes a feature indicating its class. This feature [±fem] is required in the construction of N roots sometimes referred to as the narrow lexicon (Marantz, 1997). The phonological lexicon – composed of lists of phonological forms partially specified for grammatical features (Halle & Marantz, 1993) – provides phonological content to the lexical roots. When nominal roots enter syntactic derivations, they merge with other categories to create an extended nominal projection composed of at least Number and Determiner (e.g. Ritter, 1991, 1993). So grammatical gender is a lexical property of nouns, while number and person features are a property of the determiner (e.g. Carstens, 2000; Chomsky, 1995, 2001; Ritter, 1991).

Determiners and adjectives are assigned uninterpretable gender (i.e. [*u*gender]) features on their root entries, and the [±fem] feature checks and erases the [*u*gender] features on determiners and adjectives from the derivation to LF to satisfy the principle of Full Interpretation. Determiners and adjectives are then attributed the appropriate [±fem] feature.

Determiner Phrases (DPs) thus include the functional category Number (Num) between Det and NP as well as number and gender features (e.g. Picallo, 1991; Ritter, 1993; Santoro, 2012), and adjectives are assumed to be generated from *nP* or "light" noun shell outside the NP (Carstens, 2000). This account leads to the syntactic representation in (7) as proposed and developed in Santoro (2012, p. 171):

(7) A DP's syntactic representation



The noun *pantaloni* 'pants' is base-generated in NP which enters the numeration with the interpretable gender features that must check the uninterpretable φ features of agreeing determiners, adjectives and pronouns to prevent the derivation from crashing at the PF level.

The reader may recall that there are only two types of formal features (Chomsky, 2001): (a) lexically valued and interpretable features, that is, features with semantic content; (b) lexically unvalued and uninterpretable features, that is, features without semantic content. However, grammatical gender appears to be both uninterpretable (due to its arbitrariness) and valued as a feature of the lexical properties of nouns (e.g. Carstens, 2010; Pesetsky & Torrego, 2007), while lexical gender (i.e. semantically motivated since it reflects the biological gender of the noun) is both interpretable and valued. This goes against the basic claim that the only difference between interpretable features and uninterpretable features is that the latter are not inherently valued.

If grammatical gender features are both uninterpretable and valued, they stand out as an anomaly in the current minimalist system (Carstens, 2010). They also contradict the claim made by impairment hypotheses such as the Failed Functional Features Hypothesis (i.e. features that are not present in the L1 are not acquirable in the L2 and will result in incomplete acquisition and permanent deficits; e.g. Hawkins & Chan, 1997; Hawkins & Liszka, 2003) and the Representational Deficit hypothesis (i.e. interpretable features may be acquired, but not uninterpretable features (e.g. Hawkins, 2003; Hawkins & Franceschina, 2004). Similarly, according to the Interpretability hypothesis (Hawkins & Hattori, 2006; Tsimpli & Dimitrakopoulou, 2007), uninterpretable features that are not instantiated in the L1 are vulnerable in the L2, while those bearing interpretable features are not even if they are not activated in the L1.

We adopt a different view than these impairment hypotheses by arguing that features and functional categories are in principle acquirable by adult learners because L2 acquisition is constrained and guided by Universal Grammar (for a review see e.g. Ayoun & Rothman, 2013; Rothman, 2008) as proposed by the Full Transfer/Full Access hypothesis (FA/FT) (Schwartz & Sprouse, 1994, 1996), the Full Access hypothesis (Epstein, Flynn, & Martohardjono, 1996) or the Full Functional hypothesis (Gess & Herschensohn, 2001). It does not follow that L2 acquisition will be effortless, far from it. L2 learners do experience difficulties, but they are hypothesized to stem from issues in mapping form to meaning, lexical retrieval or performance – not from representational deficits – as formalized by the Missing Inflection hypothesis (Haznedar & Schwartz, 1997), the Morphological Underspecification hypothesis (McCarthy, 2007), the Missing Surface Inflection Hypothesis (Prévost & White, 2000), and the Prosodic Transfer hypothesis (Goad, White, & Steele, 2003) as well as the Feature Reassembly Approach (FRA)(Lardiere, 2008, 2009).

The concept of assembled features originates from Chomsky (1988: 13; 2001: 4) who proposes that each language "makes a one-time selection of a subset [F] of {F} and a one-time assembly of elements of [F] as its lexicon LEX". Following the FRA, L2 learners need to reassemble features according to L2 requirements, initially relying on their L1 following the FA/FT hypothesis (Lee & Lardiere, 2019). The FRA is thus "a comparative linguistic feature-based approach" (Hwang & Lardiere, 2013, p. 58) in which the successful acquisition of L2 morphosyntactic knowledge depends on learners' ability to reassemble the feature values in functional categories and lexical items. Even if the learners' L1 and L2 exhibit the same features, they may be assembled or redistributed differently leading to acquisition difficulties.

In the case of Anglophone learners of L2 Italian, the number feature is expressed differently in the L2 (i.e. vocalic endings) than in the L1 (*-s, -es* and some irregular plurals), while the grammatical gender feature is absent altogether. Anglophone learners are familiar with a morphologically poor pronominal gender language with masculine, feminine and neuter (e.g. he/she/it), while L2 Italian is a morphologically rich language with both lexical and grammatical gender expressed with vocalic endings. According to the FRA, L2 learners first need detect how the features of gender and number are expressed, and then reassemble them. The features have to be detectable to be acquirable (Lardiere, 2009). The fact that they are conflated will render their detectability more difficult especially if Lardiere (2009) is correct in assuming that a formal contrast such as *child/children* is a prerequisite to detection. The cross-linguistic differences added to the ambiguities and idiosyncrasies described above are bound to create learning and processing difficulties for L2 learners as found in previous studies.

3. Literature review

Early non-generative studies (e.g. Berretta, 1990; Berretta & Crotta, 1991; Chini, 1995; Valentini, 1990) indicate that L2 learners tend to overuse the more regular and transparent forms and systematically assume that nouns ending in *-a* are feminine while nouns ending in *-o* are masculine; they also seem to experience difficulties with nouns ending in the ambiguous *-e*. More recent studies report similar findings.

In Oliphant (1998), American college students (n = 64) performed a written gender assignment task with familiar nouns. The first test included 24 nouns exemplifying endings in *-o, -a, -e, -i, -u*, consonant + 14 nouns with a derivational suffix (*-cida, -ista, -one, -sione/-zione, -tore, -trice*). In test 2, 18 nouns were presented with a definite article that was gender-marked (*la, il*) or not (*l'*), with nouns ending in ambiguous *-e*, or without any gender cues on either the noun or determiner. In

test 3, the stimuli exemplified grammatical and lexical gender with (in)animate nouns, with gender cues for only some of the stimuli, and participants were asked to select a modifying adjective. Findings show that the participants' performance correlates with the predictability of the vocalic endings: they performed best with *-o/-a* and less well for *-e* and various suffixes (from 74.2% for *-trice* to 1.6% for *-cida*). Although they were told that all the stimuli were singular, participants assigned feminine *-i* with the masculine (72%). They also erroneously associated *-u* with masculine (82%). Several caveats weaken these tentative findings: there were too few tokens of each type of stimuli for all the tests; participants were asked to guess when they were unsure; in the first test, although the suffixes could be both masculine and feminine, they were only classified as masculine.

Gudmundson (2012) investigated the acquisition of grammatical gender and number agreement in determiners and adjectives both within and outside the NP by Swedish college students (n = 25) who were learning Italian in an instructed setting. Student-instructor dialogues that were recorded and transcribed are part of the corpus InterIta (e.g. Bardel, 2004); 30 recordings by 15 different learners were chosen to show longitudinal development (between 4 and 29 months). The cue availability, reliability and validity of vocalic endings in the LIP corpus from which 77 100 nouns were extracted and analyzed according to their gender, number and ending. It was found that *-i* scored the highest as a masculine plural, but also as a feminine plural, while it turned out that *-e* had high values for masculine singular, but also feminine singular and plural. The findings from the oral data show that participants were sensitive to the distributional bias of the input: they obtained higher accuracy rates for predictable, transparent patterns (i.e. singular masculine) than with less predictable and more opaque patterns (i.e. feminine plural). The participants' performance on nouns with a consonantal ending was also very good because although these nouns are less frequent than nouns with a vocalic ending, they are reliably masculine singular.

Santoro (2012) set out to examine the morphological variability in the production of determiners, adjectives as well as direct object pronouns by instructed adult English learners of L2 Italian in an attempt to establish whether morphological inaccuracies may be due to computational or representational difficulties in a generative framework. Unfortunately, participants were classified as high-beginners (n = 18) or high-intermediate (n = 17) based on the number of years of instruction they had received, not on the results of an independent proficiency test. Italian native speakers (n = 12) were used as a control group. Participants completed a written agreement recognition task (cloze test with definite articles, adjectives, and direct object pronouns given in their masculine singular form had to be modified to agree as necessary) and an oral picture identification task. Findings show good accuracy percentages overall (from 75.2% to 86.5% on the oral task, and from

83.8% to 90.2% on the written task) with a better performance on determiners and adjectives than on pronouns.

Bianchi (2013) compared German-Italian bilinguals (n = 20) with advanced German learners (n = 15) of L2 Italian who completed a background questionnaire, a cloze test in German and Italian to establish language dominance for the bilingual speakers and two elicitation tasks of gender assignment and agreement: a timed acceptability judgment task (repeating or correcting aloud 40 stimuli presented aurally and on a computer screen) and a production task (describing two actions perfomed by a character). Stimuli included feminine and masculine nouns ending in *-a, -e, -o,* -consonant as well as feminine nouns ending in-*ie* or *-trice* and masculine nouns ending in *-ale* or *-one*. Findings show that bilinguals speakers with Italian as their stronger language significantly outperformed bilingual speakers with Italian as their weaker language and L2 Italian learners whose performance did not differ significantly. Participants were more accurate in gender agreement than in gender assignment with most errors occurring with masculine words ending in *-a* or -consonant, and feminine words ending in *-ie/-trice, -o* or -consonant. The masculine was preferred over the feminine, and as expected, all participants performed better with the transparent endings *-a/-e* and *-o/-i*.

To sum up, a variety of elicitation tasks – written gender assignment, dialogues, written agreement recognition, oral picture identification, timed acceptability judgment, production – show that the higher the predictability of the vocalic endings due to their transparence, the better performance and accuracy of the participants who appear to be sensitive to the distributional bias in the input. However, previous studies did not systematically test L2 learners' morphological knowledge of all vocalic endings by taking suffixes into account and clearly disambiguating between singular and plural endings. The present study addresses this caveat.

Swahili is another language that conflates gender and number on a single prefix similarly to Italian. In Spinner (2013), Anglophone learners (n = 38) of L2 Swahili were administered an oral picture description task designed to elicit verbal and adjective agreement and a written gender assignment task. Results show that participants often used singular instead of plural noun markers while being generally accurate on gender, but they were able to choose appropriate number and gender categories for nouns with various prefixes. Specifically, participants used singular marking on noun prefixes in plural contexts. However, they almost never used prefixes with incorrect gender. The author attributes the participants' errors to a failure to detect that two features are bundled into a single morpheme.

In an exploratory study using grammatical priming, a word familiarity tasks and two word repetition tasks, Spinner, Foote, and Upor (2017) found that the Swahili NSs were sensitive to incongruent gender and number effects on verbs and nouns, while Anglophone L2 learners were only sensitive to number effects. In a

gender categorization task, the L2 participants' accuracy ranged from 64% to 88% depending on the noun class. Spinner et al. did not investigate whether participants could accurately assign or produce gender and number on nouns, and they do recommend that future studies do assess gender assignment by L2 learners. They also acknowledge the impossibility to conclude whether "the different results in gender and number marking are due to L1 effects or to differences in the way gender and number are processed" (ibid, p. 469).

4. Methodology and research questions

4.1 Research questions

Our main research question is from a Feature Reassembly approach (FRA) which requires L2 learners to select vocalic endings as new features expressing both gender and number on lexical items, thereby testing their morphological competence. Can instructed learners of L2 Italian assign the appropriate gender to bare nouns (i.e. without determiners or adjectives) presented as isolated stimuli in a written gender assignment task? In other words, do they acquire a portmanteau morpheme to successfully assign the appropriate masculine or feminine grammatical gender to individual nouns?

The following additional research questions ask: (a) will participants perform better with transparent endings than with ambiguous endings? (b) will there be a number and/or gender effect?; (c) will participants be able to rely on transparent suffixes? (d) will there be a correlation between meaning and accuracy? (e) will there be a proficiency effect? (f) how native-like will the L2 learners' performance be?

The FRA predicts that learners will have difficulties detecting conflated features on a single morpheme, affecting their performance, particularly with ambiguous stimuli as opposed to transparent stimuli.

4.2 Participants

Participants (n = 87) were recruited among college students enrolled in first or second year Italian language classes at a major North-American university. They were all non-native speakers of Italian. The control group was composed of native speakers of Italian (n = 109) recruited via professional networks in Italy.

All but one L2 participants were undergraduate college students, most of them were female, they had been studying Italian between 6 to 24 months and most had not spent time in Italy; when they did, it was for short stays unlikely to have a significant impact on their proficiency.

Table 3. Participants' background information

	L2 groups	NS group
n	87	108
gender	female (n-60), male (*n* = 27)	female (*n* = 71), male (*n* = 37)
age average	21	48.8
age range	16–49	24–88
college status	freshman (*n* = 27), sophomore (*n* = 34) ; junior (*n* = 15); senior (*n* = 9); graduate (*n* = 1)	n/a
L2 study average	13 months	n/a
L2 study range	6–24 months	n/a
L2 stay	yes (*n* = 37), no (*n* = 50)	n/a
L2 stay length	1 week (*n* = 10); 2 weeks (*n* = 7); 3, 5, 6, 8, 10, 16, 24 weeks (*n* = 1 each); 4 weeks (*n* = 4); 12 weeks (*n* = 3)	n/a

4.3 Tasks and stimuli

After filling out a background information questionnaire, the L2 participants took a pre-test and completed a gender assignment task, which were both written and computerized. The pre-test was administered as an independent measure of proficiency as we cannot rely on the class(es) in which the participants may be enrolled. It consisted of a 40-item grammaticality judgment task (GJT) targeting various morpho-syntactic properties (i.e. tense, agreement, clitics) with complete sentences evenly split between grammatical and ungrammatical stimuli. Participants indicated whether the sentences were grammatical, ungrammatical, or if they did not know. They were instructed to rely on their first reaction.

Participants were then asked to complete a gender assignment task consisting of 240 Italian bare nominals selected from the written corpus of current Italian magazine and newspapers articles available online (Ayoun & Maranzana, 2020)[5] according to the following criteria: (a) nouns exhibiting the transparent vocalic endings for the masculine singular/plural with *-o*/*-i* (e.g., *albergo/alberghi* 'hotel-MSC') and for the feminine singular/plural with *-a*/*-e* (e.g., *punta/punte* 'point'); (b) nouns presenting the ambiguous vocalic ending singular/plural *-e*/*-i* (e.g., *luce/luci* 'light-FEM'; *stile* 'style-MSC' or *liti* 'litigations-FEM', *segnali* 'signals-MSC'); (c) masculine nouns with the ambiguous vocalic ending singular/plural in *a*/*-i*

5. The written corpus is composed of 40 news articles from *L'Espresso, Il Corriere della Sera, Focus, La Repubblica*, and *Oggi* accessed online. The articles covered various topics such as economy, politics, news, science, and technology, thus typical of the authentic input to which instructed L2 learners of Italian are exposed. The 40 articles generated 5,038 DPs (out of 23,047 words) which were analyzed to be categorized as either ambiguous or transparent.

(e.g., *problema* 'problem-MSC'); (d) feminine and masculine nouns exhibiting the ambiguous vocalic endings singular/plural *-e*/*-i* (e.g., *nome* 'name-FEM' or *salute* 'health-FEM'; *mesi* 'months-MSC' or *stragi* 'massacres-FEM'); (e) nouns that featured transparent suffixes such as *-ione* or *-enze*, which are feminine and *-ume* or *-ivi* which are masculine (e.g., *visione* 'vision-FEM' or *credenze* 'beliefs-FEM'; *agrume* 'citrus-MSC' or *obiettivi* 'objectives-MSC').

The stimuli were presented in random order and were composed of transparent and ambiguous nouns as displayed in Table 4.

Table 4. Composition of the stimuli

	Ambiguous nouns without a suffix	
feminine singular	10	*legge* 'law'
feminine plural	10	*mani* 'hands'
masculine singular	10	*pane* 'bread'
masculine plural	10	*ponti* 'bridge'
	Transparent nouns without a suffix	
feminine singular	10	*ora* 'hour'
masculine singular	10	*libro* 'book'
	Transparent nouns with a suffix	
feminine singular	50	*-ione*, *-ice*, *-udine* (e.g., *confusione* 'confusion', *cucitrice* 'sewing machine', *incudine* 'anvil')
feminine plural	40	*-ioni*, *-ezze*, *-ive*, *-ite*, *-enze*, *-ette*, *-ucce*, (e.g., *azioni* 'actions', *certezze* 'certainties', *iniziative* 'initiatives', *ferite* 'wounds', *interferenze* 'interferences', *cassette* 'boxes')
masculine singular	50	*-ore*, *-ume*, *-ere*, *-ame*, (e.g., *dolore* 'pain', *pattume* 'trash', *bicchiere* 'glass', *esame* 'exam')
masculine plural	40	*-agli*, *-ivi*, *-oli*, *-itti*, *-etti*, *-atti*, *-atti*, *-aggi*, *-ori*, (e.g., *dettagli* 'details', *arrivi* 'arrivals', *veicoli* 'vehicles', *conflitti* 'conflicts', *progetti* 'projects, *ricatti* 'extortions', *paesaggi* 'landscapes', *integratori* 'supplements')

We used more nominal endings with a suffix than without to reflect the input bias (Italian has hundreds of suffixes). Participants used a pull-down menu to: (a) assign a gender to each of the bare nominals as either feminine, masculine, both or if they did not know;[6] (b) indicate whether they knew the meaning of the noun or not by checking the appropriate box. The task was not timed.

6. The 'I don't know' option was included to reduce the possibility that participants would just guess when they were unsure, thereby increasing the reliability of the results and providing an indication of their confidence level.

5. Results

5.1 Pre-test

The results of the pre-test placed the participants (n = 87) into three groups of proficiency labelled Low (n = 30), Mid (n = 28) and High (n = 29) based on their accuracy scores for correctly accepted and correctly rejected stimuli ('don't know' responses were rare and not taken into account). The mean averages and ranges were as follow: Low (24.3% mean, 22.1%–26.5% range), Mid (31.9% mean, 30.3%–33.4% range), High (36.7% mean, 34.4%–39% range). There was a significant difference between groups (sum of squares = 3499.182, df = 2, mean square = 1759.591, f = 91.044, p > 0.000).

5.2 Gender assignment task

A mixed models analysis was run with all the raw data. The results are presented in individual tables because there are so many different variables. The results of the gender assignment task are presented by gender, number and transparency/ambiguity. Table 5 provides an overview of accuracy means by groups.

Table 5. Overall averages for L2 and NS groups

Stimuli		L2 groups (n = 87)	NS group (n = 108)
feminine singular	transparent	66.23%	95.7%
	ambiguous	42.9%	90.0%
feminine plural	transparent	51.43%	99.1%
	ambiguous	38.43%	91.1%
masculine singular	transparent	38.36%	98.6%
	ambiguous	31.0%	98.1%
masculine plural	transparent	72.86%	99.3%
	ambiguous	74.5%	99.7%

The L2 groups' performance varied from a low of 31.0% for ambiguous masculine singular stimuli to a high of 74.5% for ambiguous masculine plural stimuli, indicating that number has a stronger effect than transparency. L2 learners performed relatively well on only three types of stimuli: transparent feminine singular (66.23%), transparent masculine plural stimuli (72.86%) and ambiguous masculine plural stimuli (74.5%). Overall, they did slightly better on masculine stimuli (53.74%) than on feminine stimuli (49.75%). The NS group performed as expected, at 90%

or above, but their performance varies as well with a low of 90.0% on ambiguous feminine singular stimuli and 91.1% on ambiguous feminine plural stimuli, but reaches ceilings on everything else.

The next tables present more detailed results, starting with feminine singular and feminine plural stimuli in Tables 6 and 7.

Table 6. Results for feminine singular stimuli by groups and ambiguity

Feminine singular		Participants' answers by accuracy percentages			
		Feminine	Masculine	Both	DK
transparent stimuli	low group	**66.9%**	12.0%	1.6%	19.4%
	mid group	**63.6%**	14.9%	2.4%	19.1%
	high group	**68.2%**	15.6%	2.4%	13.9%
	NS group	**95.7%**	4.3%	0.0%	0.0%
Pearson $\chi^2 = 1961.14$, $df = 9$, $p < 0.001$					
ambiguous stimuli	low group	**43.8%**	28.1%	9.8%	18.3%
	mid group	**43.7%**	25.7%	12.3%	18.3%
	high group	**41.2%**	33.9%	8.8%	16.1%
	NS group	**90.0%**	9.3%	0.3%	0.0%
Pearson $\chi^2 = 724.705$, $df = 9$, $p < 0.001$					

Table 7. Results for feminine plural stimuli by groups and ambiguity

Feminine plural		Participants' answers by accuracy percentages			
		Feminine	Masculine	Both	DK
transparent stimuli	low group	**54.1%**	19.2%	4.5%	22.2%
	mid group	**50.0%**	20.4%	4.8%	24.8%
	high group	**50.2%**	24.7%	4.5%	20.5%
	NS group	**99.1%**	0.7%	0.1%	0.1%
Pearson $\chi^2 = 2595.47$, $df = 9$, $p < 0.001$					
ambiguous stimuli	low group	**38.4%**	38.9%	0.0%	22.7%
	mid group	**39.6%**	35.2%	2.6%	22.5%
	high group	**37.3%**	44.6%	1.5%	16.7%
	NS group	**91.1%**	7.6%	1.2%	0.1%
Pearson $\chi^2 = 512.47$, $df = 9$, $p < 0.001$					

There is a strong effect for ambiguity and number for the L2 learner groups who perform much better on transparent and singular stimuli (66.23% average) than on ambiguous and plural stimuli (38.43%). The performance of the NS groups is clearly significantly better than the performance of the L2 learner groups, although it is better for transparent stimuli than ambiguous stimuli for both the singular (95.7% vs 90.0%) and the plural (99.1% vs. 91.1%).

Tables 8 and 9 display the results for the masculine stimuli and again we have a strong effect for ambiguity and number, but it is reversed for the latter in that all three learner groups performed much better on masculine plural than on masculine singular stimuli.

Table 8. Results for masculine singular stimuli by groups and ambiguity

Masculine singular		Participants' answers by accuracy percentages			
		Feminine	Masculine	Both	DK
transparent stimuli	low group	40.3%	**32.0%**	3.0%	24.7%
	mid group	30.4%	**41.2%**	2.6%	25.8%
	high group	34.3%	**41.9%**	3.2%	20.5%
	NS group	0.9%	**98.6%**	0.3%	0.2%
Pearson $\chi^2 = 5005.82$, $df = 9$, $p < 0.001$					
ambiguous stimuli	low group	53.0%	**24.7%**	9.1%	13.2%
	mid group	42.2%	**33.1%**	10.0%	14.7%
	high group	45.8%	**35.2%**	9.7%	9.3%
	NS group	1.2%	**98.1%**	0.6%	0.0%
Pearson $\chi^2 = 888.11$, $df = 9$, $p < 0.001$					

Table 9. Results for masculine plural stimuli by groups and ambiguity

Masculine plural		Participants' answers accuracy percentages			
		Feminine	Masculine	Both	DK
transparent stimuli	low group	5.8%	**76.7%**	1.0%	16.5%
	mid group	8.5%	**68.0%**	2.1%	21.5%
	high group	13.1%	**73.9%**	1.6%	11.4%
	NS group	0.5%	**99.3%**	0.1%	0.1%
Pearson $\chi^2 = 1387.81$, $df = 9$, $p < 0.001$					
ambiguous stimuli	low group	4.3%	**80.6%**	1.6%	13.4%
	mid group	6.6%	**69.5%**	6.1%	17.8%
	high group	17.9%	**74.5%**	1.0%	6.6%
	NS group	0.3%	**99.7%**	0.0%	0.0%
Pearson $\chi^2 = 336.13$, $df = 9$, $p < 0.001$					

The L2 learner groups perform much better on the plural masculine than on the singular masculine stimuli with a smaller effect for ambiguity (72.86% for transparent, 74.86% for ambiguous), while NSs perform at ceiling levels for both singular and plural masculine stimuli (98.1% to 99. 7%). These effects are confirmed by highly significant differences across the board.

Since the NSs' performance was so different from the L2 learners' performance, we also ran Pearson χ^2 tests with only the L2 learner groups to see if there were significant differences between them. They appear in Table 10.

Table 10. Pearson χ^2 tests with L2 groups

		Pearson χ^2	df	*p*
Feminine singular	transparent stimuli	30.367	6	0.001
	ambiguous stimuli	7.207	6	0.302
Feminine plural	transparent stimuli	15.116	6	0.019
	ambiguous stimuli	10.081	6	0.121
Masculine singular	transparent stimuli	55.193	6	0.001
	ambiguous stimuli	10.543	6	0.104
Masculine plural	transparent stimuli	71.357	6	0.001
	ambiguous stimuli	43.731	6	0.001

The results are still highly significant for transparent feminine singular ($p < 0.000$) and plural ($p = 0.019$), transparent masculine singular ($p < 0.000$) as well as transparent masculine plural ($p < 0.000$) and singular ($p < 0.000$).

The next four tables present the results for the stimuli with and without a suffix.

Table 11. Results for feminine singular stimuli by groups and suffix

Feminine singular		Participants' answers accuracy percentages			
		Feminine	Masculine	Both	DK
without a suffix	low group	**59.4%**	17.2%	6.2%	17.2%
	mid group	**61.2%**	16.2%	8.0%	14.6%
	high group	**62.2%**	20.5%	5.8%	11.5%
	NS group	**93.6%**	5.9%	0.2%	0.2%
Pearson $\chi^2 = 811.82$, $df = 9$, $p < 0.001$					
with a suffix	low group	**64.5%**	13.7%	1.7%	20.1%
	mid group	**59.7%**	17.1%	2.4%	20.9%
	high group	**63.9%**	18.2%	2.5%	15.5%
	NS group	**95.1%**	4.8%	0.0%	0.0%
Pearson $\chi^2 = 1822.55$, $df = 9$, $p < 0.001$					

Except for the low group that performs slightly better on feminine singular stimuli with a suffix (64.5% vs 59.4%), there is no clear difference in the L2 learners' performance on stimuli with or without a suffix; the means hover around the low 60%, contrary to the NSs group whose average is slightly higher on stimuli with a suffix as opposed to without a suffix (95.1% vs 93.6%).

Table 12. Results for feminine plural stimuli by groups and suffix

Feminine plural		Participants' answers accuracy percentages			
		Feminine	Masculine	Both	DK
without a suffix	low group	**38.4%**	38.9%	0.0%	22.7%
	mid group	**39.6%**	35.2%	2.6%	22.5%
	high group	**37.3%**	44.6%	1.5%	16.7%
	NS group	**91.1%**	7.6%	1.2%	0.1%
Pearson $\chi^2 = 512.47$, $df = 9$, $p < 0.001$					
with a suffix	low group	**54.1%**	19.2%	4.5%	22.2%
	mid group	**50.0%**	20.4%	4.8%	24.8%
	high group	**50.2%**	24.7%	4.5%	20.5%
	NS group	**99.1%**	0.7%	0.1%	0.1%
Pearson $\chi^2 = 2595.45$, $df = 9$, $p < 0.001$					

The L2 learners' performance is markedly worse on feminine plural stimuli without a suffix than with a suffix (average of 38.43% vs 51.43%). Interestingly, the NS group's performance is impacted by the absence vs presence of the suffix (91.1% vs 99.1%) as well.

Table 13. Results for masculine singular stimuli by groups and suffix

Masculine singular		Participants' answers accuracy percentages			
		Feminine	Masculine	Both	DK
without a suffix	low group	26.1%	**55.6%**	5.3%	13.0%
	mid group	20.5%	**63.4%**	5.0%	11.0%
	high group	23.2%	**64.4%**	5.3%	7.1%
	NS group	0.6%	**99.1%**	0.3%	0.0%
Pearson $\chi^2 = 5090.76$, $df = 9$, $p > 0.001$					
with a suffix	low group	48.5%	**21.0%**	3.2%	27.3%
	mid group	36.7%	**30.7%**	3.0%	29.7%
	high group	41.1%	**31.4%**	3.6%	23.9%
	NS group	1.0%	**98.3%**	0.4%	0.2%
Pearson $\chi^2 = 5090.76$, $df = 9$, $p < 0.001$					

The mid and high groups outperform the low group particularly for stimuli with a suffix, but the percentages are quite low (average of 27.5%), so it appears that the presence of a suffix is not helpful for the masculine singular stimuli. The NS performance is at ceiling.

Table 14. Results for masculine plural stimuli by groups and suffix

Masculine plural		Participants' answers accuracy percentages			
		Feminine	Masculine	Both	DK
without a suffix	low group	4.3%	**80.6%**	1.6%	13.4%
	mid group	6.6%	**69.5%**	6.1%	17.8%
	high group	17.9%	**74.5%**	1.0%	6.6%
	NS group	0.3%	**99.7%**	0.0%	0.0%
Pearson $\chi^2 = 336.13$, $df = 9$, $p < 0.001$					
with a suffix	low group	5.8%	**76.7%**	1.0%	16.5%
	mid group	8.5%	**68.0%**	2.1%	21.5%
	high group	13.1%	**73.9%**	1.6%	11.4%
	NS group	0.5%	**99.3%**	0.1%	0.1%
Pearson $\chi^2 = 1387.81$, $df = 9$, $p < 0.001$					

For the masculine plural stimuli, the low group obtained their highest averages and outperformed the mid and high groups (80.6% for stimuli without a suffix, 76.7% for stimuli without). NSs are still at ceiling. All the differences are statistically significant.

Participants were also asked to indicate whether they knew the meaning of the stimuli or not. They overwhelmingly indicated that they did since the percentages for 'unknown meaning' vary between 0.1% and 0.5% regardless of the gender, number, ambiguity, suffix; moreover, the χ^2-square tests are only significant for the masculine (Pearson $\chi^2 = 8.854$, $p = 0.002$).

We also ran correlations between the participants' accuracy, whether they knew the gender of the stimuli or not ('don't know') and whether they indicated knowing the meaning of the stimuli ('meaning). These results are displayed in Table 15.

Table 15. Correlations between accuracy, 'don't know' and meaning

		Pearson correlation	Sig. (2-tailed)
Low group	accuracy / don't know	−.913	$p < 0.001$
	acccuracy / meaning	0.207	$p = 0.301$
	don't know / known meaning	−0.236	$p = 0.235$
Mid group	accuracy / don't know	−.863	$p < 0.001$
	accuracy / meaning	0.007	$p = 0.970$
	don't know / known meaning	−0.232	$p = 0.210$
High group	accuracy / don't know	−.750	$p < 0.001$
	accuracy / meaning	0.203	$p = 0.309$
	don't know / known meaning	−0.008	$p = 0.968$
NS group	accuracy / don't know	−.858	$p = 0.003$
	accuracy / meaning	−.803	$p = 0.009$
	don't know / known meaning	−.913	$p = 0.001$

There are strong negative correlations between accuracy and 'don't know' for all the groups, and they are all significant. There are weak positive correlations between accuracy and meaning, none are significant for the L2 groups. There are weak negative correlations between 'don't know' and meaning for the L2 groups and none are significant. As can be expected, there are strong, significant negative correlations for all categories for the NS group.

6. Discussion and conclusion

Our main research question asked whether instructed learners of L2 Italian could assign the appropriate gender to bare nominals as isolated stimuli presuming they would have to rely on the vocalic endings, particularly if they did not know the meaning of the nouns. The overall accuracy means showed that the L2 groups did not perform well except for masculine plural stimuli whether the vocalic endings were transparent (72.86%) or ambiguous (72.86%), and marginally so on transparent feminine singular stimuli (66.23%).

Regarding our more specific research questions: First, whether participants would perform better with transparent endings than with ambiguous ending, the answer is yes, except for the masculine singular where it only made a small difference with equally poor averages (38.36% and 31.0% for transparent and ambiguous stimuli, respectively) and the masculine plural with much better performances (72.86% and 74.5% for transparent and ambiguous stimuli, respectively)

Second, we found both number and gender effects with an apparent interaction of the two: L2 groups performed better on the masculine, but only masculine plural, whereas they performed worse on feminine plural be it transparent (51.43%) or ambiguous (38.43%) than feminine singular, but with better accuracy averages on transparent stimuli (66.23%) than ambiguous stimuli (42.9%).

Third, the presence of a suffix only helped L2 learners for masculine plural stimuli; it did not appear to have a facilitative effect for the other stimuli. This may be because our learners were not yet sufficiently advanced to have acquired the distinction between vocalic endings and suffixes.

Fourth, we did not find a correlation between meaning and accuracy, so the semantic knowledge of the stimuli did not translate into a morphological knowledge of the vocalic endings or of the grammatical gender of nouns. Alternatively, learners may be aware that gender is part of the lexical information they need to acquire about a noun along with its spelling, meaning and pronunciation, but they struggle with it due to the ambiguity of the input. The lack of communicative value of grammatical gender may be another explanation.

Fifth, there was no proficiency effect for the feminine or masculine stimuli, although the Low group outperformed the others on masculine plural stimuli.

Proficiency generally appears to be a poor predictor of the participants' performance which is unusual and may be due to the complexity and ambiguity of vocalic endings expressing gender and number.

Finally, the L2 groups' performance was significantly different from the NS group's performance. Their best average (80.6% for the Low group on ambiguous masculine plural) is well below the NS performance that is consistently at ceiling except for ambiguous feminine singular and plural stimuli, but still acceptable (90% and 91.1%, respectively), given that NS participants are expected to perform at least 90% accuracy (e.g. Dronjic & Helms-Park, 2014).

Similarly to previous studies conducted with L2 learners of Italian, the transparency of vocalic endings was a significant factor in the performance of our L2 learners, but only for masculine plural (72.86% average), and marginally for feminine singular (66.23%). The contrast with transparent masculine singular (38.36%) is striking and puzzling. We do not have an explanation for it aside from the fact that our participants had been studying Italian for only 6 to 24 months. More advanced learners may perform better. They may also have a better command of morphology that would allow them to utilize the gender and number information provided by the suffixes. Our participants' performance appears to be erratic ranging from 27.7% (masculine singular) to 72.87% (masculine plural) on stimuli with a suffix; they actually did much better on masculine singular stimuli without a suffix (61.13%), but obtained a higher accuracy on feminine plural stimuli with a suffix (51.43%) than without (38.43%).

Our mixed findings with low proficiency learners are inconclusive regarding the FRA, but other studies have argued it can account for their findings such as Spinner (2013) and Spinner et al. (2017) reviewed above.[7] As Spinner et al. point out, the FRA is not yet sufficiently elaborated to predict how particular features may be acquired. In the case of conflated features such as gender and number in Italian and Swahili, L2 learners have to both detect how these two features are expressed and then reassemble them. If Lardiere (2009, p. 214) is correct is assuming that "any feature contrast that is detectable is, in principle, ultimately acquirable", L2 learners should be eventually successfully, but we need to be more specific about what what detection may mean. Lardiere (ibid) adds that "the basis for detectability is the observation of any formal contrast, such as the difference between *student/students*". In the case of Italian vocalic endings, detectability may not even be the issue for Anglophone learners because strong final vowels are bound to stand out for native speakers of a language that lost them due to phonetic and morphological changes

7. See also El-Ghazoly (2013) for gender and number in Arabic as well as Hwang & Lardiere (2013) for plural in Korean.

resulting in word-final schwa (Minkova, 1991). However, the input is ambiguous at best, as described above.

We would like to propose that L2 learners' difficulties may be explained by the one-to-one principle as defined in Andersen (1984, p. 79): "Stated simply, the one-to-one principle specifies that an IL system should be constructed in such a way that an intended underlying meaning is expressed with one clear invariant surface form (or construction)". Andersen (*ibid*: 89) cites the case of an Anglophone learner of L2 Spanish who initially "ignored the gender and number encoded in articles and used one simple definite article *la* (without intended feminine meaning) for both singular and plural and one indefinite article *un* (without any intended masculine meaning)". Similarly, L2 Italian learners may initially assume that each vocalic ending expresses only one gender or one number. As they progress in their interlanguage, they eventually learn that one form (i.e. one vocalic ending) may express not only both gender and number, but also different gender and number combinations.

Future studies with more L2 learners at more advanced proficiency levels and different elicitation tasks may obtain better findings to indicate whether they are ultimately able to acquire grammatical gender and number. For instance, a written production task may reveal that learners have acquired the appropriate gender for the nouns they use. However, one would face the methodological problem of distinguishing between gender assignment errors and gender agreement errors. This could be solved by administering two different elicitation tasks to the same L2 learner groups: a gender assignment task followed by a production task that would use only the stimuli to which learners attributed the appropriate gender.

Pending future studies, we would like to conclude by stressing the impact of the learning difficulties created by the sheer number of vocalic endings conflating gender and number features on a single portmanteau morpheme as well as the necessity to disambiguate potentially conflicting cues within a DP (e.g. *le belle foto* 'the beautiful photos'), particularly for learners who have a low tolerance for ambiguity. Research in individual differences among L2 learners has underscored how important learning (or cognitive) styles, strategies and affective variables are in their development (e.g. Dörnyei & Skehan, 2003). Tolerance of ambiguity is one of these personality variables (see Ehrman, Leaver, & Oxford, 2003 for a review). We follow Dewaele and Wei (2012) in adopting Budner's (1962) definition of tolerance of ambiguity as "the tendency to perceive ambiguous situations as desirable" and ambiguity itself as: (a) an ambiguous situation or stimulus as "one which cannot be adequately structured or categorized by the individual because of the lack of sufficient cues"; and either (b) "a completely new situation in which there are no familiar cues"; (c) "a complex situation in which there are a great number of cues to be taken into account"; or (d) "a contradictory situation in which different elements

or cues suggest different structures" (Budner, 1962, p. 30). All these definitions are relevant to grammatical gender in Italian. Empirical studies have found a link between tolerance of ambiguity and L2 learners' performance. For instance, in a large-scale study (n = 2 158) with monolingual, bilingual and multilingual speakers, Dewaele and Wei (2012) found positive links between multilingualism and tolerance of ambiguity, while ESL studies suggested that tolerance of ambiguity was a good predictor of proficiency (e.g. Atamanova & Bogomaz, 2014; Ezzati, 2016).

References

Acquaviva, P. (2009). The structure of the Italian declension system. In F. Montermini, G. Boyé, & J. Tseng (Eds), *Selected proceedings of the 6th Décembrettes* (pp. 50–62). Cascadilla Press.

Andersen, R. W. (1984). The one to one principle of interlanguage construction. *Language Learning*, 34(4), 77–95. https://doi.org/10.1111/j.1467-1770.1984.tb00353.x

Atamanova, I., & Bogomaz, S. (2014). Ambiguity tolerance as a psychological factor of foreign language communicative competence development. *Procedia – Social and Behavioral Sciences*, 154, 342–352. https://doi.org/10.1016/j.sbspro.2014.10.161

Ayoun, D. (2007). The second language acquisition of grammatical gender and agreement. In D. Ayoun (Ed.), *French applied linguistics* (pp. 130–170). John Benjamins. https://doi.org/10.1075/lllt.16.10ayo

Ayoun, D. (2010). Corpus data. Shedding the light on French grammatical gender…or not. *Eurosla Yearbook*, 10, 119–141. https://doi.org/10.1075/eurosla.10.08ayo

Ayoun, D. (2018). Grammatical gender assignment in French: Dispelling the native speaker myth. *Journal of French Language Studies*, 28, 113–148. https://doi.org/10.1017/S095926951700014X

Ayoun, D., & Maranzana, S. (2020). Italian grammatical gender: A corpus study from a second language perspective. *Italica*, 97(2), 306–333.

Ayoun, D., & Rothman, J. (2013). Generative approaches to the L2 acquisition of temporal-aspectual-mood systems. In R. Salaberry & L. Comajoan (Eds.), *Research design and methodologies in studies on L2 aspect* (pp. 119–156). Mouton de Gruyter. https://doi.org/10.1515/9781934078167.119

Bardel, C. (2004). Il progetto InterIta: l'Apprendimento dell'italiano L2 in un contesto svedese. In B. Erman, J. Falk, G. Magnusson, & B. Nilsson (Eds.), *Second language acquisition* (pp. 11–30). Almqvist and Wiksell International.

Berretta, M. (1990). Morfologia in italiano lingua seconda. In E. Banfi & P. Codin (Eds.), *Storia dell'italiano e forme dell'italianizzazione: Atti del XXIII Congresso internazionale di studi: Trento – Rovereto 18–20 maggio 1989* (pp. 181–201). Bulzoni.

Berretta, M., & Crotta, G. (1991). Italiano L2 in un soggetto plurilingue (cantonese-malese-inglese): Sviluppo della morfologia. *Studi Italiani di Linguistica Teorica e Applicata*, 20, 285–331.

Bianchi, G. (2013). Gender in Italian-German bilinguals: A comparison with German L2 learners of Italian. *Bilingualism: Language and Cognition*, 16(3), 538–557. https://doi.org/10.1017/S1366728911000745

Budner, S. (1962). Intolerance of ambiguity as a personality variable. *Journal of Personality*, 30(1), 29–50. https://doi.org/10.1111/j.1467-6494.1962.tb02303.x

Carstens, V. (2000). Concord in minimalist theory. *Linguistic Inquiry*, 31, 319–355. https://doi.org/10.1162/002438900554370

Carstens, V. (2010). Implications of grammatical gender for the theory of uninterpretable features. In M. Putman (Ed.), *Exploring crash-proof grammars* (pp. 31–57). John Benjamins. https://doi.org/10.1075/lfab.3.03car

Chini, M. (1995). *Genere grammaticale e acquisizione. Aspetti della morfologia nominale in italiano L2*. Angeli.

Chini, M. (1998). *Genuserwerb des Italienischen durch deutsche Lerner*. In H. Wegener (Ed.), *Eine zweite Sprache lernen: Empirische Untersuchunge zum Zweitspracherwerb* (pp. 39–58). Narr.

Chomsky, N. (1988). *Language and problems of knowledge: The Managua lectures*. The MIT Press.

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

Chomsky, N. (2001). Derivation by phase. In K. Hale & M. Kenstowicz (eds), *Ken Hale: A life in language* (pp. 1–52). The MIT Press.

D'Achille, P., & Thornton, A. M. (2003). La flessione del nome dall'italiano antico all'italiano contemporaneo. In N. Maraschio & T. Poggi Salani (Eds.), *Italia linguistica anno Mille, Italia linguistica anno Duemila. Atti del XXXIV Congresso internazionale di studi della Società di Linguistica Italiana, Firenze, 19–21 ottobre 2000* (pp. 211–230). Bulzoni.

D'Achille, P., & Thornton, A. M. (2008). I nomi femminili in *-o*. In E. Cresti (Ed.), *Prospettive nello studio del lessico italiano. Atti del IX Congresso Internazionale della Società di Linguistica e Filologia Italiana* (pp. 473–481). Firenze University Press.

De Martino, M., Bracco, G., & Laudanna, A. (2011). The activation of grammatical gender information in processing Italian nouns. *Language and Cognition*, 26, 645–776.

De Mauro, T., Mancini, F., Vedovelli, M., & Voghera, M. (1993). *Lessico di frequenza dell'italiano parlato*. Etatslibri.

Dewaele, J-M. (2015). Gender errors in French interlanguage. The effect of initial consonant versus initial vowel of the head noun. *Arborescence*, 5, 7–27. https://doi.org/10.7202/1032661ar

Dewaele, J-M., & Wei, L. (2012). Is multilingualism linked to a higher tolerance of ambiguity? *Bilingualism: Language and Cognition*, 16, 231–240. https://doi.org/10.1017/S1366728912000570

Dominguez, A., Cuetos, F., & Segui, J. (1999). The processing of grammatical gender and number in Spanish. *Journal of Psycholinguistic Research*, 28, 485–498. https://doi.org/10.1023/A:1023216326448

Dörnyei, Z., & Skehan, P. (2003). Individual differences in second language learning. In M. Long & C. Doughty (Eds), *The Handbook of second language acquisition* (pp. 589–630). Wiley Blackwell. https://doi.org/10.1002/9780470756492.ch18

Dressler, W., & Thornton, A. M. (1996). Italian nominal inflection. *Wiener Linguistische Gazette*, 55–57, 1–26.

Dronjic, V., & Helms-Park, R. (2014). Fixed choice word-association tasks as second language lexical tests: What native-speaker performance reveals about their potential weaknesses. *Applied Psycholinguistics*, 35, 193–221. https://doi.org/10.1017/S014271641200029X

Ehrman, M., Leaver, B. L., & Oxford, R. (2003). A brief overview of individual differences in second language learning. *System*, 31, 313–330. https://doi.org/10.1016/S0346-251X(03)00045-9

El-Ghazoly, B. (2013). Feature reassembly and forming syntactic ties: The acquisition of non-canonical agreement in Arabic L2 (Unpublished PhD dissertation). Indiana University.

Epstein, S., Flynn, S., & Martohardjono, G. (1996). Second language acquisition: Theoretical and experimental issues in contemporary research. *Behavioral and Brain Sciences*, 19, 677–714. https://doi.org/10.1017/S0140525X00043521

Ezzati, M. (2016). Exploring tolerance of ambiguity and grammar achievement of advanced EFL learners. *Journal for the Study of English Linguistics*, 4(2), 1–12. https://doi.org/10.5296/jsel.v4i2.9919

Foote, R. (2015). The production of gender agreement in native and L2 Spanish: The role of morphophonological form. *Second Language Research*, 31, 1–31. https://doi.org/10.1177/0267658314565691

Foundalis, H. (2002). Evolution of gender in Indo-European languages. In W. D. Gray & C. D. Schunn (Eds.), *Proceedings of the 24th annual meeting of the Cognitive Science Society* (pp. 304–309). Lawrence Erlbaum Associates.

Gess, R., & Herschensohn, J. (2001). Shifting the DP parameter: A study of anglophone French L2ers. In C. R. Wiltshire & J. Camps (Eds), *Romance syntax, semantics and their L2 acquisition* (pp. 105–119). John Benjamins. https://doi.org/10.1075/cilt.216.10ges

Goad, H., White, L., & Steele, J. (2003). Missing inflection in L2 acquisition: Defective syntax or L1-constrained prosodic representations? *Canadian Journal of Linguistics*, 48, 243–263. https://doi.org/10.1353/cjl.2004.0027

Gracanin-Yuksek, M. (2006). V-N compounds in Italian. A case of agreement in word formation. In C. Nishida & J-P. Montreuil (Eds.), *New perspectives on Romance linguistics, Vol. 1: Morphology, syntax, semantics and pragmatics. Selected Papers from the 35th Linguistic Symposium on Romance Languages (LSRL), Austin, Texas, February*, pp. 113–126. John Benjamins. https://doi.org/10.1075/cilt.275.10gra

Gudmundson, A. (2012). *L'accordo nell'italiano parlato da apprendenti universitari svedesi: Uno studio sull'acquisizione del numero e del genere in una prospettiva funzionalista*. Stockholm University.

Halle, M., & Marantz, A. (1993). Distributed morphology and the pieces of inflection. In K. Hale & J. S. Keyser (Eds), *The view from building 20: Essays in linguistics in honor of Sylvain Bromberger* (pp. 111–176). The MIT Press.

Hawkins, R. (2003). Representational deficit theories of adult SLA: Evidence, counterevidence and implications. Plenary talk given at EUROSLA, Edinburgh, UK, September 2003.

Hawkins, R., & Chan, C. (1997). Partial availability of Universal Grammar in second language acquisition: The failed formal features hypothesis. *Second Language Research*, 13, 187–226. https://doi.org/10.1191/026765897671476153

Hawkins, R., & Franceschina, F. (2004). Explaining the acquisition and non-acquisition of determiner-noun gender concord in French and Spanish. In P. Prévost & J. Paradis (Eds.), *The acquisition of French in different contexts. Focus on functional categories* (pp.175–207). John Benjamins. https://doi.org/10.1075/lald.32.10haw

Hawkins, R., & Hattori, H. (2006). Interpretation of English multiple wh-questions by Japanese speakers: A missing uninterpretable feature account. *Second Language Research*, 22, 269–301. https://doi.org/10.1191/0267658306sr269oa

Hawkins, R., & Liszka, S. (2003). Locating the source of defective past tense marking in advanced L2 English speakers. In R. van Hout, A. Hulk, F. Kuiken, & R. Towell (Eds), *The interface between syntax and lexicon in second language acquisition* (pp. 21–44). John Benjamins. https://doi.org/10.1075/lald.30.03haw

Haznedar, B., & Schwartz, B. (1997). Are there optional infinitives in child L2 acquisition? In E. Hughes, M. Hughes & A. Greenhill (Eds.), *Proceedings of the 21st Annual Boston University Conference on Language Development* (pp. 257–268). Cascadilla Press.

Hwang, S. H., & Lardiere, D. 2013. Plural-marking in L2 Korean: A feature-based approach. *Second Language Research*, 29(1), 57–86. https://doi.org/10.1177/0267658312461496

Küpisch, T., Müller, N., & Cantone, K. F. (2002). Gender in monolingual and bilingual first language acquisition: Comparing Italian and French. *Lingue e Linguaggio*, 1, 107–150.

Lardiere, D. (2008). Feature assembly in second language acquisition. In J. M. Liceras, H. Zobl, & H. Goodluck (Eds.), *The role of formal features in second language acquisition* (pp. 106–140). Lawrence Erlbaum Associates.

Lardiere, D. (2009). Some thoughts on the contrastive analysis of features in second language acquisition. *Second Language Research*, 25, 173–227. https://doi.org/10.1177/0267658308100283

Lee, E., & Lardiere, D. (2019). Feature reassembly in the acquisition of plural marking by Korean and Indonesian bilinguals. *Linguistic Approaches to Bilingualism*, 9(1), 73–119. https://doi.org/10.1075/lab.16018.lee

Loporcaro, M., Faraoni, V., & Gardani, F. (2014). The third gender of Old Italian. *Diachronica*, 31(1), 1–22. https://doi.org/10.1075/dia.31.1.01gar

Lyster, R. (2014). Using form-focused tasks to integrate language across the immersion curriculum. *System*, 54, 4–13. https://doi.org/10.1016/j.system.2014.09.022

Maiden, M., & Robustelli, C. (2000). *A reference grammar of modern Italian*. NTC Publishing group.

Marantz, A. (1997). No escape from syntax: Don't try morphological analysis in the privacy of your own lexicon. *University of Pennsylvania Working Papers in Linguistics*, 4, 201–225.

Masine, F., & Scalise, S. (2012). Italian compounds. *Probus*, 24, 61–91. https://doi.org/10.1515/probus-2012-0004

McCarthy, C. (2007). Morphological variability in second language Spanish (Unpublished PhD dissertation). McGill University.

Micheli, M. S. (2016). Sul plurale delle parole composte nell'italiano contemporaneo. *Studi di Lessicografia Italiana*, 33, 229–256.

Minkova, D. (1991). *The history of final vowels in English: The sound of muting*. Mouton de Gruyter. https://doi.org/10.1515/9783110889512

Montermini, F. (2003). Appunti sulla cancellazione di vocale in derivazione. In A. Bisetto, C. Iacobini, & A. M. Thornton (Eds), *Scritti di morfologia in onore di Sergio Scalise in occasione del suo 60° compleanno* (pp. 171–188). Caissa Italia.

Oliphant, K. (1998). Acquisition of grammatical gender in Italian as a foreign language. *The Canadian Modern Language Review*, 54, 239–262. https://doi.org/10.3138/cmlr.54.2.239

Pesetsky, D., & Torrego, E. (2007). The syntax of valuation and the interpretability of features. In S. Karimi, V. Samiian, & W. K. Wilkins (Eds.), *Phrasal and clausal architecture: Syntactic derivation and interpretation. In honor of Joseph E. Emonds* (pp. 262–294). John Benjamins. https://doi.org/10.1075/la.101.14pes

Picallo, C. (1991). Nominals and nominalization in Catalan. *Probus*, 3(3), 279–316. https://doi.org/10.1515/prbs.1991.3.3.279

Prévost, P., & White, L. (2000). Missing surface inflection or impairment in second language acquisition? Evidence from tense and agreement. *Second Language Research*, 16, 103–134. https://doi.org/10.1191/026765800677556046

Riente, L. (2003). Ladies first: The pivotal role of gender in the Italian nominal inflection system. *McGill Working Papers in Linguistics*, 17, 1–53.

Ritter, E. (1991). Two functional categories in noun phrases: Evidence from modern Hebrew. In S. D. Rothstein (Ed.), *Syntax and semantics 25: Perspectives on phrase structure* (pp. 37–62). Academic Press. https://doi.org/10.1163/9789004373198_004

Ritter, E. (1993). Where's gender? *Linguistic Inquiry*, 24, 795–803.

Rothman, J. (2008). Aspectual morphology use in adult L2 Spanish and the competing systems hypothesis: When pedagogical and linguistic rules conflict. *Languages in Contrast*, 8(1), 74–106. https://doi.org/10.1075/lic.8.1.05rot

Santoro, M. (2012). Morphological variability in interlanguage grammars: New evidence from the acquisition of gender and number in Italian determiner phrases and direct object pronouns. *The Canadian Journal of Applied Linguistics*, 15(1), 167–189.

Scalise, S. (1994). *Morfologia*. Il Mulino.

Schwartz, B., & Sprouse, R. (1994). Word order and nominative case in nonnative language acquisition: A longitudinal study of (L1 Turkish) German interlanguage. In T. Hoekstra & B. Schwartz (eds), *Language acquisition studies in generative grammar* (pp. 317–368). John Benjamins. https://doi.org/10.1075/lald.8.14sch

Schwartz, B., & Sprouse, R. (1996). L2 cognitive states and the Full Transfer/Full Access model. *Second Language Research*, 12, 40–72. https://doi.org/10.1177/026765839601200103

Sgroi, S. C. (2008). La mozione: Problemi teorici, storici e descrittivi. *Quaderni di Semantica*, XXIX(1), 55–118.

Spinner, P. (2013). The second language acquisition of number and gender in Swahili: A Feature Reassembly approach. *Second Language Research*, 29, 455–479. https://doi.org/10.1177/0267658313477650

Spinner, P., Foote, R., & Upor, R. A. (2017). Gender and number processing in second language Swahili. *Linguistic Approaches to Bilingualism*, 8(4), 446–476. https://doi.org/10.1075/lab.16042.spi

Stark, E. (2008). The role of the plural system in Romance. In U. Detges & R. Waltereit (Eds), *The paradox of grammatical change: Perspectives from Romance* (pp. 57–84). John Benjamins. https://doi.org/10.1075/cilt.293.04sta

Thornton, A. M., Iacobini, C., & Burani, C. (1997). *BDVDB. Una base di dati per il vocabolario di base della lingua italiana* [BDVDB. A database for the basic vocabulary of Italian], (2nd ed.) Bulzoni.

Tsimpli, I. M., & Dimitrakopoulou, M. (2007). The interpretability hypothesis: Evidence from *wh*-interrogatives in second language acquisition. *Second Language Research*, 23, 215–242. https://doi.org/10.1177/0267658307076546

Valentini, A. (1990). Genere e numero in italiano L2. In M. Berretta, P. Molinelli, & A. Valentini (Eds.), *Parallela 4: Morfologia/morphologie* (pp. 335–345). Gunter Narr.

Appendix A

Nominal classes based on LIP: gender, number and frequency of each noun (partially adapted from Gudmundson, 2012 who used the 77 100 nouns from the LIP oral corpus (De Mauro, Mancini, Vedovelli, & Voghera, 1993).

Class	Gender	SG	PL	Example	# Tokens in LIP	Token in %	Cum. freq.	# Lemmas
A	MSC	*-o*	*-i*	*libro*	30685	39.8%	39.8%	2595
B	FEM	*-a*	*-e*	*sedia*	24209	31.4%	71.2%	2051
C1	FEM	*-e*	*-i*	*chiave*	16568	21.5%	92.7%	1566
C2	MSC			*mare*				
C3	MSC&FEM			*insegante*				
D	FEM	*-à*		*abilità*	2281	2.96%	95.6%	236
E	MSC	*-a*	*-i*	*problema*	1122	1.46%	97.1%	48
F	MSC	cons.		*film*	540	0.70%	97.8%	54
G1	MSC	*-a*	*-i*	*artista*	340	0.44%	98.2%	91
G2	FEM	*-a*	*-e*	*artista*				
H1	FEM	*-i*		*analisi*	305	0.40%	98.6%	35
H2	MSC	*-i*		*domani*				
I	FEM	*-o*		*foto*	187	0.24%	98.9%	6
J	MSC	*-i*		*lunedi*	225	0.29%	99.2%	5
K	MSC-SG & FEM-PL	*-o*	*-a*	*paio*	193	0.25%	99.2%	9
L	MSC	*-è*		*caffè*	67	0.09%	99.5%	6
M	MSC	*-a*		*cinema*	51	0.07%	99.6%	7
N	FEM	*-ù*		*virtù*	12	0.02%	99.6%	4
except.					319	0.41%	100%	25
Total					**77104**	**100%**		**6738**

CHAPTER 6

Grammatical gender and article use in beginning learners of German

Peter Ecke
University of Arizona

Acquiring nouns' grammatical gender, corresponding articles, and their morphological variants is one of the biggest challenges for adult learners of German. In this study, we analyze determiner noun phrases (NPs) produced orally by Spanish L1/English L2 beginning learners of German L3. We explore the extent to which L3 learners correctly produce determiner NPs including the appropriately marked grammatical gender, the extent to which they have become sensitive to nouns' formal gender cues, and the factors that help them cope with the complexity of the German article and gender systems. We pay special attention to the possibility of cross-linguistic influence (CLI) in L3 learners' article/gender choices according to predictions made by the Parasitic Model of vocabulary acquisition. The results suggest that learners reduce the complexity of the article/gender selection task by resorting to CLI, the adoption of gender from assumed lexical equivalents, and by over-using the most frequent forms (*ein*-MASC/NEU and *die*-FEM) which also are the phonologically least complex and most similar forms to equivalents in the L1 and/or L2. Formal gender cues only have a limited effect on article choice for the learners at this early stage of acquisition. Sensitivity to gender cues emerges only for the most frequent cue types and in association with definite (not indefinite) article use.

Keywords: articles, grammatical gender, errors, German, second language, third language

1. Introduction

Grammatical gender is an inherent part of the lexical entry of nouns in native speakers of languages like German, French or Spanish (Ayoun, 2007; Vigliocco, Antonini, & Garrett, 1997). It is part of the lemma or syntactic frame (the grammatical specification) of the entry (Hall & Schultz, 1994; Hohlfeld, 2004; Jackendoff, 2002; Levelt, Roelofs, & Meyer, 1999). For learners of German as a second language (L2), acquiring nouns' grammatical gender, corresponding articles, and

https://doi.org/10.1075/sibil.63.06eck

their morphological variants is one of the biggest learning challenges. Even at the intermediate and advanced levels of proficiency, they show persistent deviations from L1 speakers when they are asked to assign gender to nouns through the corresponding articles (Grauberg, 1971; Rogers, 1987; Walter & MacWhinney, 2015). Some researchers claim that even advanced learners of German are, for the most part, unable to represent gender in the lexical entry and that they cannot process it as automatically and subconsciously as native speakers do (Bordag, Opitz, & Pechmann, 2006), and that even native speakers can experience retrieval difficulties in gender assignment tasks (Ayoun, 2018).

Relatively little is known about how novice L2 learners of German cope with the complex gender and article system in the first few months of learning German. This study seeks to contribute to our knowledge about how first-semester students of German acquire and productively use grammatical gender and determiner noun phrases (NPs). We analyze NPs produced in relatively free speech by learners of German L3 who have Spanish L1 and English L2 and explore the degree to which the novice L3 learners correctly produce determiner NPs including the appropriately marked grammatical gender, the extent to which the L3 learners have become sensitive to nouns' morphological and phonological gender cues, and the factors that help learners cope with the complexity of definite and indefinite article usage in determiner NP production. We pay special attention to the possibility of cross-linguistic influence (CLI) in learners' gender assignment and article choice according to predictions made by the Parasitic Model of vocabulary acquisition (Hall & Ecke, 2003).

2. Grammatical gender and articles in German

In German, nouns are marked for case (nominative, genitive, dative, accusative), number (singular, plural), and grammatical gender (masculine, feminine, neuter). The three grammatical categories are expressed in a complex morphological system of articles, adjectives, and noun endings (Baten & Lochtman, 2014; Diel, Albrecht, & Zoch, 1991; Walter & van Compernolle, 2017). Because German places more importance on morphology than on word order, compared to English for example, acquiring the declension system including noun gender is of crucial importance to understanding as well as being understood (MacWhinney, Bates, & Kliegl, 1984). Table 1 illustrates the declension paradigm of German indefinite and definite articles.

Knowing or inferring the grammatical gender of nouns is a precondition for the declension system to work and a challenging learning task by itself. L2 learners of German perceive noun gender as the greatest impediment to the accurate use of

Table 1. Declension paradigm of indefinite and definite articles in German

Article	Case	Masculine	Feminine	Neuter	Plural
Indefinite	Nominative	*ein*	*eine*	*ein*	*(keine)*
	Genitive	*eines*	*einer*	*eines*	*(keiner)*
	Dative	*einem*	*einer*	*einem*	*(keinen)*
	Accusative	*einen*	*eine*	*ein*	*(keine)*
Definite	Nominative	*der*	*die*	*das*	*die*
	Genitive	*des*	*der*	*des*	*der*
	Dative	*dem*	*der*	*dem*	*den*
	Accusative	*den*	*die*	*das*	*die*

the declension system (Ritterbusch, LaFond, & Agustin, 2006). At the beginning stages of acquisition, L2 German learners only use nominative forms; accusative forms appear later as markers of non-subject forms (Baten, 2013; Diel et al., 2000). However, even second- and third-year students of German were shown to be unable to process accusative case marking in simple noun-verb-noun sequences when relevant case cues contradicted animacy and word order cues without having received explicit instruction (Jackson, 2007; VanPatten & Borst, 2012).

The low availability of accusative case marking and its restriction to masculine nouns result in a low cue validity that delays novice L2 learners' use of accusative article forms in German compared to L2 learners of Russian, for example, where case marking cues are more readily available and more reliable (Kempe & MacWhinney, 1998, 1999). Most textbooks of German do not introduce genitive and dative cases in first semester chapters although there has been a successful attempt to explicitly teach these to first semester students in a research setting (Walter & van Compernolle, 2017). Because of the (normally) late acquisition and introduction of genitive and dative cases, we focus in this study on the use of articles and gender in nominative and accusative forms of determiner NPs.

According to the Duden dictionary project, 46% of the German nouns listed in the dictionary are feminine (FEM), 34% are masculine (MASC) and 20% are neuter (NEU) (Duden, 2021). However, the frequency of occurrence of lexemes *der, die, das, ein* and *eine* in corpora is different because of the multiple functions the forms can assume. *Der*, for example, can occur as MASC/nominative but also FEM/genitive, FEM/dative, and plural (PLU)/all genders/genitive. The form *die* can occur as FEM/nominative or PLU/all genders/nominative. *Ein* can occur as MASC/nominative, NEU/nominative, and NEU/accusative.

Table 2 presents the frequency ranks and numbers of occurrences of the article forms (albeit used in different functions). The nominative definitive article forms are frequently used to teach and memorize the noun's gender through paired association.

Table 2. Word forms *der, die, das, ein* and *eine*, their frequency rank and number of occurrences in million according to Corpora Collection Leipzig University (2018)

	der	*die*	*das*	*ein*	*eine*
Rank	1	2	7	16	21
N of occurrences	21	19.7	6.3	4.6	4

Note. Based on material crawled in 2018 with 46,843,422 sentences

Grammatical gender in German has long been assumed to be completely arbitrary and extremely difficult to learn. While views about the arbitrariness of German gender persist, it has been shown that morphological, phonological, and semantic cues contribute to the assignment of gender in German (Corbett, 1991; Köpcke, 1982; Köpcke & Zubin, 1983, 1984; Kraiss, 2014; Mills, 1986). These cues occasionally compete with one another. The relation between cues and gender assignment is complex and the predictability of gender based on cues varies. Semantic cues include natural (biological) gender and different kinds of semantic categories. For example, *der Mann* 'the man' is masculine whereas *die Frau* 'the woman' is feminine. However, inanimate nouns are not predominantly neuter, and animate nouns are neuter when used in diminutive forms as in *das Männlein* 'the little man' and *das Mädchen* 'the girl'. Semantic categories can determine gender. For example, alcoholic beverages (except beer), days, months, and seasons are all masculine.

Below is a compilation of morphological and phonological gender cues that were used and coded in the present study's analyses. One question of interest was to what extent the morphological and phonological cues listed below as well as natural gender relate to rates of accurate grammatical gender markings. Cues were adopted from Menzel (2003) and Walter and MacWhinney (2015). Both studies relied heavily on Köpcke and Zubin's work (1983, 1984).

Phonological and morphological gender cues are as follows:

- Feminine: Nouns ending in *-e*,*-ei*[100], *-heit*[100], *-ie*[100], *-ik*[100], *-in*, *-keit*[100], *-schaft*[100], *-tät*[100], *-tion*[100], *-ung*, a subgroup/exception of words ending in *-el*, *-er*
- Masculine: Monosyllabic nouns ending in a consonant, nouns ending in *-el*, *-er*, *-ich*[100], *-ist*[100], *-(m)us*[100], *-ling*[100], a subgroup/exception of words ending in *-e*, *-en*
- Neuter: Nouns ending in *-a*, *-at*, *-chen*[100], *-en*, *-ment*, *-lein*[100], *-o*, *-um*, nouns beginning with *Ge-*, a subgroup/exception of nouns ending in *-er*, *-e*, *-el*

The reliability of these cues varies considerably. For example, the ending *-heit* (e.g. *Schönheit* 'beauty') is always feminine, whereas the ending *-e* (e.g. *Rose* 'rose') is feminine in the majority of cases, but there are exceptions (e.g. *Name* 'name-MASC').

Endings with a 100% reliability according to Walter and MacWhinney (2015) are marked with the superscript [100].

The grammatical gender and article systems of our participants' other languages, Spanish and English, differ from German in significant ways. Gender classes and corresponding indefinite and definite articles in nominative and accusative cases in the three languages are illustrated in Table 3.

Table 3. Grammatical gender, indefinite and definite articles in Spanish, English, and German

Article		Spanish	English	German
Indefinite	Singular	*un*-MASC *una*-FEM	*a*/*an*	*ein* (*einen**)-MASC *eine*-FEM *ein*-NEU
	Plural	*unos*-MASC *unas*-FEM	–	(*keine***)
Definite	Singular	*el*-MASC *la*-FEM	*the*	*der* (den*)-MASC *die*-FEM *das*-NEU
	Plural	*los*-MASC *las*-FEM	*the*	*die* (not marking gender)

Note.
* changed article form in accusative case
** negation used as article word in plural

Spanish nouns are either masculine or feminine. Indefinite articles, *un*-MASC and *una*-FEM, and definite articles, *el*-MASC and *la*-FEM, mark the nouns' gender in singular, whereas definite articles *los*-MASC and *las*-FEM mark it in the plural. There is no change of the articles' forms depending on grammatical case. English does not have grammatical gender, but semantic gender is marked in the pronoun system, for example, someone may refer to a ship as *she*. Thus, there is only the indefinite article *a* (*an* used in prevocalic position). The only definite article is *the* which is both singular and plural.

3. The acquisition and processing of gender and articles in German L2/L3

3.1 Gender processing in advanced learners of German

Non-native speakers of German differ from L1 speakers of German in how they assign and process grammatical gender. In experiments in which they had to assign grammatical gender to German pseudo-words and anglicisms, L1 speakers of German used both formal and semantic cues to assign gender, whereas L2 learners

of German primarily relied on semantic cues (Delisle, 1985). However, in a similar (off-line) study, Levine (1999) demonstrated that near-native speakers of German could assign gender to words with affixed gender cues quite similarly to native speakers.

Walter and MacWhinney (2015) tested advanced English L1 learners of German L2 for their knowledge of phonological, morphological, and semantic gender cues, and found that most learners had not attained full control over even the most reliable cues. The highest rates of correct nominal gender assignment were obtained when natural and grammatical gender coincided. This suggests that learners transfer semantic gender from L1 English to L2 German. Phonological cues were the most difficult ones for these learners followed by morphological and semantic cues.

Using picture naming and grammaticality judgment tasks, Bordag et al. (2006) investigated how the phonological endings of German words affected grammatical gender processing in groups of German L1 speakers and English L1 learners of German L2. Whereas they did not find any effects of noun endings on gender processing in L1 speakers, they found evidence that L2 word endings play a significant role in L2 gender processing. L2 learners were faster when producing gender-marked NPs that contained a noun with a gender typical ending and slower when the noun had a gender atypical ending. Analogous results were found in the grammaticality judgment tasks.

In other timed picture naming tasks with Czech L1 learners of German L2, Bordag and Pechmann (2007) found evidence for both the influence of L1 grammatical gender and phonological cues on the processing of L2 German NPs. When Czech and German nouns were congruent in gender classes, they were processed faster than when they had incongruent genders. L2 learners had also less difficulty naming nouns with a gender-typical ending (reliable cue) than nouns with a gender-atypical ending.

Studies into the acquisition and processing of gender in closely related language pairs (German L1 and Dutch L2) and less related language pairs (Greek L1 and German L2) provided further evidence for L1 influence. In particular, cognate status and gender compatibility for translation equivalents across languages have been shown to affect the times needed to produce, comprehend (Lemhöfer, Spalek, & Schriefers, 2008; Lemhöfer, Schriefers, & Hanique, 2010) and translate (Salamoura & Williams, 2007) determiner NPs. In addition, it has been shown that German L1 learners of Dutch L2 frequently transfer German gender to Dutch nouns, especially when the translation equivalents are similar in form (cognates), but also to a lesser extent when they are not. Cognates that differ in their gender frequently result in L2 gender assignment errors as reflected by incorrect article choices (Lemhöfer et al., 2010). In the next section, we review what has been found about beginning learners' acquisition and processing of gender in German.

3.2 Gender processing in beginning learners of German

Speakers of languages such as Turkish and Japanese that do not have grammatical gender tend to avoid the use of articles in German (Jaensch, 2009; Pfaff, 1984; Menzel & Tamaoka, 1995). Speakers of languages that have grammatical gender such as French but whose gender system is configured differently have been assumed to either rely on cues from the L1 system or transfer the gender of the L1 translation equivalent (Diel et al., 1991) although the picture seems to be more complicated.

Menzel (2003) investigated the extent to which Japanese beginning learners of German L2 already recognized gender regularities based on formal cues. She presented first and second-year students of German with nouns that were known, unknown, or pseudowords. Upon presentation of single (pseudo)nouns on a computer screen, the participants had to name the appropriate definite article. Menzel found that accuracy of gender assignment was highest for nouns with gender cues *-er*-MASC (66%), *-e*-FEM (64%), *-o*-NEU (54%) and *-in*-FEM (54%) whereas correctness was lowest for nouns with cues *-ik*-FEM (21%), *-tion*-FEM (24%), *-ung*-FEM (25%) and *-chen*-NEU (27%) in spite of the high reliability of the latter cues. She also found that words ending in *-e*-FEM and *-er*-MASC were the most frequent words in the textbook used by the learners and argued that frequency of occurrence in the input must have been a major factor determining the results.

Kraiss (2014) conducted another study about gender cues using two groups of beginning English L1 learners of German L2. One group received explicit instruction about semantic, morphological, and phonological gender cues and gender frequency patterns, whereas the control group received no explicit instruction about gender cues. In a series of tasks, the participants of both groups had to assign the correct gender to bare nouns. In three of the five tasks, the experimental group outperformed the control group which Kraiss took as evidence for the effectiveness of explicit instruction of gender cues and regularities.

Thomoglou (2007) studied the potential influence of L1 gender in Greek L1/English L2 learners of German L3. The author had assumed that L2 influence could be ruled out because the L2 lacked grammatical gender and did not analyze it further. Greek has three grammatical gender classes: masculine, feminine and neuter like German. However, there is no systematic correspondence in gender assignment between Greek and German translation equivalents. Moreover, neuter is the most frequent gender in Greek whereas it is the least frequent one in German. In a role-play production task with first- and second-year students of German, Thomoglou found limited evidence for L1 influence which was mostly restricted to nouns whose gender was difficult to predict or contradicted what formal cues would suggest. An overgeneralization of the most frequently occurring neuter gender was not found.

In a study of lexical errors in spoken L3 German, Hall and Ecke (2003) reported instances of L1 influence on L3 article selection and gender assignment. Participants were Spanish L1, English L2 beginning learners of German L3. In a production task, the L3 learners occasionally produced NPs suggestive of L1 gender influence. Examples (1) to (3) illustrate such L1 influence.

(1) **der bad*
'the-MASC bathroom' (target: *das*-NEU *bad*)
LI CLI source: *el*-MASC *baño*

(2) **die haus hat drei schlafzimmer*
'the-FEM house has three bedrooms' (target: *das*-NEU *haus*)
LI CLI source: *la*-FEM *casa*

(3) **die italienische essen ist sehr gut*
'the-FEM Italian food is very good' (target: *das*-NEU *essen*)
L1 CLI source: *la*-FEM *comida*

Hall and Ecke (2003) coded these errors as indicators of CLI at the syntactic frame (lemma) level along with other deviations, e.g. violations of subcategorization rules, subcategorized prepositions, count/non-count specification etc. They were interested in how frequently syntactic frames of relatively new L3 words were affected by CLI from the L1, L2, and L3 as potential sources and how frequent this was in relation to CLI and its sources at form and meaning levels of representation. Whereas they found most CLI overall to come from the typologically close L2 English affecting meaning and combined form/meaning levels, the syntactic frame level (including gender), in contrast, was influenced primarily by L1 Spanish structures.

3.3 The Parasitic Model and gender transfer

The model that Hall and Ecke used to interpret CLI-based errors is the Parasitic Model of L2 and L3 vocabulary acquisition (henceforth PM) (Ecke & Hall, 2014: Hall & Ecke, 2003; Hall & Schultz, 1994). The PM acknowledges the pervasiveness of CLI in L2/L3 acquisition and assumes that learners (unconsciously) detect and use similarity between new word forms and already known forms. Once learners detect a new word's similarity in (phonological or orthographic) form and/or meaning with an already known item, they will (by default) assume equivalence and create connections between the two forms using the stable one as a host representation to anchor the new form in the lexicon. That way it can be accessed as soon as possible for communicative purposes. In conjunction with assumed similarity of form and meaning, learners will also expect similarity in the items' syntactic frame, i.e. their grammatical specification which includes, e.g. word class,

subcategorization frames, and grammatical gender in the case of nouns (Hall & Reyes Duran, 2009; Hall et al., 2009). Assuming equivalence and forming parasitic connections will usually reduce the learning effort and speed up learning, i.e. establishing new representations and access routes. If equivalence is only partial (e.g. in phonological form or meaning and not in syntactic frame), the formed connections will lead to errors (deviations) like those illustrated in Examples (1)–(3), (4)–(5) and Figure 1. In Example 1 and Figure 1, the learner assumed equivalence between the new L3 form *bad* 'bathroom' and the L1 form *baño* 'bathroom' including its masculine gender leading to an incorrect use of the definite article *der*-MASC. The German word *bad* is neuter and requires the definite article *das*-NEU.

Importantly, according to the PM, forms of any previously learned language can serve as host representations if recognized as similar to the new form. In other words, a new form may be connected in parasitic fashion with forms and frames from the L1, the L2, or from within the L3 providing they are found to be similar and deemed equivalents.

To sum up, it appears that advanced learners of German differ from native speakers in the processing of gender in that (a) it is affected by similarity/cognate status of equivalent nouns, congruence of gender in L1 and L2 equivalents; (b) learners not recognizing all gender cues and (c) needing more time to process them compared to native speakers who do so more accurately and more automatically. From the relatively few studies with learners at the novice level, there is some evidence that (d) beginning learners of German transfer article and gender specification from known languages to new L2 and L3 lexical items and that (e) they start becoming sensitive only to a limited set of gender cues that frequently occur in the input. (f) Explicit instruction of gender cues appears to have a positive effect on gender assignment in L2 German, at least for a short time.

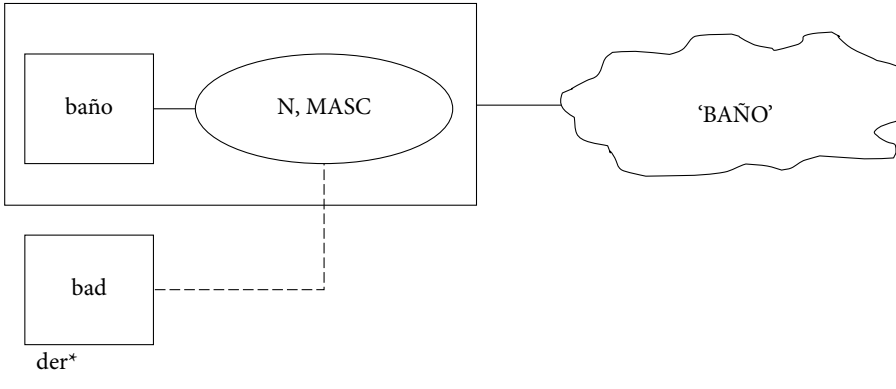


Figure 1. Example of a parasitic connection between a new L3 form and the syntactic frame (including grammatical gender) of its L1 equivalent

4. Methods

The present study contributes to our understanding of article/gender development in German by collecting spontaneous speech production data from an under-researched learner group (L1 Spanish learners of L3 German) at the beginning (novice) stage of learning and by using the PM as a theoretical framework to interpret the data.

We investigated L3 learners' productions of determiner NPs to learn how frequently gender errors are influenced by transfer (CLI, as predicted by the PM) and how frequent these CLI-based errors are in relation to correct gender assignments and gender errors that appear to be affected by factors other than CLI. Since the participants' L3 vocabulary knowledge can be assumed to be at the initial stage of vocabulary acquisition, we expected to detect a substantial amount of CLI as source of article/gender errors.

In addition to CLI, we analyzed to what extent the L3 German learners' sensitivity to formal (phonological and morphological) cues impacted correct gender assignment and if learners used (over)generalization as a strategy to reduce the complexity of article choice and gender assignment in oral production. Particularly, we explored the following questions.

4.1 Research questions

1. To what degree are definite and indefinite articles and grammatical gender used correctly in spoken production by novice learners of German L3?
2. What is the role of cross-linguistic influence (CLI) from previously acquired words' syntactic frames, particularly L1 gender, in the causation of errors with definite and indefinite L3 articles?
3. To what extent do other factors (the reduction/incomplete production and over-generalization of articles) affect L3 learners' selection and erroneous production of definite and indefinite articles including gender markings?
4. To what extent have learners of L3 German become sensitive with respect to nouns' morphological and phonological cues and their correspondences with grammatical gender classes?

4.2 Participants

Participants were 123 Mexican L1 Spanish-speaking learners of L3 German with advanced L2 English skills. All participants were undergraduate students enrolled at a private university in central Mexico. At the time of the oral production interview,

they had completed about 60 hours of classroom instruction in German. Their English proficiency was established through the institution's placement policy: to receive permission for enrollment in beginning German courses, students had to demonstrate a TOEFL score of at least 500 or have completed the institution's advanced English course. The screening of students' performance in a German chapter test including a free writing task at the beginning of the semester did not reveal any false beginners, that is, students who had received prior instruction in German.

4.3 Tasks and procedure

We recorded, transcribed, and analyzed L3 learners' speech produced in final oral examinations that were conducted as five-minute interviews by the instructor of first semester courses of German. The courses adopted a communicative approach and focused on meaning. The nominative forms of indefinite and definite articles were introduced along with the three gender classes.

Students were told that they had to memorize the gender/definite article of new nouns but were not instructed about gender cues. Accusative forms of the masculine articles were introduced in the second part of the course in a chapter on eating and drinking (purchasing, ordering, and talking about food). Dative and genitive cases and corresponding forms of articles were not introduced and taught explicitly at this early stage of acquisition.

In the interviews, participants were prompted to talk about their families, studies, hobbies, daily routine, eating preferences, and their home, which were all main topics covered in the course. The interviewer asked questions such as: *Hast du Geschwister?* 'Do you have siblings?' *Was studierst du?* 'What do you study?' *Was machst du am Wochenende (morgens, mittags und abends)?* 'What do you do on weekends (in the morning at noon and in the evening)?' *Wo wohnst du?* 'Where do you live?' *Welche Zimmer sind in deinem Haus?* 'What rooms does your house have?'

All NPs were identified and coded with respect to whether they were used correctly or incorrectly with or without the required article. Consistent with what has been taught, only NPs in nominative and accusative case positions were produced and analyzed. Definite and indefinite articles marked for accusative feminine (*die, eine*) and accusative neuter (*das, ein*) are identical to corresponding forms in the nominative. Only definite and indefinite articles for masculine nouns change from nominative *der* and *ein* to *den* and *einen* in the accusative case. These cases were extremely rare in the data (0.9% of NPs), but were coded because accusative forms had been introduced in the course.

NPs were coded with respect to the accuracy of the used determiner based on the context and factors that may have affected the use of articles in incorrectly

produced determiner NPs. We computed the accuracy rates of nouns used without any determiners, with an indefinite article as well as with a definite article. We also established the accurate uses of masculine, feminine, and neuter nouns with the definite articles: *der*-MASC, *die*-FEM, and *das*-NEU as well as nouns used in the plural with the definite article *die* which does not mark grammatical gender (see Table 3).

We established accuracy rates for the indefinite articles *ein*-MASC, *ein*-NEU, and *eine*-FEM. In addition to these indefinite articles, we included other '*ein*-words' that mark gender exactly like the indefinite articles: (1) possessive determiners were used almost exclusively in the first person: *mein* 'my-MASC'/'my-NEU' and *meine* 'my-FEM'. (2) the pronouns *kein*-MASC/-NEU, *keine*-FEM, and *keine*-PLU express negation of a noun, e.g. *ein Haus* 'a-NEU house' becomes *kein Haus* 'no-NEU house' when negated, and *Häuser* 'houses' become *keine Häuser* 'no houses' in negated form.

We analyzed all incorrect choices of determiner NPs to determine for each individual case if any of the following factors could have led to the incorrect use of the determiner:

a. L1 influence: The gender of the L1 equivalent was used. See Examples (4–5).
b. L2 influence: A determiner or a form closely resembling it from the L2 was used (e.g. *the/de, e/a/an*). See Examples (6–7).
c. L3 influence: A determiner of an L3 word, similar in form, was used. See Examples (8–9).
d. Reduced forms: *de* used instead of *der*; *e, a, an* or *n* used instead of *ein*; and *ne* used instead of *eine*. See Examples (10–11).
e. Avoidance: No determiner was used in a context where one was required.
f. Overgeneralization of the indefinite articles (*ein*-MASC/NEU or *eine*-FEM).
g. Overgeneralization of the definite articles (*der*-MASC, *die*-FEM/PLU or *das*-NEU).

Examples of L1 Spanish influence

(4) **der esszimmer*
'the-MASC dining room' (target: *das*-NEU *esszimmer*)
LI CLI source: *el*-MASC *comedor*

(5) **eine tisch*
'a-FEM table' (target: *ein*-MASC *tisch*)
L1 CLI source: *una*-FEM *mesa*

Examples of L2 English influence

(6) **the wohnzimmer*
'the living room' (target: *das*-NEU *wohnzimmer*)
L2 CLI source: *the living room*

(7) **de bilder*
'the pictures' (target: *die*-PLU *bilder*)
L2 CLI source: *the pictures*

Examples of L3 German influence

(8) **der küchen*
'the-MASC kitchen (blend with cake) (target: *die*-FEM *küche*)
L3 CLI source: *der*-MASC *kuchen* 'cake'

(9) **die kinderzimmer*
'the-FEM children's room' (target: *das*-NEU *kinderzimmer*)
L3 CLI source: *die*-PLU *kinder* 'children'

Examples of reduced forms

(10) **de treppe*
'the staircase' (target: *die*-FEM *treppe*)

(11) **e slafzimmer*
'a bedroom' (target: *ein*-NEU *schlafzimmer*)

Notice that some errors were coded into two or more categories. For example, a form, such as *de* (as in Examples 7 and 10) was coded as 'reduced form' and as 'L2 influence' whereas *der*-MASC in Example 4 was coded as 'L1 influence' and 'over-generalization of *der*'. Consequently, the percentages for factors that affect error production in the results section add up to more than 100 percent.

4.4 Statistical analyses

Accuracy rates of correct article use were tested for significant differences via two-dimensional c*hi-square* tests. The dependent variable was accuracy of article production (the frequencies of correct and incorrect article uses). Independent variables were the gender/article classes (masculine, feminine, neuter, plural) and the cues by gender (no cue, masculine, feminine, neuter). The null hypotheses assumed that there was no association between the accuracy of article choice and article/gender class, and no association between accuracy of article choice and cues by gender. We adopted $p < 0.05$ as confidence level for significance for all analyses including the display of factors contributing to errors (the incorrect uses of articles).

5. Results

Overall, 2,218 nouns/NPs were produced by the 123 participants. In a first general analysis, we examined the NPs in context to determine whether no determiner, a definite article or an indefinite article was required. Further we established if the NP had to be marked as masculine, neuter, feminine, or as plural. Then we determined the percentages of correct uses of these attributes in the given context. Results are illustrated in Table 4.

Table 4. Number of required determiner types and gender classes and overall accuracy of determiner use (n = 2,218 NPs)

	No article	Definite article	Indef. article	Masculine	Feminine	Neuter	Plural
n	511	182	1525	663	509	495	40
accuracy in %	98%	84%	80%	67%	54%	66%	50%

Of the 2,218 NPs, 1,567 (70.6%) were used appropriately (i.e. without an article, with definite or indefinite article, and with the correct gender as required by the context and the noun). The L3 learners experienced almost no problems producing NPs when no determiner had to be used. They were also quite accurate when a NP required a definite article (84% of the times) and an indefinite article (80% of the times). The high number of required indefinite articles compared to definite articles is most likely the result of question prompts that called for indefinite articles in the answers. As far as gender is concerned, masculine and neuter nouns were overall more often marked correctly than feminine nouns. Learners rarely produced plural nouns, and only half of them were target-like. We now turn to the analysis of indefinite article use.

5.1 Gender and indefinite article use

Table 5 presents the frequency and accuracy of indefinite article production.

Of the 1,225 NPs with indefinite articles, 900 (73.5%) were used with the correct article/grammatical gender. *Chi-square* testing showed significant differences in the frequencies of article accuracy, χ_2 (3) = 119.43, p < .001. The articles *ein*-NEU and *ein*-MASC were used significantly more often correctly than *eine*-FEM and *keine*-PLU. The 325 incorrectly produced indefinite articles were further analyzed with respect to the factors that may have affected the incorrect use of the article. Figure 2 presents the results.

Consistent with the high accuracy rates of *ein* with neuter and masculine nouns, *ein* is the indefinite article most frequently used in inaccurate NP productions (49%

Table 5. Incorrect and correct uses of indefinite articles

		Indefinite article (n = 1,225)			
		ein- MASC	*eine-*FEM	*ein-* NEU	*keine-* PLU
Incorrect	n	89_a	172_b	48_a	16_b
	%	18.6%	44.9%	14.4%	55.2%
Correct	n	390_a	211_b	286_a	13_b
	%	81.4%	55.1%	85.6%	44.8%

Note. Each subscript letter denotes a subset of Article categories whose column proportions do not differ significantly from each other at the .05 level.

of the times). The difference between the overuse of *ein* and L1 influence as well as all other factors is significant. In other words, the most frequent single source of error for indefinite articles is the overgeneralization of *ein*. The second most likely error source is L1 influence (38%) which is significantly more frequent than the overuse of *eine*-FEM and all other remaining factors. However, if cross-linguistic influences from L1, L2 and L3 are taken together in one category, then All CLI (62%) becomes the most frequent factor.

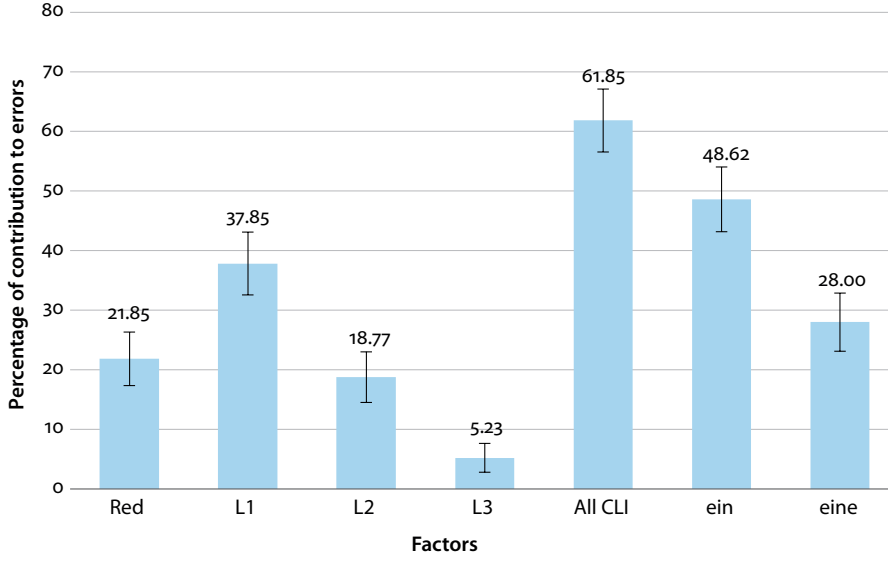


Figure 2. Factors contributing to errors with indefinite articles in percent and with 95% confidence interval

Note. N = 325 indefinite articles. Red = reduced forms; L1, L2, L3 = CLI from corresponding language; ALL CLI = CLI from L1, L2, and L3; ein, eine = generalization of corresponding articles.

5.2 Gender and definite article use

Table 6 illustrates the numbers and percentages of incorrect and correct uses of definite articles.

Table 6. Incorrect and correct uses of definite articles

		Definite articles (*n* = 265)			
		der-MASC	*die*-FEM	*das*-NEU	*die*-PLU
Incorrect	n	26_a	13_a	52_b	$2_{a, b}$
	%	30.2%	18.3%	52.0%	25.0%
Correct	n	60_a	58_a	48_b	$6_{a, b}$
	%	69.8%	81.7%	48.0%	75.0%

Note. Each subscript letter denotes a subset of Article categories whose column proportions do not differ significantly from each other at the .05 level.

Out of the 265 NPs produced with definite articles, 64.9% were used with the correct definite article/gender. A *chi-square* analysis showed significant differences in accuracy rates between article categories, χ_2 (3) = 22.58, $p < .001$. The accuracy of definite articles *die*-FEM (81.7%) and *der*-MASC (69.8%) were not significantly different from each other but differed significantly from *das*-NEU (48%). The accuracy of *die*-PLU (100%) was not found to be significantly different from other articles because of the small data sample. The 93 definite articles that were incorrectly used were further analyzed with respect to factors that may have led to the deviant use. The obtained findings are illustrated in Figure 3.

The single factors that most frequently contributed to L3 learners' errors with definite articles and gender are the overgeneralization of *die*-FEM/PLU (45% of the times) and L1 influence (44%). There are no significant differences in frequency between the two. Generalizing *der*-MASC (32%) as well as L2 influence (24%) were significantly less frequent than the top two factors. Using reduced forms (13%), overgeneralizing *das*-NEU (10%), and L3 influence (4%) were the least frequent factors. However, if influence from L1, L2 and L3 are taken together in one category, then All CLI (71%) becomes the most frequent factor.

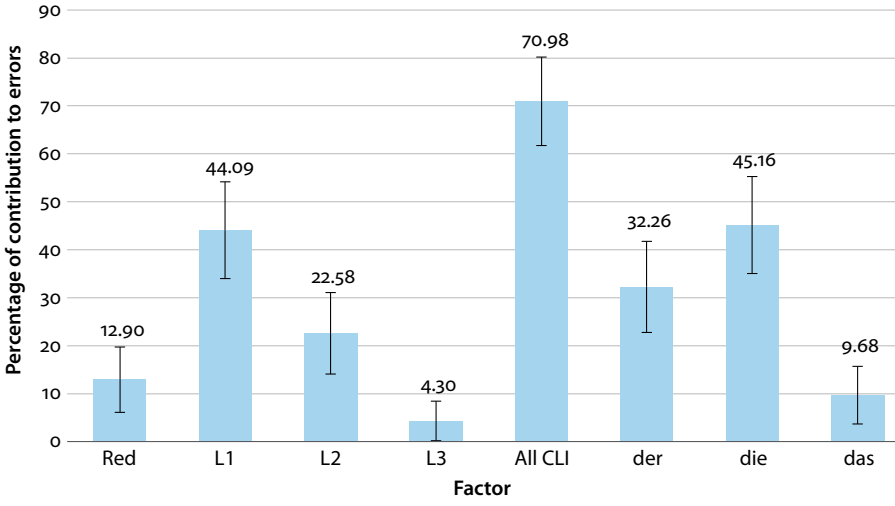


Figure 3. Factors contributing to errors with definite articles in percent and with 95% confidence interval

Note. $N = 93$ definite articles. Red = reduced forms; L1, L2, L3 = CLI from corresponding language, ALL CLI = CLI from L1, L2, and L3; der, die, das = generalization of corresponding articles.

5.3 Learners' sensitivity to gender cues

Our last analysis concerned L3 German learners' sensitivity to nouns' morphological and phonological cues and their correspondences with grammatical gender classes. It is important to note that the individual cue type comparisons must be taken with caution because some cue types (e.g. *-ung*, *-er*, *-en*) were produced sparsely, especially with definite articles. Nouns with feminine cues *-in*, *-ik*, *-ei*, *-heit*, *-ie*, *-keit*, *-schaft*, *-tät*, *-tion*, with masculine cues *-ist, -ling*, *-(m)us*, and with neuter cues *-a*, *-at*, *-chen*, *Ge-*, *-ment*, and *-um* were not produced at all or so rare (produced less than five times across both article types) that they were excluded from the results. For comparative purposes, we display the correct use of indefinite and definite articles by gender cue type together in Figure 4 (see the Appendix for more detailed results).

The accuracy of indefinite articles is highest for nouns with neuter cues *-en* and neuter exception *-er*, masculine exception *-en*, masculine *-ich*, *-el*, natural masculine gender, and masculine single-syllabic nouns ending in consonants, but only *-en* and exception *-er* differ significantly from all other cue types. On the other hand, exception *-er*-FEM and most importantly the frequently produced *-e*-FEM are used significantly less often correctly than all masculine cue types.

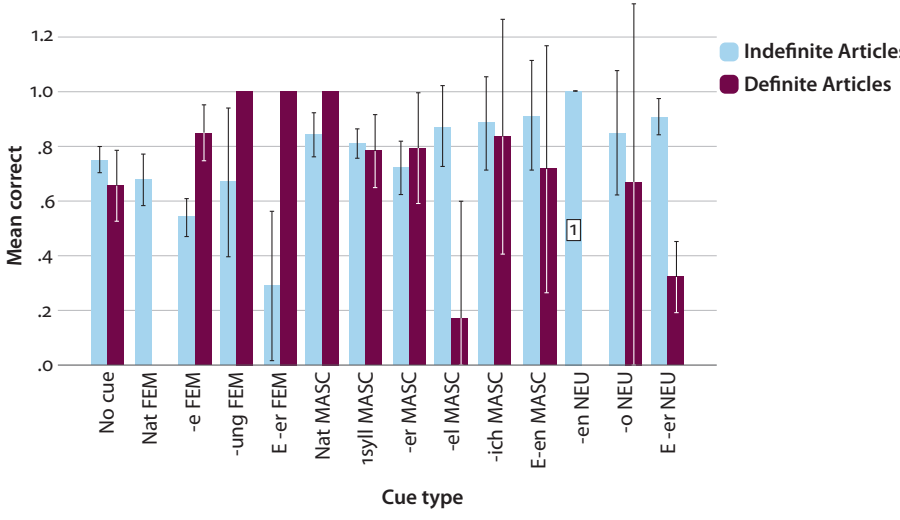


Figure 4. Accuracy of indefinite and definite article uses for specific gender cues

Overall feminine cue types show a significantly lower accuracy rate than masculine and neuter cue types that require *ein* as illustrated in Table 7. These data are consistent with the high accuracy percentages of indefinite articles *ein*-MASC/NEU compared to the lower accuracy of *eine*-FEM reported above.

Table 7. Incorrect and correct uses of indefinite and definite articles across cues for each gender class

			Cue by gender				
			No cue	Feminine	Masculine	Neuter	Other
Indefinite	Incorrect	n	83_{a}	136_{b}	81_{a}	10_{c}	$3_{a, c}$
		%	25.2%	42.4%	19.3%	9.1%	17.6%
	Correct	n	247_{a}	185_{b}	338_{a}	100_{c}	$14_{a, c}$
		%	74.8%	57.6%	80.7%	90.9%	82.4%
Definite	Incorrect	n	19_{a}	8_{b}	$21_{a, b}$	38_{c}	$2_{a, b}$
		%	34.5%	14.8%	26.3%	66.7%	20.0%
	Correct	n	36_{a}	46_{b}	$59_{a, b}$	19_{c}	$8_{a, b}$
		%	65.5%	85.2%	73.8%	33.3%	80.0%

Note. Each subscript letter denotes a subset of Cue by Gender categories whose column proportions do not differ significantly from each other at the .05 level value. For indefinite articles: $\chi_2(4) = 71.19$, $p < .001$, for definite articles: $\chi_2(4) = 38.76$, $p < .001$.

The accuracy percentages for definite articles show a different picture: feminine nouns with gender cues *-ung*, exception *-er* and *-e* were among the ones most frequently produced accurately along with natural masculine nouns (see Figure 4). Most masculine cues yielded high accuracy rates as well with definite articles except *-el*. The accuracy of neuter exception *-er*, which was high for indefinite articles, was very low (and significantly different from feminine *-e* and No cues), probably due to learners' assumption that *-er* is masculine and/or the general underuse of neuter *das*. In other words, the data from indefinite and definite article use are opposing in part as far as gender assignment accuracy is concerned. Masculine *-el* and neuter exception *-er* differ in accuracy percentages for indefinite and definite articles as feminine *-e* nouns do. The latter are mostly accurate when used with the definite article (85%), but much less so when used with the indefinite article (54%).

If gender cues are taken together for each gender class, the following picture emerges for definite article use: nouns with feminine gender cues are more often produced with the correct definite article than masculine nouns/cues (albeit not reaching significance) and masculine nouns/cues are significantly more often used with the correct definite article than neuter nouns/cues (see Table 7 and Figure 5).

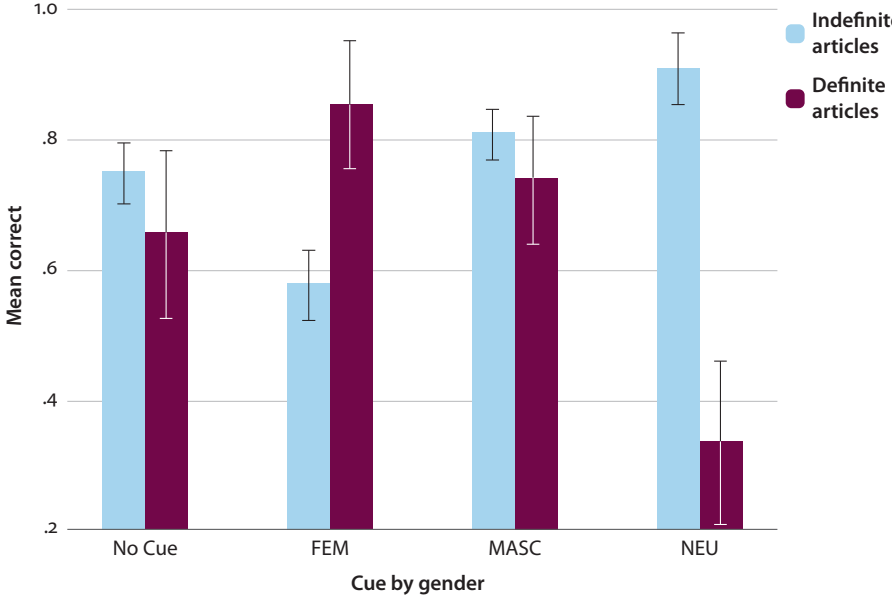


Figure 5. Correctness of indefinite and definite article use by gender cue type

6. Discussion

We investigated the strength of CLI on gender assignment and determiner use in relation to other factors, such as overgeneralization of articles and gender cue use. As a theoretical framework, we adapted the Parasitic Model of L2 and L3 vocabulary acquisition (Ecke, 2015; Hall & Ecke, 2003). This model assumes that CLI plays an import role in the early stages of lexical learning and describes how lexical form, syntactic frame, and meaning representations are formed, revised, and automatized. Initially, through the detection and use of similarity between new word forms and already represented ones, learners establish connections between assumed form equivalents which results in the adoption of the syntactic frame of the host representation (Hall & Reyes Duran, 2009; Hall et al., 2009). The syntactic frame of the host's lexical entry includes grammatical gender in the case of nouns (Ayoun, 2007; Vigliocco, Antonini, & Garrett, 2007). Grammatical gender and article influence from the L1 or the L2 can be expected to result in errors when L3 learners assume equivalence with items that are incongruent in their grammatical gender.

Thus, our research questions considered: (1) the extent to which definite and indefinite articles and their grammatical gender are used correctly in L3 production by novice learners of German; (2) the relative strength of CLI from previously acquired words' syntactic frames, particularly L1 grammatical gender; (3) other factors that affect L3 learners' production of definite and indefinite articles; and (4) the extent to which learners of L3 German have become sensitive to nouns' morphological and phonological cues and their correspondences with grammatical gender classes.

6.1 Accuracy of article use and sensitivity to gender cues

As far as gender assignment through indefinite article use is concerned, we found that *ein*-NEU and *ein*-MASC were most often used correctly (86% and 81% of the times respectively). The article *eine*-FEM was used significantly less accurately (55% of the times only). The use of the plural form *keine* (which negates the corresponding nouns) was rare and correct only 43% of the times. Gender assignment by means of definite articles showed a different picture: Articles *die*-FEM and *die*-PLU were accurately used 82% and 75% of the times respectively. *Der*-MASC, when used, was accurate 70% of the times whereas *das*-NEU was accurate only 48% of the times. The analysis of incorrectly used articles shows a pattern that is consistent with correct article uses: The most frequently over-used indefinite article was *ein*-NEU/MASC whereas the most frequently over-used definite article was *die*-FEM/PLU although the latter was significantly different only from *das*-NEU.

The opposing pattern – overgeneralization of masculine/neuter in indefinite contexts, vs. overgeneralization of feminine in definite contexts – suggests that grammatical gender has not yet been acquired, and that it does not significantly affect the selection and use of articles in beginning L3 learners. Since learners have not yet established gender representations, they need to process gender each time anew on the basis of a variety of cues and constraints that they are still quite unfamiliar with (cf. Bordag & Pechmann, 2007). The results of previous studies suggest that intermediate and advanced learners of German also have difficulties with gender assignment (Presson, MacWhinney, & Tokowicz, 2014; Rogers, 1987; Walter & MacWhinney, 2015).

As far as learners' sensitivity to gender cues is concerned, we also found differing results for indefinite and definite article use. Masculine *-el*, neuter exception *-er* and feminine *-e* differed widely in accuracy percentages for indefinite and definite articles. Only masculine cues *-er*, *-ich* and natural masculine gender showed consistently high accuracy percentages over 70% with both types of articles.

Spontaneous language production studies like this one have limitations with respect to what they can reveal about learners' sensitivity towards gender cues. While participants in experimental studies may be asked to associate an article/gender with all possible cues as stimuli, participants of naturalistic studies will only provide nouns, corresponding cues, and articles that their productive vocabulary and the context allow for. Consequently, only a subset of cues may be used by the participants as in the present study. Despite that limitation, the data on gender cues suggest that learners either start marking gender preferably (or exclusively) with the definite article and not (yet) with the indefinite article or that grammatical gender has not yet become a relevant category for learners to express. Articles may be used first and primarily to mark (in)definiteness (and not gender) as has been suggested for child L2 learners of German (Wegener, 1999). Only definite articles may begin being used to also express gender. The cues feminine *-e*, masculine *-er*, and masculine monosyllabic nouns ending in a consonant appear to be the ones that this study's participants start using first in association with the definite article. Feminine *-e* and masculine *-er* were also the cues with the highest accuracy percentages reported in Menzel (2003), and the cues that occurred most frequently in the nouns in the textbook used by the participants of her study. Our data on cue sensitivity are also consistent with the observed general overuse of *ein*-MASC/NEU and *die*-FEM and the underuse of *das*-NEU corresponding to general frequency of use patterns.

6.2 CLI and other factors affecting article use and gender assignment

Let us now turn to a discussion of CLI and other factors and their impact on L3 learners' erroneous production of indefinite and definite articles and corresponding genders. For incorrectly produced indefinite articles, the following hierarchy of factors evolved from the data. It displays the single factors that were potential error sources in order of strength with percentages of likelihood: generalization of *ein*-MASC/NEU (50%), L1 influence (38%), generalization of *eine*-FEM (28%), reduction of form (22%), L2 influence (19%), and L3 influence (5%). For the incorrectly produced definite articles, the hierarchy of factors looked as follows: generalization of *die*-FEM (45%), L1 influence (44%), generalization of *der*-MASC (32%), reduction of form (13%), generalization of *das*-NEU (10%) and L3 influence (4%).

The L3 learners in the present study appear to be subject to input frequency constraints. They tend to (over)use the articles that are most frequent in the input: indefinite article *ein* used both for masculine and neuter and *die* used both for feminine and plural. It has to be emphasized that it appears to be the determiners' form and not grammatical gender that is overgeneralized. If it was the latter, one would expect similar patterns of over-use of only one gender across definite and indefinite article productions which we did not find. Generalization based on overall frequency effects, particularly generalized *ein*-MASC/NEU and *die*-FEM/PLU have also been reported for child L1 acquisition (MacWhinney, 1978; Mills, 1986) and child L2 acquisition (Wegener, 1999) of German. The overgeneralization of a single article to reduce complexity may be a general factor that modulates the acquisition and use of articles across different learner groups. Note, however, that word form factors also interact with and potentially contribute to the (over)use of articles *ein*-MASC/NEU and *die*-FEM. This concerns the phonological complexity of the articles and their difficulty to be produced. Notice the consonant (C) – vowel (V) combinations of the three definite and two indefinite articles: *die* = CV, *der* = CV(C), *das* = CVC, *ein* = VVC, *eine* = VVCV. *Die* and *ein* are the least complex articles and can be expected to be the least difficult in spoken production. Finally, the phonological similarity of the indefinite article *ein* with its L1 equivalent *un* 'a-MASC' and L2 equivalent *a(n)* may further contribute to its identification and use as the unmarked form.

The second most frequent potential single error source was L1 influence: 38% of deviations with indefinite articles and 44% of errors with definite articles could be traced to L1 influence. We expected prior language influence to occur based on assumptions of the PM (Ecke & Hall, 2014): the detection of similarity between a new L3 form and a known form will lead to the assumption of equivalence of the two forms and the adoption of the syntactic frame (including gender) of the host. Note that the adoption of a frame may simply be achieved through a connection

to the host's representation. It does not require an independent gender node in a lexical entry of the L3 form. Quite to the contrary: the links to the L1 (or L2) frame may primarily reflect instances of what Hall and Ecke (2003) described as processing CLI: connections that help establish lexical access routes through mediating host representations for comprehension and production.

As an individual factor, L1 influence came in second in the hierarchy of factors affecting errors with article choice. However, if CLI is conceived as a more general category that subsumes influence from L1, L2 and L3 equivalents, then it becomes the most frequent error source that significantly exceeds all other factors and potentially affects a substantial 201 of 325 errors (62%) with indefinite articles and 72 of 93 errors (77%) with definite articles (cf. Figures 2 and 3). In other words, in about two thirds of all determiner NP errors, CLI is a possible error source.

An assumption of the Parasitic Model of vocabulary acquisition is that L3 learners can form cross-lexical connections between new L3 words with already known items from any language (L1, L2 or from within L3) providing that they detect and use similarity and assume equivalence of pairs in form and/or meaning (Hall & Ecke, 2003). The assumption of equivalence and creation of connecting links between lexical items leads to the adoption of the syntactic frame (including grammatical gender in the case of nouns).

Form similarity (i.e. a new item's phonological or orthographic similarity with an already known item), in particular, is expected to lead to a cross-lexical connection and syntactic frame adoption. In the case of L3 German, one would predict a substantial amount of form connections with words from the typologically closer L2 English than from L1 Spanish. Both Rothman's (2015) Typological Primacy Model and Bardel and Falk's (2007) L2 Status assumption would make similar predictions.

Some cases in the data illustrate how lexical form similarity can trigger L2 influence in the production of determiners. Cognates (words similar in sound and meaning) often trigger the overt intrusion of L2 articles in the L3 NP, for example: *a bruder* 'a brother' (target *ein*-MASC), *a problem* 'a problem' (target: *ein*-NEU) and *my vater* 'my father' (target: *mein*-MASC).

Clyne (1987) noticed similar cases of cognates that triggered the intrusion of L1 function words during code-switching. Hall (2002) illustrated the pervasiveness of cognate status and how it affects learners' assumptions about word meaning in a judgment task with pseudo-cognates. Hall et al. (2009) demonstrated cognate effects on L3 learners' initial assumptions about the syntactic frames (idiosyncratic grammatical properties) of recently encountered L3 words.

Lemhöfer and colleagues illustrated how cognate status affects learners' assumptions about grammatical gender in closely related L1 German and L3 Dutch and how cognate status and gender congruence reduces retrieval times in L3

production and comprehension tasks (Lemhöfer et al., 2008; Lemhöfer et al., 2010). In terms of the PM, these are all potential manifestations of processing CLI and a consequence of L2/L3 learners' detection and use of similarity at the initial stage of individual word acquisition.

The question is why there are not more cases of L2 influence from typologically similar English, but instead more frequent cases of L1 Spanish influence among the L3 determiner errors. One possibility is that in the absence of apparent L2-L3 word form similarity, learners automatically resort to and adopt L1 gender as part of the syntactic frame of the L1 equivalent. Spanish and German articles are more similar in that they mark grammatical gender whereas English articles do not. The recognition and use of this similarity would be consistent with the finding that beginning L1 Spanish/L2 English/L3 German learners' errors at the syntactic frame level in general are most frequently influenced by the L1 (and not the L2) (Ecke & Hall, 2000; Hall & Ecke, 2003).

Another possibility is that the L1 influence is conditioned by lexicon-external factors, such as the learners' meta-linguistic awareness of the L1, L2 and L3 gender and article systems. Let us presume that L3 learners of German in their very first attempts to acquire individual words assume equivalence with L2 words based on the more pronounced form similarity with L2 words than with L1 equivalents, but that they soon realize through explicit instruction or negative feedback that German articles encode gender unlike English and similar to Spanish. In subsequent word learning attempts, they may assume equivalence of L1 and L3 syntactic frames, genders, and articles. Once learners notice that L1-L3 frame equivalence is rare and that its assumption frequently results in errors, the learners may abandon or suppress CLI again and turn to a frequency-based strategy: They generalize and overuse just one article (usually the most frequent one of its type) to reduce the cognitive effort for dealing with the complex and (for learners) non-transparent German article/gender systems.

While this is pure speculation at this point, it may help illustrate the point we want to make: Learners may use a CLI-based default learning mechanism at the lexical level (at some early point in time) that is modulated or superseded by other factors (Hall & Ecke, 2003, p. 73). Longitudinal case studies combined with introspective methods (see e.g. Williams & Hammarberg, 1998) could help investigate the predominant strategies that learners use and perhaps replace with other more effective strategies over time. They could also shed light on individual learner differences and preferences with respect to strategy use to cope with the complexity of the German gender and article systems – an important issue that was not investigated here.

7. Conclusion

To conclude, we found that cross-linguistic influence (mostly from L1 and to a lesser extent L2 and L3) is a significant factor affecting article and gender selection (often resulting in errors) in beginning learners' L3 production of German. CLI from all learner languages represents the most frequent potential source for article selection errors. CLI interacts and competes with other factors, most importantly the overgeneralization of articles (not gender classes) based on frequency of occurrence in the input and possibly phonological complexity, similarity with L1/L2 equivalents, and pronounceability. Gender cues only seem to have a limited effect on article choice and gender assignment for learners at this very early stage. Learners seem to start associating the gender cues most frequent in the input with definite articles first. Indefinite articles appear to be used primarily or exclusively to express (in)definiteness, not gender (Wegener, 1999). Thus, our data support the view that in beginning L2/L3 learners of German, grammatical gender is only sparsely acquired and not yet represented as a feature in an independent L3 lexical entry's syntactic frame (Bordag & Pechmann, 2007). Instead, it may either be expressed through parasitic connections with the gender nodes in frames of (mostly) L1 host representations (Hall & Ecke, 2003), not be specified at all, or be consciously suppressed because of learners' awareness of the complexity and incompatibility of L1, L2, and L3 gender and article systems.

Future research should attempt to disentangle the impact of lexicon-internal processes, such as CLI, from lexicon-external factors, such as metalinguistic awareness or frequency-based generalizations by adding longitudinal case studies and introspective studies to on-line gender processing/assignment studies and error analyses. These would also facilitate investigating individual learner differences in the acquisition of the German article/gender systems.

The learners of German in this study reduce the complexity of the task to select articles and assign gender by resorting to CLI, the adoption of gender from assumed lexical equivalents, and by over-using the most frequent forms in the input which are also the (phonologically) least complex and most similar forms to equivalents in the L1 and/or L2. While the focus of this discussion has been on the interpretation of patterns of learners' erroneous uses of articles and genders, it needs to be emphasized that the L3 learners of this study were overall quite successful in producing accurate determiner NPs for the nominative and to some extent accusative case constructions that they had been exposed to thus far in a first semester course of German.

Acknowledgements

I would like to thank Dalila Ayoun and three anonymous reviewers for their helpful suggestions, Chris Hall for collaboration on earlier research that this project has built on, and Mark Borgstrom from the University of Arizona's Research Computing/Statistical Consulting team for his help with the statistical analyses.

Funding

Part of this research was funded through a grant from CONACYT [No. 25850-H] to the author.

References

Ayoun, D. (2007). The second language acquisition of grammatical gender and agreement. In D. Ayoun (Ed.), *French applied linguistics* (pp. 130–170). John Benjamins. https://doi.org/10.1075/lllt.16.10ayo

Ayoun, D. (2018). Grammatical gender assignment in French: Dispelling the native speaker myth. *Journal of French Language Studies*, 28, 113–148. https://doi.org/10.1017/S095926951700014X

Bardel, C., & Falk, Y. (2007). The role of the second language in third language acquisition. The case of Germanic syntax. *Second Language Research*, 23(4), 459–484. https://doi.org/10.1177/0267658307080557

Baten, K. (2013). *The acquisition of the German case system by foreign language learners*. John Benjamins. https://doi.org/10.1075/palart.2

Baten, K., & Lochtman, K. (2014). Das deutsche Kasussystem im Fremdsprachenerwerb: Ein Forschungsüberblick. *Muttersprache*, 124(1), 1–25.

Bordag, D., Opitz, A., & Pechmann, T. (2006). Gender processing in first and second languages: The role of noun termination. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 32(5), 1090–1101. https://doi.org/10.1037/0278-7393.32.5.1090

Bordag, D., & Pechmann, T. (2007). Factors influencing L2 gender processing. *Bilingualism: Language and Cognition*, 10(3), 299–314. https://doi.org/10.1017/S1366728907003082

Clyne, M. (1987). Constraints on codeswitching: How universal are they? *Linguistics*, 25, 739–764. https://doi.org/10.1515/ling.1987.25.4.739

Corbett, G. G. (1991). *Gender*. Cambridge University Press. https://doi.org/10.1017/CBO9781139166119

Corpora Collection Leipzig University. (2018). *German news corpus based on material crawled in 2018 with 46,843,422 sentences*. Retrieved on 10 June 2021 from: https://corpora.uni-leipzig.de/en?corpusId=deu_newscrawl-public_2018&word=

Delisle, H. (1985). The acquisition of gender by American students of German. *The Modern Language Journal*, 69(1): 55–63. https://doi.org/10.1111/j.1540-4781.1985.tb02527.x

Diehl, E., Albrecht, H., & Zoch, I. (1991). *Lernerstrategien im Fremdsprachenerwerb: Untersuchungen zum Erwerb des deutschen Deklinationssystems*. De Gruyter. https://doi.org/10.1515/9783111632285

Diehl, E., Christen, H., Leuenberger, S., Pelvat, I., & Studer, T. (2000). *Grammatikunterricht: alles für der Katz? Untersuchungen zum Zweitsprachenerwerb Deutsch*. Niemeyer.

Duden. (2021). Die Verteilung der Artikel (Genusangaben) im Rechtschreibduden. Retrieved on 10 June 2021 from https://www.duden.de/sprachwissen/sprachratgeber/Die-Verteilung-der-Artikel-Genusangabe-im-Rechtschreibduden

Ecke, P. (2015). Parasitic vocabulary acquisition, cross-linguistic influence, and lexical retrieval in multilinguals. *Bilingualism: Language and Cognition*, 18, 145–162. https://doi.org/10.1017/S1366728913000722

Ecke, P., & Hall, C. J. (2000). Lexikalische Fehler in Deutsch als Drittsprache: Translexikalischer Einfluss auf drei Ebenen der mentalen Repräsentation. *Deutsch als Fremdsprache*, 37(1), 30–36.

Ecke, P., & Hall, C. J. (2014). The Parasitic Model of L2 and L3 vocabulary acquisition: Evidence from naturalistic and experimental studies. *Fórum Linguístico*, 11(3), 360–372. https://doi.org/10.5007/1984-8412.2014v11n3p360

Grauberg, W. (1971). An error analysis in German of first-year university students. In G. E. Perren & J. L. M. Trim (Eds), *Applications of linguistics* (pp. 257–263). Cambridge University Press.

Hall, C. J. (2002). The automatic cognate form assumption: Evidence for the parasitic model of vocabulary development. *International Review of Applied Linguistics in Language Teaching (IRAL)*, 40(2), 69–87. https://doi.org/10.1515/iral.2002.008

Hall, C. J., & Ecke, P. (2003). Parasitism as a default mechanism in L3 vocabulary acquisition. In J. Cenoz, B. Hufeisen & U. Jessner (Eds), *The multilingual lexicon* (pp. 71–85). Kluwer. https://doi.org/10.1007/978-0-306-48367-7_6

Hall, C. J., Newbrand, D., Ecke, P., Sperr, U., Marchand, V., & Hayes, L. (2009). Learners' implicit assumptions about syntactic frames in new L3 words: The role of cognates, typological proximity, and L2 status. *Language Learning*, 59(1), 153–202. https://doi.org/10.1111/j.1467-9922.2009.00503.x

Hall, C. J., & Reyes Duran, A. (2009). Cross-linguistic influence in L2 verb frames: The effects of word familiarity and language proficiency. In A. G. Benati (Ed.), *Issues in language proficiency* (pp. 24–44). Continuum.

Hall, C. J., & Schultz, M. (1994). Los errores de marco sintáctico: Evidencia del modelo parasitario del léxico mental en un segundo idioma. *Estudios en Lingüística Aplicada, Memorias del II Congreso Nacional de Lingüística* (pp. 376–389). UNAM, INAH and ECM.

Hohlfeld, A. (2004). Accessing grammatical gender in German: The impact of gender-marking regularities. *Applied Psycholinguistics*, 27, 127–142. https://doi.org/10.1017/S0142716406060218

Jackendoff, R. (2002). *Foundations of language: Brain, meaning, grammar, evolution*. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780198270126.001.0001

Jackson, C. N. (2007). The use and non-use of semantic information, word order, and case markings during comprehension by L2 learners of German. *The Modern Language Journal*, 91(3), 418–432. https://doi.org/10.1111/j.1540-4781.2007.00588.x

Jaensch, C. (2009). Article choice and article omission in the L3 German of native speakers of Japanese with L2 English. In M. del P. García-Mayo & R. Hawkins (Eds), *Second language acquisition of articles: Empirical findings and theoretical implications* (pp. 233–263). John Benjamins. https://doi.org/10.1075/lald.49.14jae

Kempe, V., & MacWhinney, B. (1998). The acquisition of case marking by adult learners of Russian and German. *Studies in Second Language Acquisition*, 20(4), 543–587. https://doi.org/10.1017/S0272263198004045

Kempe, V., & MacWhinney, B. (1999). Processing of morphological and semantic cues in Russian and German. *Language and Cognitive Processes*, 14(2), 129–171. https://doi.org/10.1080/016909699386329

Köpcke, K.-M. (1982). *Untersuchungen zum Genussystem der deutschen Gegenwartssprache*. Tübingen: Niemeyer. https://doi.org/10.1515/9783111676562

Köpcke, K.-M., & Zubin, D. A. (1983). Die kognitive Organisation der Genuszuweisung zu den einsilbigen Nomen der deutschen Gegenwartssprache. *Zeitschrift für Germanistische Linguistik*, 11(2), 166–182. https://doi.org/10.1515/zfgl.1983.11.2.166

Köpcke, K.-M., & Zubin, D. (1984). Sechs Prinzipien für die Genuszuweisung im Deutschen: Ein Beitrag zur natürlichen Klassifikation. *Linguistische Berichte*, 93, 26–50.

Kraiss, A. (2014). Taking the awful out of the German language: A study of a new way to teach German gender and plural assignment. *Die Unterrichtspraxis/Teaching German*, 47(1), 23–41. https://doi.org/10.1111/tger.10153

Lemhöfer, K., Schriefers, H., & Hanique, I. (2010). Native language effects in learning second-language grammatical gender: A training study. *Acta Psychologica*, 135(2), 150–158. https://doi.org/10.1016/j.actpsy.2010.06.001

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

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

Levine, G. (1999). Gender assignment to German nonsense nouns: What does the native speaker know that the non-native speaker doesn't? In A. Greenhill, H. Littlefield, & C. Tano (Eds), *Proceedings of the 23rd Annual Boston University Conference on Language Development* (Vol. 2, pp. 397–406). Cascadilla Press.

MacWhinney, B. (1978). *The acquisition of morphophonology*. University of Chicago Press.

MacWhinney, B., Bates, E., & Kliegl, R. (1984). Cue validity and sentence interpretation in English, German, and Italian. *Journal of Verbal Learning and Verbal Behavior*, 23, 127–150. https://doi.org/10.1016/S0022-5371(84)90093-8

Menzel, B. (2003). Genuserwerb im DaF-Unterricht. *Deutsch als Fremdsprache*, 40(4), 233–237.

Menzel, B., & Tamaoka, K. (1995). Der? Die?? Das??? Genuszuweisung bei Anfängern: Zufall, Pauken oder Strategie. *Deutsch als Fremdsprache*, 1, 12–22.

Mills, A. E. (1986). *The acquisition of gender: A study of English and German*. Springer. https://doi.org/10.1007/978-3-642-71362-0

Pfaff, C. W. (1984). On input and residual L1 transfer effects in Turkish and Greek children's German. In R. W. Andersen (Ed.), *Second languages. A cross-linguistic perspective* (pp. 271–298). Newbury House.

Presson, N., MacWhinney, B., & Tokowicz, N. (2014). Learning grammatical gender: The use of rules by novice learners. *Applied Psycholinguistics*, 35(4), 709–737. https://doi.org/10.1017/S0142716412000550

Ritterbusch, R., LaFond, L., & Agustin, M. (2006). Learner difficulties with German case: Implications from an action research study. *Die Unterrichtspraxis/Teaching German*, 39, 30–45. https://doi.org/10.1111/j.1756-1221.2006.tb00014.x

Rogers, M. (1987). Learners' difficulties with grammatical gender in German as a foreign language. *Applied Linguistics*, 8(1), 48–74. https://doi.org/10.1093/applin/8.1.48

Rothman, J. (2015). Linguistic and cognitive motivations for the Typological Primacy Model (TPM) of third language (L3) transfer: Timing of acquisition and proficiency considered. *Bilingualism: Language and Cognition*, 18(2): 179–190. https://doi.org/10.1017/S136672891300059X

Salamoura, A., & Williams, J. N. (2007). The representation of grammatical gender in the bilingual lexicon: Evidence from Greek and German. *Bilingualism: Language and Cognition*, 10(3), 257–275. https://doi.org/10.1017/S1366728907003069

Thomoglou, P. (2007). Muttersprachеinfluss beim Genuserwerb: Beobachtungen an griechischen Lernern des Deutschen. *Deutsch als Fremdsprache*, 44(1), 20–25.

VanPatten, B., & Borst, S. (2012). The roles of explicit information and grammatical sensitivity in processing instruction: Nominative-accusative case marking and word order in German L2. *Foreign Language Annals*, 45(1), 92–109. https://doi.org/10.1111/j.1944-9720.2012.01169.x

Vigliocco, G., Antonini, T., & Garrett, M. F. (1997). Grammatical gender is on the tip of Italian tongues. *Psychological Science*, 8(4), 314–317. https://doi.org/10.1111/j.1467-9280.1997.tb00444.x

Walter, D., & MacWhinney, B. (2015). US German majors' knowledge of grammatical gender. *Die Unterrichtspraxis: Teaching German*, 48(1), 25–40. https://doi.org/10.1111/tger.10179

Walter, D. R., & van Compernolle, R. A. (2017). Teaching German declension as meaning: A concept-based approach. *Innovation in Language Learning and Teaching*, 11(1), 68–85. https://doi.org/10.1080/17501229.2015.1041961

Wegener, H. (1999). German gender in children's second language acquisition. In B. Unterbeck, M. Rissanen, T. Nevalainen, & M. Saari (Eds), *Gender in Grammar and Cognition* (pp. 511–544). Mouton de Gruyter.

Williams, S., & Hammarberg, B. (1998). Language switches in L3 production: Implications for a polyglot speaking model. *Applied Linguistics*, 19(3), 295–333. https://doi.org/10.1093/applin/19.3.295

Appendix. Number and percentage of incorrect and correct use of indefinite and definite articles by cue type

			Cue Type														Total
			No cue	Nat FEM	-e FEM	-ung FEM	E-er FEM	Nat MAS	1syll MAS	-er MAS	-el MAS	-ich MAS	E-en MAS	-en NEU	-o NEU	E-er NEU	Other	
Indefinite	Incorrect	n	83a, b	33b	88c, d	5a, b, d	10c	14a, e, f	39a, f	22a, b	3a, b, e, f	2a, b, e, f	1a, b, e, f	0e, f	2a, b, e, f	8e	3a, b, e, f	313
		%	25.2	32.7	46.1	33.3	71.4	16.1	19.3	27.8	13	11.8	9.1	0	15.4	9.6	17.6	26.1
	Correct	n	247a, b	68b	103c, d	10a, b, d	4c	73a, e, f	163a, f	57a, b	20a, b, e, f	15a, b, e, f	10a, b, e, f	14e, f	11a, b, e, f	75e	14a, b, e, f	884
		%	74.8	67.3	53.9	66.7	28.6	83.9	80.7	72.2	87	88.2	90.9	100	84.6	90.4	82.4	73.9
	All	n	330	101	191	15	14	87	202	79	23	17	11	14	13	83	17	1197
		%	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
Definite	Incorrect	n	19a		8b	0a, b, c, d	0a, b, c, d	0a, b, c, d	9a, b	4a, b	5d	1a, b	2a, b	1a, c, d	1a, b, c, d	36c, d	2a, b	88
		%	34.5		15.4	0	0	0	22	21.1	83.3	16.7	28.6	100	33.3	67.9	20	34.4
	Correct	n	36a		44b	1a, b, c, d	1a, b, c, d	1a, b, c, d	32a, b	15a, b	1d	5a, b	5a, b	0a, c, d	2a, b, c, d	17c, d	8a, b	168
		%	65.5		84.6	100	100	100	78	78.9	16.7	83.3	71.4	0	66.7	32.1	80	65.6
	All	n	55		52	1	1	1	41	19	6	6	7	1	3	53	10	256
		%	100		100	100	100	100	100	100	100	100	100	100	100	100	100	100

Note. Each subscript letter denotes a subset of Cue Type categories whose column proportions do not differ significantly from each other at the .05 level. For indefinite articles: $\chi^2(14) = 90.11$, $p < .001$, for definite articles: $\chi^2(13) = 50.77$, $p < .001$.

CHAPTER 7

The non-default gender category in additional-language French

Amanda Edmonds, Aarnes Gudmestad and Thomas Metzger
Université Paul-Valéry Montpellier 3 / Virginia Polytechnic Institute and State University / The Ohio State University

Whereas much research has examined the acquisition of grammatical gender in additional-language French, previous studies have all focused on targetlikeness. In other words, previous research reveals how accurate learners are in expressing gender and what factors may play a role in accuracy. Such research has found, for example, that learners of French tend to be less accurate in their expression of gender with feminine (as opposed to masculine) nouns and when the modifier used is an adjective (vs. a determiner). In the current chapter, we offer a complementary approach to the study of the acquisition of grammatical gender. We begin by establishing that masculine-marked modifiers are the default form in the corpus under study. Then, working within variationism, we seek to determine what factors influence the use of modifiers marked for feminine (the non-default) versus masculine (the default) gender, without focusing on accuracy. Our analysis is based on 5,992 instances of the expression of grammatical gender on determiners and modifiers in oral longitudinal data collected from 20 university learners of French. The data were modeled using a generalized linear mixed-effects model, which identified eight factors that influenced the use of the non-default gender category in oral additional-language French. The current study provides new insights into what impacts learners' choice between modifier forms and, specifically, into what leads to the selection of the non-default (i.e. feminine) gender forms.

Keywords: French, additional language, accuracy, modifier use, grammatical gender

Introduction

Within the field of second language acquisition (SLA), few linguistic features have inspired as much research as grammatical gender. Although the studies within this large literature are diverse, they share a tendency to consider the expression

https://doi.org/10.1075/sibil.63.07edm

of grammatical gender in an additional language (AL)[1] in terms of assessments of targetlike use. In other words, these studies show a preoccupation with how close learners may come to expressing gender in a way that respects prescriptive descriptions or native-speaker norms of the AL.[2] Previous research is thus often based on percentages of correct use of grammatical gender, and this body of research has identified a host of variables that appear to play a role in correctness. To name but one, numerous studies on AL French have found that learners' marking of gender is more accurate on determiners than on adjectives (Ayoun, 2007; Bartning, 2000; Dewaele & Véronique, 2001; Edmonds, Gudmestad, & Metzger, 2020; Granfeldt, 2005; Harley, 1979).

Instead of focusing on accuracy, in the current chapter we adopt a form-based approach informed by variationist SLA (e.g. Gudmestad, 2021) and show how it contributes novel insights into the understanding of the acquisition of gender. Specifically, we offer a reanalysis of data that we originally explored by examining targetlikeness (Edmonds et al., 2020). In our original study, we reported rates of targetlike use and fit a regression model to explain targetlike gender marking on modifiers (i.e. determiners and adjectives) in oral production data from 20 Anglophone learners of French collected before, during, and after an academic year in France (data were taken from the LANGSNAP corpus; Mitchell, Tracy-Ventura, & McManus, 2017). In the current reanalysis, we consider gender marking to constitute an example of interlanguage variation, and seek to identify the factors that may explain how the same 20 learners from the LANGSNAP corpus choose between feminine- and masculine-marked modifiers. More specifically, we focus our analysis on the non-default (feminine) gender category and identify the linguistic and extralinguistic factors that influence the choice of feminine over masculine modifier forms. This change in focus – from accuracy to the choice of modifier forms – allows us to contribute new knowledge concerning the acquisition of grammatical gender in AL French. We demonstrate how the two analyses of the same dataset complement each other and help to enrich the understanding of the complex nature of defaults in interlanguage more generally.

---

1. An AL is one that learners acquire after their language(s) of first socialization (see The Douglas Fir Group, 2016). One of our justifications for adopting this term, instead of the more widely used second language, is that its use avoids all confusion with actual order of acquisition (second language vs. third, fourth, etc. language). This is particularly relevant for multilingual learners, like those in this study: 18 of the 20 participants have learned at least one (and up to three) other ALs besides French (see also Dewaele, 2018).

2. In this chapter, the terms *targetlike, correct,* and *accurate* will be used as synonyms. This is because for the expression of gender, language use by native speakers and prescriptive rules about that use are largely consistent.

1. Background

We begin with a concise overview of the grammatical gender system in French. We then offer a review of the importance of a focus on errors in SLA in general terms before examining more specifically the notion of defaults, which plays an important role in the current analysis. We then review past studies that have investigated the acquisition of grammatical gender by AL learners of French, focusing our attention on those studies that examined how learners mark grammatical gender in oral and written production.

1.1 Grammatical gender in French

French has two gender categories – masculine and feminine –, with a greater proportion of nouns in French being masculine (61%, according to Surridge & Lessard, 1984, p. 46). While the vast majority of French nouns have a single gender, a small set – including, for example, *colocataire* 'roommate' and *propriétaire* 'owner' – can be used in the masculine or the feminine, depending on the biological gender of the referent. Such nouns illustrate the presence of biological gender in French, such that nouns that refer to female humans are feminine (e.g. *nièce* 'niece') and those that refer to males are masculine (e.g. *neveu* 'nephew'). All other nouns are generally held to show arbitrary gender, although certain scholars (e.g. Nelson, 2005) have suggested that a small set of other semantic characteristics may dictate gender assignment, insofar as nouns belonging to certain semantic classes, such as forms of precipitation (feminine) or colors (masculine), are associated with one gender.

For nouns whose gender is arbitrary, the shape of the noun – and, in particular, the ending of the noun – may offer valuable clues to its gender. There are, in French, a myriad of endings strongly associated with either masculine or feminine gender, although none of them has the canonical status of, for example, the *-o* (masculine) and *-a* (feminine) endings in Spanish and Italian. Whereas scholars such as Guiraud (1973) have noted that certain phonological forms are associated with one gender (i.e. consonant-final nouns tend to be feminine, vowel-final nouns masculine) and others have focused on the importance of derivational suffixes in predicting the gender of a noun (Surridge, 1989), most researchers have explored the relationship between grammatical gender and noun endings by examining the orthographic form of endings (see Ayoun, 2018 for an assessment of analyses concerning noun endings). This body of research has led to several lists of potentially predictive orthographic endings, of which we will offer one example. Lyster's (2006) account of predictability in gender attribution was based on an analysis of almost 10,000 common nouns contained in the *Robert Junior Illustré* dictionary. This particular

dictionary, whose target audience was French-speaking children between the ages of 8 and 12, was selected for analysis by Lyster because it was hypothesized to provide an appropriate target for AL learners, as it contains fewer rare and archaic words than one would find in a dictionary for adults. For the analysis, he sought to determine the proportion of nouns whose ending was predictive of its gender. To do so, Lyster classified each noun by its orthographic ending (meaning here the spelling corresponding to the final rhyme), and then he calculated the percentage of all nouns with a given ending that have the same gender. Any ending for which at least 89% of nouns had the same gender was considered predictive of grammatical gender, and Lyster found that "of all 9,649 nouns with only one gender, 7,760 or 80 percent have endings that reliably predict their gender" (p. 84). While it appears from this statement that the gender of the majority of French nouns is predictable on the basis of its ending, it must be kept in mind that Lyster classified as predictive a total of 273 different orthographic endings. Thus, while noun endings do offer clues to noun gender in French, the sheer number of different endings may obscure them.

So far, we have addressed the attribution of grammatical gender in French. We now turn to the marking of grammatical gender. The gender of a noun may be audible and/or visible via syntactic agreement on any modifier (determiner or adjective). We say *may be* audible and/or visible because many modifiers in French do not actually reveal the gender of the noun they modify. Ayoun (2010) offered a concrete characterization of this state of affairs with an analysis of a set of 5,016 determiner phrases from a magazine and newspaper corpus. She reported that only 50.24% of phrases in the corpus contained a modifier overtly marked for gender, highlighting the fact that input often does not offer an overt indication of a noun's grammatical gender. The relatively low level of overt gender marking in French is due to several characteristics of the language. First, numerous modifiers are invariable with respect to gender, meaning that the same explicit form is used with masculine and feminine nouns. This is particularly the case in the plural, as seen in the following examples where the determiner *des* 'some' and the adjective *sympathiques* 'nice' are both invariable: *des remarques sympathiques* 'some nice remarks$_{fem}$', *des commentaires sympathiques* 'some nice comments$_{masc}$'. Second, numerous modifiers show suppletive or elided forms with vowel-initial nouns, many of which are invariable (i.e. not overtly marked) for gender when used with vowel-initial nouns: *mon ami*-MASC 'my friend', *mon amie*-FEM 'my friend' and *l'ami*-MASC 'the friend', *l'amie*-FEM 'the friend'. Finally, there are more cases of overt gender marking in written than in oral French. This is because certain written forms that are distinct for gender – for example, *cet*-MASC versus *cette*-FEM 'this' and *têtu*-MASC versus *têtue*-FEM 'stubborn' – are pronounced identically, as in *cet ami têtu* and *cette amie têtue*, which are both pronounced [sɛ ta mi te ty]. Given these characteristics of

gender marking in French, Ayoun (2007, 2010) has noted that for AL learners, the oral and written input they receive with respect to grammatical gender is both complex and ambiguous (for a more thorough treatment of grammatical gender in French, see Ayoun, 2007, 2018).

1.2 Analyzing targetlikeness: Errors and defaults

Since its beginnings, the field of SLA has often focused on learner errors: describing them, analyzing their source(s), explaining them, offering remediation, and so on. In the 1960s and 1970s, researchers such as Corder and Selinker argued eloquently for the importance of learner errors in analyses of learner language. According to Corder (1967, p. 167), learner errors are significant for at least three reasons:

> First to the teacher, in that they tell him, if he undertakes a systematic analysis, how far towards the goal the learner has progressed and, consequently, what remains for him to learn. Second, they provide to the researcher evidence of how language is learned or acquired, what strategies or procedures the learner is employing in his discovery of the language. Thirdly (and in a sense this is their most important aspect) they are indispensable to the learner himself, because we can regard the making of errors as a device the learner uses in order to learn. It is a way the learner has of testing his hypotheses about the nature of the language he is learning.

In line with the second reason expressed by Corder, errors have long been considered by researchers from a variety of theoretical and methodological frameworks to constitute a window into both the learner's interlanguage system and the psycholinguistic mechanisms that underpin the acquisition of an AL. For example, certain researchers working within variationist SLA (e.g. Young, 1991) observed variation in the expression of morphology in interlanguage and sought to determine whether that variation is systematic. In so doing, such researchers identify the factors that favor one form (e.g. feminine modifiers) over another (e.g. masculine modifiers), allowing the analyst to characterize the interlanguage system that underlies a given case of variation (see Section 2.3 for more details on variationism).

Whereas variationist approaches to interlanguage variation focus on understanding the choice between two or more forms, most analyses of errors focus on accuracy. Approaching interlanguage through the prism of accuracy has received numerous criticisms. From a methodological point of view, Schachter (1974) empirically demonstrated that learners may avoid certain structures (and, thus, certain errors), potentially undermining comparisons of error profiles across groups of learners, whereas researchers such as Pica (1983) offered evidence of the difficulties involved in devising appropriate and transparent error measures. From a conceptual point of view, it has been suggested that analyses centered on errors lead researchers to focus on what learners cannot do in their AL, leaving to the side what

they are capable of (Ortega, 2017). Moreover, adopting a targetlikeness perspective forces analysts to identify the target against which to compare interlanguage and to make difficult (and often fundamentally ambiguous) decisions about the source of different errors (Flick, 1979).

Despite these challenges and criticisms, a focus on accuracy is very much alive in current SLA research, to the point that AL studies not referring to errors constitute an exception to the rule (see Ortega, 2016, for a discussion of the current situation). Such analyses take many forms, relying on notions like targetlike rates of use, accuracy, nativelikeness, or – the notion to which we now turn – defaults. Among research on the acquisition of inflectional morphology, the notion of default is relatively widespread. However, there are different understandings of the term. Whereas some SLA researchers consider default forms to be synonymous with unmarked forms, in the sense developed within general typological theories of markedness (Wiberg, 1996), others (especially those working within a Universal Grammar perspective) have used default to mean forms whose features (number, person, etc.) are un(der)specified, "which allows them to be inserted in particular positions without creating a mismatch with the feature requirements of the host position" (Prévost, 2002, p. 88). Yet other researchers use the notion of a default to describe a particular nontargetlike pattern seen in language acquisition data, and it is this definition that we will adopt in the current chapter. In this use of default, researchers analyzing, for example, grammatical gender (Sabourin, Stowe, & de Haan, 2006) and past-tense marking (Comajoan, 2005; Salaberry, 1999, 2003) have identified as the default a single form in a given inflectional paradigm that is overgeneralized by AL learners, thus leading to certain patterns of nontargetlike behavior.

Blom et al. (2016, p. 389) provide a clear description of this approach to defaultness: "To identify default forms across languages based on language acquisition data, it is important to determine which inflectional paradigmatic form has the highest accuracy and whether the same form is also more overused than the other forms." With respect to the acquisition of grammatical gender, the concept of default has been mentioned by numerous researchers. For example, one of the main findings from Sabourin et al.'s (2006, p. 25) study of transfer effects in the acquisition of gender in AL Dutch is that "learners make use of a default strategy in which they assign the more frequent gender (common gender) when they do not know the gender of a particular noun." The idea of a default gender category in AL French has also been addressed in previous research: "The finding that the foreign learners committed more gender mistakes in which they made feminine adjectives masculine rather than the other way around indicates that they tended to use the masculine form as a default" (Holmes & Dejean de la Bâtie, 1999, p. 500).

While observations about default use with respect to grammatical gender are not rare, we know of no previous research on French that has explored the factors that influence (non)-default form use.[3] Such an investigation is the goal of the current chapter.

1.3 Targetlikeness and the acquisition of grammatical gender

Table 1 summarizes how accurate learners of AL French have been found to be in expressing grammatical gender in oral or written production. In the cases of Ayoun (2007), Bartning (2000), Dewaele and Véronique (2001), and Granfeldt (2005), the figures presented are our own calculations. This is because these studies presented accuracy or error rates with respect to determiners and adjectives separately, whereas we have opted for providing an overall view of targetlikeness on modifiers (determiners and adjectives combined). Before commenting on these results, three caveats must be made. First, the majority of studies concerns English-speaking learners of French; it is thus unclear to what extent patterns may be generalized to other groups of learners. Second, with the exception of Ayoun (2007), all studies focused on learners at a relatively advanced level of proficiency, leaving open the question of what earlier stages of acquisition may look like. Finally, because of differences across the studies, it is problematic to attempt direct comparisons. With these caveats in mind, the findings in Table 1 demonstrate relatively high accuracy rates: Although Granfeldt (2005) found that one participant – Lisa – reached 66.7 percent accuracy during her third month of university study, most results were above 80 percent. Despite overall high accuracy rates, variability in gender marking was observed in these results. As concerns development, the four longitudinal studies showed a somewhat mixed picture: Granfeldt's two learners demonstrated different patterns, with Lisa's accuracy rate decreasing over a semester and Sama's increasing (to 100%) over the same period. The other longitudinal research examined change in gender marking over the course of a stay abroad, and all three studies (Edmonds & Gudmestad, 2018; Edmonds et al., 2020; Godfrey, Treacy, & Tarone, 2014) found an increase in accurate gender marking after a one- or two-semester stay in France. Similarly, results from cross-sectional studies were mixed, with Bartning (2000) showing clear development between pre-advanced and advanced levels, but Harley (1979) and Ayoun (2007) reporting patterns that did not indicate linear improvement. In addition to reporting rates of targetlikeness, researchers have sought explanations for the patterns seen in Table 1, insofar as they attempted to identify the factors that influence targetlikeness in production. We turn now to the findings from this research.

3. For research that examines the variable use of modifier gender in AL Spanish (rather than accuracy in grammatical gender marking), see Gudmestad, Edmonds, and Metzger (2021).

Table 1. Summary of grammatical gender accuracy results for production studies in AL French

Study	Modality	Participants	Longitudinal	Accuracy results
Edmonds and Gudmestad (2018)	Written	20 learners enrolled in a French degree program in the United Kingdom (UK)	Yes	Before a year in France: 82% At the end of the year abroad: 87.5% Eight months after return to the UK: 86.9%
Godfrey et al. (2014)	Written	4 American leaners who spent 1 semester in France / 4 American learners who remained in the US	Yes	Beginning of semester: 87.6% (stay abroad learners) / 90.7% (at-home learners) End of semester: 96.4% (stay abroad learners) / 91.2% (at-home learners)
Ayoun (2007)	Written	35 American learners at three levels of proficiency	No	Low-level learners: 93% Intermediate-level learners: 92.2% Advanced learners: 97.5%[a]
Granfeldt (2005)	Oral	2 tutored Swedish learners in their first semester of university study; both had studied French in secondary school (Lisa, Sama)[b]	Yes	Month 1: 78.4% (Lisa) 82.1% (Sama) Month 2: 76% (Lisa) / 85.5% (Sama) Month 3: 66.7% (Lisa) / 90.8% (Sama) Month 4: 71.6% (Lisa) / 100% (Sama)
Bartning (2000)	Oral	6 advanced and 9 pre-advanced Swedish learners	No	Pre-advanced learners: 74.1% Advanced learners: 88%
Harley (1979)	Oral	11 Grade 2 and 5 Grade 5 English-speaking children attending a French immersion school in Toronto	No	Grade 2: 83.1% Grade 5: 79.1%[c]
Edmonds et al. (2020)	Oral	20 learners enrolled in a French degree program in the UK	Yes	Before a year in France: 79.1% At the end of the year abroad: 85.6% Eight months after return to the UK: 87%
Dewaele (1994)	Oral	39 Flemish speakers enrolled in an advanced-level French course at a Belgian university	No	Debate: 80.1% Exam: 80.9% Interview: 85%

Table 1. (*continued*)

Study	Modality	Participants	Longi-tudinal	Accuracy results
Dewaele and Véronique (2001)[d]	Oral	27 Flemish learners at a pre-advanced or advanced level	No	Above 90%

Note.

a The total tokens identified included modifiers that were invariable for gender, which logically makes the rate of targetlikeness higher for this study.

b This study also reported on child bilinguals and untutored adults. We focused only on the tutored learners.

c We omitted the data Harley reported from children with at least one French-speaking parent.

d The exact accuracy rates could not be calculated because accuracy results were provided in bar graphs (the total number of errors is provided – 519 – but not the total number of tokens). This explains why the results are expressed differently from the other studies reviewed.

We have already mentioned that four of the studies presented in Table 1 reported separate accuracy rates for determiners and adjectives. This choice highlights the first important factor that appears to characterize patterns of targetlikeness: learners tend to be more accurate when marking gender on determiners than on adjectives. Eight other potentially important factors identified in previous research are presented in (1).

(1) a. General proficiency: Several studies in Table 1 (e.g. Bartning, 2000) demonstrated that learners at a higher level of overall proficiency made fewer errors in expressing grammatical gender.

b. Task: Dewaele (1994) found that Flemish university students showed different rates of gender errors as a function of task: the lowest level was found in informal discussions, whereas debates and narrative discourse led to higher levels.

c. First phoneme of the noun: Dewaele (2015) examined gender marking on the indefinite article with vowel- and consonant-initial nouns among 36 first-language Dutch students for whom French was either a second or a third language. Results showed that accuracy rates were lower with vowel-initial nouns, but only for the lower proficiency, third-language French speakers. See Edmonds et al. (2020) for further evidence of the importance of this factor.

d. Final phoneme of the noun: Of the three different factors we originally used to explore the impact of noun ending, only one was found to significantly predict targetlikeness: learners from the LANGSNAP corpus were less likely to mark gender in a targetlike manner when the noun ended with a consonant (Edmonds et al., 2020).

e. Noun class: We also found evidence of the significance of noun class, insofar as learners were significantly less likely to produce targetlike gender marking when a noun had arbitrary (versus biological) gender (Edmonds et al., 2020).
f. Distance: Several studies have reported that targetlike use on gender marking diminishes as the linear distance between the noun and the modifier increases (Edmonds & Gudmestad, 2018; Edmonds et al., 2020; Prodeau, 2005).
g. Noun frequency: Using an offline task, Surridge and Lessard (1984) reported that Canadian learners of French were more accurate in assigning gender to frequent nouns, where *frequent* referred to how often a noun was used by native French speakers. In our own research, we have found that noun frequency significantly predicts targetlikeness in oral production when frequency is defined with respect to individual usage. More specifically, we found that the more learners used a noun with a gender-marked modifier, the more they tended to be targetlike when expressing gender with that noun (Edmonds et al., 2020).
h. Noun gender: Several studies have reported that learners tend to be more targetlike when marking gender with masculine versus feminine nouns (Bartning, 2000; Dewaele & Véronique, 2001; Edmonds & Gudmestad, 2018; Edmonds et al., 2020; Harley, 1979).

This final observation concerning the potential role of noun gender indicates that, overall, learners tend to use more masculine modifier forms than feminine ones. As stated in the previous sub-section, this observation has led to the suggestion that AL learners of French treat the masculine as the default gender (Holmes & Dejean de la Bâtie, 1999). Because accuracy has dominated previous research, little is known about what factors may significantly encourage a learner to opt for the default option (i.e. a modifier marked for masculine) or may push the speaker to use the non-default form (a modifier marked for feminine). Thus, in the current analysis, we set out to answer the following three questions: (a) Are masculine modifier forms the default and feminine modifier forms the non-default in the corpus under study? (b) What factors predict the use of the non-default gender category in AL French, and (c) How might these factors change over the course of 21 months?

2. Method

2.1 Data and participants

We analyzed data from the publicly available LANGSNAP[4] corpus. This corpus comes from a longitudinal project on AL development conducted at the University of Southampton for which 27 Spanish majors and 29 French majors were followed over the course of 21 months, including an academic year spent in a target-language community. Participants met with researchers on six occasions: in the month of May prior to their year abroad, three times during their academic year abroad, and twice after their return to the United Kingdom. Each data collection point yielded both oral (semi-guided interview and oral narration) and written (argumentative essay) data. We limited the current project to a subset of these data.

We report on the data provided by 20 of the 29 French majors. Socio-biographic details for these 20 participants are provided in Table 2.

Table 2. Socio-biographic details for 20 LANGSNAP participants

Characteristic	Details for LANGSNAP corpus
Age	*M* = 20, *SD* = 0.38 (range: 19–21)
Gender	18 women, 2 men
First language	English = 18 English and French = 1 Finnish = 1
Number of years spent studying French	*M* = 10.45, *SD* = 2.3 (range: 6–20)
Initial proficiency (elicited imitation task)	*M* = 59.25, *SD* = 14.6 (range: 36–97 points)
Main activity during the year abroad	Language assistant = 12 Exchange student = 5 Workplace intern = 3
Other languages studied at university	Spanish = 3 German = 3 Chinese = 1 Italian = 1

We chose to analyze data from three of the six data collection points for this project, which allowed us to explore the full 21-month period covered by the corpus. More specifically, we analyzed (a) the session that took place before the students went abroad (henceforth pre-stay), (b) the session conducted one year later, at the end of their stay in France (henceforth in-stay), and (c) the final data collection session,

4. http://langsnap.soton.ac.uk/; the corpus is also available on Talkbank (https://slabank.talkbank.org/access/).

which occurred 8 months after the participants' return to the United Kingdom, and a total of 21 months after the first session (henceforth post-stay). We explore the evolution in the use of modifiers marked for feminine or masculine gender using data from the two oral tasks: semi-guided interviews and oral picture narrations (see Edmonds & Gudmestad, 2018 for an analysis of the written data). For the semi-guided interviews, participants interacted one-on-one with a researcher and discussed topics related to their stay in France, their future plans, etc. For the oral picture narrations, participants were provided with a picture book. They were given a few minutes to familiarize themselves with the story before retelling it in their own words.

2.2 Data coding

Using the transcripts provided on the LANGSNAP website, all instances of a noun modified by a determiner and/or an adjective were noted for coding. In cases where a noun was found to be accompanied by more than one modifier (e.g. one noun modified by a determiner and adjective or by multiple adjectives), each noun-modifier combination was considered a separate token. A total of 11,218 noun-modifier combinations were identified. Of these, approximately half involved modifiers with distinct masculine and feminine forms in the oral modality (k = 5,992). The remaining occurrences involved modifiers whose aural form is invariable for gender and, as such, were excluded. An example from the data analyzed is provided in (2). All modifiers are underlined in the extract in (2).

(2) *je voudrais euh euh vivre en France à l'avenir et je ne veux pas étudier les maths tout la vie euh oui et ainsi j'ai choisissai choisi à étudier le français* […] *mes mes* professeurs mes profs de français ne pensaient pas que j'ai des des bonnes capacités *français* (Participant 112, pre-stay, interview)

'I would like uh uh to live in France in the future and I don't want to study math all my life uh yes and thus I have chosed chosen to study French […] my my teachers my French teachers didn't think that I had some some good French abilities'

This extract contains 10 examples of a noun modified by a determiner and/or an adjective. Five of these examples involve invariable modifiers and were thus excluded from the analysis: *l'avenir, les maths, mes professeurs, mes profs de français, des capacités*. The remaining five occurrences were included in the analysis, and here we see three instances of a masculine modifier (*tout, le, français*) and two examples of a feminine modifier (*la, bonnes*).

Each noun-modifier combination was coded for the dependent variable and for 14 independent (linguistic and extralinguistic) variables based on previous research. We also included a random effect for participant, which allowed us to account for variability in the population of AL French speakers.

The binary dependent variable was the gender of the modifier: feminine or masculine. The first independent variable coded for task (interview vs. oral narration), whereas the second variable coded for time (pre-stay, in-stay, post-stay). The third variable identified the initial proficiency of the speaker (corresponding to the score on an elicited imitation test administered at pre-stay). Modifier type, the fourth independent variable, coded for whether the modifier was a determiner or an adjective. The fifth variable – syllable distance – reflected the relationship between the modifier and the noun by recording the number of syllables that separated the two.

The remaining nine independent variables coded for characteristics of nouns that may impact gender marking. The first of these concerned the gender of the noun, and each token was coded for whether the noun was masculine or feminine. The variable noun-initial phoneme coded for whether the noun began with a vowel or a consonant. Additionally, we chose to include three different variables to code for noun ending, reflecting the diversity of descriptions of this characteristic available in the literature. The first such variable – noun-final phoneme – coded for whether the noun ended with a vowel or a consonant. The second variable targeting noun ending – noun derivational morphology – showed whether the noun included a derivational ending linked to a given gender (according to the list provided by Surridge, 1989). The final noun-ending variable (noun-final orthographic rhyme) identified whether the ending of each noun was predictive or not of grammatical gender on the basis of Lyster's (2006) categorization of orthographic noun endings. The sixth noun-specific independent variable was noun class, for which we distinguished between nouns with biological and arbitrary gender. The seventh variable was semantic class, which coded for whether the noun belonged to a semantic class predictive of gender (following Nelson, 2005). The final two variables coded for different aspects of noun frequency. Noun log-frequency (language) coded for the frequency of each noun in modern-day French. For this variable, we used lemmatized frequency counts per million words taken from the film portion of the *Lexique 3.8* database (New et al., 2007). Each score was log-transformed to reduce skewness in the distribution. The variable noun frequency (individual) reflected how often each participant used a given noun with a modifier audibly marked for gender, as individual usage may impact the patterns of gender marking. For every participant, we thus calculated how often they used each noun at each data collection session and in each task. The list of 14 independent variables explored is provided in (3).

(3) Task
Time
Initial proficiency
Modifier type
Syllable distance
Noun gender
Noun-initial phoneme
Noun-final phoneme
Noun derivational morphology
Noun-final orthographic rhyme
Noun class
Semantic class
Noun log-frequency (language)
Noun frequency (individual)

2.3 Data analysis

We worked within the variationist framework, which recognizes that languages – including interlanguages – show systematic variation. Systematic (as opposed to random) variation refers to instances where two or more forms can be used in a single context to express the same meaning or function and where their usage patterns are impacted by an underlying probabilistic grammar. Variationist analyses attempt to explain cases of systematic variation by identifying linguistic and extralinguistic factors that significantly influence variable behavior in language (see Preston, 2000). To do so, such analyses generally rely on multivariate modeling. A hallmark of variationism is the recognition that multiple factors (both internal and external to a language system) work together to predict or explain why one form – for example, feminine-marked modifiers – may be more or less likely to occur in a given context. Within SLA, there exists a large body of research conducted within this perspective (see Bayley & Tarone, 2012), with most studies having focused on cases of sociolinguistic variation (i.e. variation that is instantiated in language use by native speakers, see Regan, Howard, & Lemée, 2009, for an example), and a small set of research looking at cases of interlanguage variation (e.g. Young, 1991). By focusing on the variation between feminine and masculine modifiers among learners of AL French, the current chapter contributes to variationist research on interlanguage variation.

The goal of our analysis is to explore variable gender marking in AL French by providing insight into factors predicting the use of the non-default gender category. Before conducting our multivariate analysis, we determined whether the masculine modifier form could be considered the default form in the LANGSNAP corpus. Although the ultimate focus of the current project is not on documenting or

explaining accuracy, but instead on explaining (non-)default form use, examining accuracy in order to demonstrate that the masculine forms are indeed the default in this corpus (and feminine forms, the non-default) is an indispensable step.

The identification of a possible default form was carried out following Blom et al.'s (2016) characterization of defaults in language acquisition data. More specifically, we examined whether the marking of gender involved: (a) more accurate use of one – the default – form in contexts where it is expected, and (b) the overextension of that same form into contexts where it is not targetlike. Following this verification, we carried out a generalized linear mixed-effects model to determine what factors significantly predict the use of modifiers marked for feminine (versus masculine) gender. This analysis consisted of four steps. We began by using bootstrapping to explore the independent variables. Variables that did not appear to significantly impact the use of gender-marked modifiers were removed at this stage. Second, we fit a generalized linear mixed-effects model using a backward selection strategy. Non-significant effects were removed, and potential interactions between each significant independent variable and the variable of time were explored. These interactions were of particular interest as they reveal whether the factors influencing the use of gender-marked modifiers changed over the 21-month period covered by the LANGSNAP corpus. For the third step, potential strong correlations (i.e. > 0.6) between retained fixed effects were explored. Had any such correlations been identified, one of the two factors in question would have been removed using previous research and the goals of the current project to make the decision. For the final step, McFadden's R^2 and the Bayesian Information Criterion were calculated. Both measures indicate the quality of the overall model. The analysis was conducted using R software (for the bootstrapping phase and the calculation of McFadden R^2) and SAS ® software[5] (for all other aspects of the analysis).

3. Results

We begin by examining whether the masculine form of modifiers may be considered the default form in this dataset. To do so, we first calculated the overall rate of targetlikeness in the corpus, which revealed that in 84.5% (5,036 out of 5,992) of cases, the gender of the modifier matched the gender of the noun. We then looked at masculine and feminine nouns separately and found evidence of the masculine being treated as a default (see Table 3).

5. The output, code, and analysis for this paper were generated using SAS software, Version 9.4 of the SAS System for Windows. Copyright © 2018 SAS Institute Inc. SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc., Cary, NC, USA.

Table 3. Accuracy as a function of noun and modifier gender

	Modifier gender			
	Masculine		Feminine	
Noun gender	***k***	**%**	***k***	**%**
Masculine	3094	90.7%	318	9.3%
Feminine	638	24.7%	1942	75.3%

Following Blom et al. (2016), this evidence is based on the higher accuracy rates with masculine (90.7%) versus feminine (75.3%) nouns, and the associated greater tendency to use masculine modifiers with feminine nouns (24.7%) as opposed to feminine modifiers with masculine nouns (9.3%). Taken together, these results indicate a greater tendency to overextend masculine modifier forms.

We conclude that these patterns are consistent with a characterization of masculine-marked modifiers as the default form in this corpus. For our analysis, the less frequent, non-default category of feminine was used as the reference category for the dependent variable, meaning that the model shows how the significant factors positively or negatively influenced the use of feminine-marked modifiers. However, the binary nature of the dependent variable means that information regarding the use of masculine modifiers is also contained in the model, as we will illustrate in our presentation of the results.

The first phase of the analysis (bootstrapping) led to the removal of one variable: noun-final phoneme. We then fit the generalized linear mixed-effects model, using the 13 remaining independent variables and the random effect for participants. Five variables were found to be non-significant and were thus removed from the model (task, time, noun derivational morpheme, initial proficiency, noun-log frequency [language]).

Although time was not significant as a main effect, we explored the potential interaction between time and the eight significant fixed effects in order to determine whether the influence of the significant factors showed change over time. Two significant interactions were observed: time × noun frequency (individual) and time × noun gender. However, because the interactions were each strongly correlated with the factor time, they were removed from the model. The final model is provided in Table 4 (see Appendix for details concerning the random effect).

In Table 4, the reference categories for the nominal independent variables are indicated in square brackets (except for the variables syllable distance and noun frequency [individual], which have no reference category because they are continuous). The directionality (positive or negative) of the parameter estimate indicates the impact of each fixed effect on the log-odds that a feminine modifier will be used.

Table 4. Generalized linear mixed-effects model

Factors	Estimate	*SE*	df	*p* value	Confidence intervals	
					Lower	Upper
(Intercept)	−3.3136	0.4038	19	< .0001	−4.188	−2.4684
Noun gender [masculine]						
feminine	3.5265	0.0824	19	< .0001	3.3541	3.6989
Modifier type [determiner]						
adjective	−1.1727	0.0967	19	< .0001	−1.3750	−0.9704
Syllable distance	−.1797	0.0455	5964	< .0001	−0.2689	−0.0906
Noun-final orthographic rhyme [predictive]						
not predictive	.1742	0.0826	19	.0484	0.0013	0.3471
Noun-initial phoneme [consonant]						
vowel	0.5887	0.1206	19	.0001	0.3363	0.8412
Noun class [grammatical gender]						
semantic gender	0.3608	0.1430	19	0.0207	0.0616	.6601
Semantic class [yes]						
no	1.2852	0.3999	19	0.0046	0.4483	2.1222
Noun frequency (individual)	−0.0692	0.0166	5964	< .0001	−0.1017	−0.0367

We begin by presenting the results in terms of the non-default, the feminine modifier. The positive parameter estimate for the first factor, noun gender, indicates that these learners were more likely to use a feminine-marked modifier with feminine nouns. For modifier type, the use of an adjective (as opposed to a determiner) resulted in a lower likelihood of a feminine modifier (visible in the negative parameter estimate). Syllable distance also shows a negative parameter estimate, which indicates that as the distance between the modifier and the noun increased, the log-odds of using a feminine modifier decreased. We found that nouns that end with a predictive orthographic rhyme were more likely to be used with a modifier marked for the feminine. These learners also showed sensitivity to the beginning of nouns in their selection of modifier form: nouns that begin with vowels were more likely to be accompanied by a feminine-marked modifier than those beginning with consonants. Nouns that reflect biological gender were more likely to be used with feminine-marked modifiers than those with arbitrary gender, whereas nouns that do not belong to a semantic class associated with a specific gender had higher log-odds of being used with a modifier marked for the feminine than with the masculine gender than those that belong to a semantic class that are connected to a particular gender. Finally, as individuals used a particular noun more frequently with a modifier overtly marked for gender, the log-odds of using a feminine modifier with that noun significantly decreased.

As mentioned previously, the results from this model can also be interpreted with respect to the use of the masculine (or the default) gender. This requires simply reversing the directionality associated with the parameter estimate. For example, where feminine modifiers were less likely to be used with adjectives (versus determiners), masculine modifiers showed the opposite pattern: as compared to feminine modifiers, masculine ones were more likely to be used with adjectives than with determiners. Overall, the results from Table 3 indicate that the log-odds of using masculine (versus feminine) modifiers were higher when the modifier was an adjective, as the distance between the noun and the modifier increased, and as individual speakers used each noun more often with a gender-marked modifier. They were, however, lower when the noun was feminine, when it ended in an orthographic rhyme not predictive of gender, when the noun began with a vowel, when the noun referred to a referent with biological gender, and when the noun did not belong to a semantic class associated with a specific gender.

The two metrics used to assess the quality of this model showed that the final model fit the data well. Starting with the Bayesian Information Criterion, this measure compares the log-likelihoods of the final model (Table 4) with a null model (which contains only the dependent variable). According to Kass and Raftery (1995), a difference of greater than 10 is strong evidence in favor of the model with a lower score. In this analysis, the generalized linear mixed-effects model has a Bayesian Information Criterion score of 4745.93, whereas the null model has the much higher score of 7950.033, indicating that the proposed model provides a better fit for the data. The second metric we used – the McFadden's R^2 – also confirmed the quality of the proposed model. The value obtained for this metric was $R^2_{McFadden} = 0.4061$, which indicates a strong model fit.

4. Discussion

Whereas previous research on the acquisition of grammatical gender in AL French has reported on how accurate learners are and on what factors may contribute to improved targetlikeness, the current study adopted a different and potentially complementary perspective. We focused on the choice learners make between modifier forms and, more specifically, between the non-default and default forms. Thus, our goal was to better understand what extralinguistic and linguistic factors predict the variable use of modifiers marked for feminine versus masculine gender and how these factors may change over the course of 21 months.

We begin by responding to our three research questions. In response to the first question, we demonstrated that the masculine form appears to be treated as the default in the dataset analyzed. The second research question asked which factors significantly predict the choice between modifier forms. Our generalized linear

mixed-effects model identified eight (of the 14 fixed effects investigated) as significant (see Table 4). The remaining six factors – time, task, initial proficiency, noun-final phoneme, noun derivational morphology, and noun-log frequency (language) – did not appear to be important for predicting the variable use of modifier gender in this dataset. Our third research question focused on potential change over time. Our analysis revealed that the impact of these eight factors was stable over the three data-collection periods investigated, as no significant interactions with time were retained in the final model. In other words, it appears that this participant group's use of the non-default modifier form did not change over the 21-month period investigated.

We turn now to what a focus on the choice between default and non-default gender-marked forms brings to a discussion that has been dominated by a focus on accuracy, before then reflecting on how this project informs discussions about default use in interlanguage more generally. Previous research has suggested that masculine modifier forms may constitute the default strategy in AL French based on the general observation that more errors in gender marking involve a feminine noun with a masculine modifier, as opposed to a masculine noun with a feminine modifier (Holmes & Dejean de la Bâtie, 1999). While not using the term default, other researchers, such as Dewaele & Véronique (2001, p. 290), have made similar observations: "gender errors most often involved the use of masculine gender instead of feminine." Although our own findings are in line with these observations, insofar as masculine modifier forms were overgeneralized, the results from our multivariate analysis suggest that referring to the masculine default as a general phenomenon is too simplistic. More specifically, our analysis demonstrates that (non-)default-form use is impacted by numerous factors. If a general tendency towards the use of masculine forms as defaults is indeed visible, the strength of the masculine default may be variable across different contexts of use. From the perspective of the learner, this means that the acquisition of grammatical gender in French involves, in part, overcoming the overuse of modifiers marked for the masculine, but that the extent of this overextension may vary as a function of different linguistic and extralinguistic factors. As we will demonstrate, the need to nuance the observation that the masculine is used as a default by learners of AL French can be further illustrated by bringing together the results from the current variationist analysis and previous (targetlike-oriented) research.

To do so, we draw on the results from our previous analysis of the same LANGSNAP dataset (see Edmonds et al., 2020). We focus on what insights can be derived from the six factors that were found to be significant in both the previous and current analyses: noun gender, noun-initial phoneme, noun frequency (individual), modifier type, syllable distance, and noun class. Table 5 offers a summary of the findings for these six variables.

Table 5. Summary of results for the six independent variables significant in both analyses

Factor	Edmonds et al. (2020)	Current study
Noun gender	Less likely to mark gender in a targetlike manner with feminine (versus masculine) nouns	More likely to use feminine-marked modifiers with feminine (versus masculine) nouns
Noun-initial phoneme	Less likely to mark gender in a targetlike manner when the noun began with a vowel (versus a consonant)	More likely to use feminine-marked modifiers when the noun began with a vowel (versus a consonant)
Noun frequency (individual)	Less likely to mark gender in a targetlike manner as the number of individual uses of a noun decreased	More likely to use feminine-marked modifiers as the number of individual uses of a noun decreased
Modifier type	Less likely to mark gender in a targetlike manner on adjectives (versus determiners)	Less likely to use feminine-marked modifiers with adjectives (versus determiners)
Syllable distance	Less likely to mark gender in a targetlike manner as the distance between the noun and modifier increased	Less likely to use feminine-marked modifiers as the distance between the noun and modifier increased
Noun class	Less likely to mark gender in a targetlike manner when the noun had arbitrary (versus biological) gender	Less likely to use feminine-marked modifiers when the noun had arbitrary (versus biological) gender

The first commonality between the two analyses concerns the factor of noun gender. The results revealed that the same category – feminine nouns – led both to a lower likelihood of targetlike use and to a higher likelihood of using feminine-marked modifiers. More specifically, these two results highlight the fact that although the masculine form of modifiers was overused in this dataset (resulting in lower rates of targetlikeness with feminine nouns), feminine modifiers remained more likely with feminine nouns. In other words, the overextension of masculine modifier forms appears relatively slight. This finding underscores the fact the notion of default, as defined by Blom et al. (2016), covers a wide variety of usage patterns, ranging from the use of a default form to the exclusion of all others to a slight overextension of the default category. We note that this possible range of patterns has been reported in previous SLA research on different inflectional paradigms. For example, whereas researchers such as Salaberry (1999) have reported evidence that beginner learners may initially rely on a strong default (i.e. perfective forms, to the exclusion of all other past-tense forms, to express past tense), the patterns uncovered in the current dataset suggest a weaker tendency to have recourse to the default masculine form. To the best of our knowledge, clear operationalizations of default strength within SLA research that focuses on language use have yet to be proposed. However, we

may hypothesize, for example, that strength of a default strategy changes along the developmental trajectory or may vary according to task (see Salaberry, 2003). Interestingly, in the current analysis, we saw no significant change in the use of feminine- versus masculine-marked modifiers over a 21-month period, including an academic year abroad. In other words, non-default patterns of use neither expanded nor evolved. Exploring whether and how default strategies change over the developmental trajectory is a promising direction for future research.

Returning to the results presented in Table 5, the findings with respect to the five remaining significant independent variables show two patterns. First, there are independent variables for which gender-marking behavior with one category results in lower likelihood of targetlikeness and a concomitant higher likelihood of using the non-default (feminine) form (noun-initial phoneme and noun frequency [individual]). Second, we note independent variables for which lower likelihood of targetlikeness goes hand-in-hand with lower likelihood of using feminine-marked modifiers in the same category (modifier type, syllable distance, noun class).

We suggest that these different patterns of convergence between the accuracy-based and form-based analyses reflect the fact that the masculine default is not uniformly in place across all contexts of use in a given interlanguage system. More specifically, when the same category of an independent variable leads to a lower likelihood of targetlikeness and a higher likelihood of feminine-modifier form use, this suggests that the use of a masculine default in these categories is not particularly strong (or perhaps not even present). This is because if a uniform masculine default were in operation, we would expect the use of feminine modifiers to be less likely in categories where likelihood of targetlikeness is lower, as the overextension of the default form (and not the greater likelihood of the non-default form) would be expected to be largely responsible for nontargetlike behavior. This is precisely the pattern seen for noun-initial phoneme and noun frequency (individual), suggesting that the likelihood of marking gender in a nontargetlike manner with vowel-initial nouns and with nouns used less frequently may not be due to a strong overgeneralization of the masculine default.

The second pattern, on the other hand, is consistent with what we would expect in the presence of a strong(er) masculine default. More specifically, there is a lower likelihood of targetlikeness with the same categories where the likelihood of feminine modifiers is low, presumably because feminine nouns paired with masculine (or default) modifiers are prevalent in these particular categories. This second pattern is seen for the variables modifier type, syllable distance, and noun class. Taken together, the concordance between the current study and past, accuracy-oriented research with respect to six independent linguistic factors allows us to nuance and reinforce previous findings with respect to the use of (non-)default forms in grammatical gender marking, both by questioning the strength of the default strategy

overall and by pinpointing more precisely those contexts in which the overuse of the masculine default may be most prevalent.

Whereas defaults for inflectional paradigms are often presented as an overall phenomenon, which can imply that the default form is preferred in all (or most) contexts of use, our analysis of feminine- versus masculine-marked modifiers in AL French clearly shows that this conception may be misleading. Although in this dataset accuracy in expressing masculine gender is higher overall and masculine modifier forms are used more often with feminine nouns than the reverse, the choice between the default (masculine) and non-default (feminine) modifier forms is more complex than a general, undifferentiated preference for the default. This complexity was revealed thanks to a multivariate analysis. We argue that such statistical analyses, which are common in variationist approaches, may be an important tool to uncovering the factors that explain (non-)default use in other inflectional paradigms.

5. Conclusion

To conclude, we return to the citation by Corder (1967, p. 167) provided in the literature review, where he sets forth three arguments in favor of dedicating attention to errors in interlanguage. Corder's second argument identifies errors as essential to the analyst, insofar as "they provide to the researcher evidence of how language is learned or acquired, [and] what strategies or procedures the learner is employing in his discovery of the language." Much previous SLA research has focused on learner errors by documenting and explaining accuracy. We hope that the current study has convincingly demonstrated how a variationist perspective on (non-)default forms can offer insight into interlanguage systems, while avoiding some of the challenges that have been identified for analyses focused on accuracy. Notably, our variationist analysis takes into consideration the full set of noun-modifier combinations produced (instead of focusing solely on errors) and offers new knowledge about the acquisition of gender. By focusing on the learner's choice between feminine- and masculine-marked modifier forms, we identified factors that significantly contributed to explaining the variable choice between the feminine (non-default) and masculine (default) forms. We moreover showed that accuracy-based and form-based analyses can fruitfully complement each other. In particular, the combination of these two types of approaches to learner language provided new insight into the observation that masculine forms tend to be used as defaults in the marking of gender in AL French. It will be interesting for future research to explore factors that predict (non-)default use across a wider span of the developmental trajectory than we examined in the current study. Indeed, most past research on gender marking in AL French has focused on relatively proficient

users of the language, leaving open the question of whether and to what extent (non-)default use characterizes gender marking at earlier stages.

As highlighted in our discussion section, the notion of default in interlanguage has been raised for various inflectional paradigms. We argue that the reflections developed in this chapter on two particular issues deserve additional research. First, we highlighted the fact that the notion of default has been used to refer to a range of situations, running the gambit from the exclusive use of a single form in a given inflectional paradigm to a slight overextension of a form. Providing further empirical evidence of the different configurations that respect Blom et al.'s (2016) characterization of defaults, in a variety of inflectional paradigms and involving various language combinations, would contribute to helping researchers to more clearly identify the variable strength of defaults in interlanguage. Second, the multivariate analysis presented in this chapter has importantly revealed that the learners studied did not appear to have recourse to the masculine default to the same extent throughout the interlanguage system and that, instead, the reliance on the default is influenced by several linguistic factors. This strongly suggests that, at least in this dataset, the notion of default is more complex than implied by previous characterizations. Future research that seeks both to verify this finding with respect to gender marking and to explore whether defaults in other inflectional paradigms show similar complexity should lead to better understanding of the nature of defaults in interlanguage systems.

References

Ayoun, D. (2007). The second language acquisition of grammatical gender and agreement. In D. Ayoun (Ed.), *French applied linguistics* (pp. 130–170). John Benjamins. https://doi.org/10.1075/lllt.16.10ayo

Ayoun, D. (2010). Corpus data: Shedding the light on French grammatical gender…or not. *Eurosla Yearbook*, 10, 119–141. https://doi.org/10.1075/eurosla.10.08ayo

Ayoun, D. (2018). Grammatical gender assignment in French: Dispelling the native speaker myth. *Journal of French Language Studies*, 28, 113–148. https://doi.org/10.1017/S095926951700014X

Bartning, I. (2000). Gender agreement in L2 French: Pre-advanced vs. advanced learners. *Studia Linguistica*, 54, 225–237. https://doi.org/10.1111/1467-9582.00062

Bayley, R., & Tarone, E. (2012). Variationist approaches to second language acquisition. In P. Robinson (Ed.), *Routledge encyclopedia of second language acquisition* (pp. 3–56). Routledge.

Blom, E., Chondrogianni, V., Marinis, T., & Vasic, N. (2016). The acquisition of verbal paradigms in Dutch and Greek L2 children: Cross-linguistic differences and inflectional defaults. *International Journal of Bilingualism*, 20(4), 386–402. https://doi.org/10.1177/1367006915609237

Comajoan, L. (2005). The early L2 acquisition of past morphology: Perfective morphology as an aspectual marker or default tense marker? In D. Eddington (Ed.), *Selected proceedings of the 6th Conference on the Acquisition of Spanish and Portuguese as First and Second Languages* (pp. 31–43). Cascadilla Proceedings Project.

Corder, S. P. (1967). The significance of learner's errors. *International Review of Applied Linguistics and Language Teaching*, 5(4), 161–170. https://doi.org/10.1515/iral.1967.5.1-4.161

Dewaele, J.-M. (1994). Variation synchronique des taux d'exactitude: Analyse de la fréquence d'erreurs morpholexicales dans trois styles oraux d'interlangue française. *International Review of Applied Linguistics and Language Teaching*, 17(4), 277–302.

Dewaele, J.-M. (2018). Why the dichotomy 'L1 versus LX user' is better than 'native versus non-native speaker'. *Applied Linguistics*, 39(2), 236–240.

Dewaele, J.-M. (2015). Gender errors in French interlanguage: The effects of initial consonant versus initial vowel of the head noun. *Arborescences: Revue d'Études Françaises*, 5, 7–27. https://doi.org/10.7202/1032661ar

Dewaele, J.-M., & Véronique, D. (2001). Gender assignment and gender agreement in advanced French interlanguage: A cross-sectional study. *Bilingualism: Language and Cognition*, 4, 275–297. https://doi.org/10.1017/S136672890100044X

The Douglas Fir Group. (2016). A transdisciplinary framework for SLA in a multilingual world. *The Modern Language Journal*, 100(Supplement 2016), 19–47. https://doi.org/10.1111/modl.12301

Edmonds, A., & Gudmestad, A. (2018). Gender marking in written L2 French: Before, during, and after residence abroad. *Study Abroad Research in Second Language Acquisition and International Education*, 3(1), 59–82. https://doi.org/10.1075/sar.16018.edm

Edmonds, A., Gudmestad, A., & Metzger, T. (2020). A longitudinal study of grammatical gender marking in French as an additional language. *Applied Linguistics*, 41(5), 733–755. https://doi.org/10.1093/applin/amz036

Flick, W. (1979). A multiple component approach to research in second language acquisition. In R. Andersen (Ed.), *The acquisition and use of Spanish and English as first and second languages* (pp. 59–74). TESOL.

Godfrey, L., Treacy, C., & Tarone, E. (2014). Change in French second language writing in study abroad and domestic contexts. *Foreign Language Annals*, 47(1), 48–65. https://doi.org/10.1111/flan.12072

Granfeldt, J. (2005). The development of gender attribution and gender concord in French: A comparison of bilingual first and second language learners. In J.-M. Dewaele (Ed.), *Focus on French as a foreign language: Multidisciplinary approaches* (pp. 164–190). Multilingual Matters. https://doi.org/10.21832/9781853597688-009

Gudmestad, A. (2021). Variationist approaches. In N. Tracy-Ventura & M. Paquot (Eds.), *The Routledge handbook of second language acquisition and corpora* (pp. 230–241). Routledge.

Gudmestad, A., Edmonds, A., & Metzger, T. (2021). Moving beyond the native-speaker bias in the analysis of variable gender marking. *Frontiers in Communication*, 6, 165.

Guiraud, P. (1973). *Le Français populaire*. P.U.F.

Harley, B. (1979). French gender 'rules' in the speech of English-dominant, French-dominant, and monolingual French speaking children. *Working Papers in Bilingualism*, 19, 129–156.

Holmes, V. M., & Dejean de la Bâtie, B. (1999). Assignment of grammatical gender by native speakers and foreign learners of French. *Applied Psycholinguistics*, 20, 479–506. https://doi.org/10.1017/S0142716499004026

Kass, R. E., & Raftery, A. E. (1995). Bayes factors. *Journal of the American Statistical Foundation*, 90(430), 773–795. https://doi.org/10.1080/01621459.1995.10476572

Lyster, R. (2006). Predictability in French gender attribution: A corpus analysis. *Journal of French Language Studies*, 16, 69–92. https://doi.org/10.1017/S0959269506002304

Mitchell, R., Tracy-Ventura, N., & McManus, K. (2017). *Anglophone students abroad: Identity, social relationships, and language learning*. Routledge. https://doi.org/10.4324/9781315194851

Nelson, D. (2005). French gender assignment revisited. *Word*, 56, 19–38. https://doi.org/10.1080/00437956.2005.11432551

New, B., Brysbaert, M., Veronis, J., & Pallier, C. (2007). The use of film subtitles to estimate word frequencies. *Applied Psycholinguistics*, 28, 661–677. https://doi.org/10.1017/S014271640707035X

Ortega, L. (2016). Multi-competence in second language acquisition: Inroads into the mainstream? In V. Cook & L. Wei (Eds.), *The Cambridge handbook of multi-competence* (pp. 50–76). Cambridge University Press. https://doi.org/10.1017/CBO9781107425965.003

Ortega, L. (2017). *The bi/multilingual turn in SLA: How far have we (not) come, and why?* Presented at the 36th Second Language Research Forum, The Ohio State University, Columbus, OH, 12–15 October.

Pica, T. (1983). Methods of morpheme quantification: Their effect on the interpretation of second language data. *Studies in Second Language Acquisition*, 6(1), 69–78. https://doi.org/10.1017/S0272263100000309

Preston, D. (2000). Three kinds of sociolinguistics and SLA: A psycholinguistic perspective. In B. Swierzbin, F. Morris, M. E. Anderson, C. A. Klee, & E. Tarone (Eds.), *Social and cognitive factors in SLA: Selected proceedings of the 1999 Second Language Research Forum* (pp. 3–30). Cascadilla Press.

Prévost, P. (2002). Truncation and missing inflection in initial child L2 German. *Studies in Second Language Acquisition*, 25, 65–97. https://doi.org/10.1017/S0272263103000032

Prodeau, M. (2005). Gender and number in French L2: Can we find out more about the constraints on production in L2? In J.-M. Dewaele (Ed.), *Focus on French as a foreign language*, (pp. 135–163). Multilingual Matters. https://doi.org/10.21832/9781853597688-008

Regan, V., Howard, M., & Lemée, I. (2009). *The acquisition of sociolinguistic competence in a study abroad context*. Multilingual Matters. https://doi.org/10.21832/9781847691583

Sabourin, L., Stowe, L. A., & de Haan, G. J. (2006). Transfer effects in learning a second language grammatical gender system. *Second Language Research*, 22(1), 1–29. https://doi.org/10.1191/0267658306sr259oa

Salaberry, R. (2003). Tense aspect in verbal morphology. *Hispania*, 86(3), 559–573. https://doi.org/10.2307/20062909

Salaberry, M. R. (1999). The development of past tense verbal morphology in classroom L2 Spanish. *Applied Linguistics*, 20(2), 151–178. https://doi.org/10.1093/applin/20.2.151

Schachter, J. (1974). An error in error analysis. *Language Learning*, 24, 205–214. https://doi.org/10.1111/j.1467-1770.1974.tb00502.x

Surridge, M. E. (1989). Le genre grammatical en Français fondamental: Données de base pour l'enseignement et l'apprentissage. *La Revue Canadienne des Langues Vivantes*, 45, 664–674. https://doi.org/10.3138/cmlr.45.4.664

Surridge, M. E., & Lessard, G. (1984). Pour une prise de conscience du genre grammatical. *La Revue Canadienne des Langues Vivantes*, 41, 44–52.

Wiberg, E. (1996). Reference to past events in bilingual Italian-Swedish children of school age. *Linguistics*, 34, 1087–1114. https://doi.org/10.1515/ling.1996.34.5.1087

Young, R. (1991). *Variation in interlanguage morphology*. Peter Lang.

Appendix. Details for the random effect for participants

Effect	Participant	Estimate	*SE*
Intercept	100	0.1169	0.1307
Intercept	101	−0.1776	0.1646
Intercept	102	0.4322	0.1374
Intercept	104	0.06808	0.1482
Intercept	105	0.1056	0.1509
Intercept	106	0.09017	0.1430
Intercept	107	−0.1183	0.1563
Intercept	108	0.2899	0.1413
Intercept	109	0.001955	0.1565
Intercept	110	−0.2238	0.1630
Intercept	111	−0.1726	0.1433
Intercept	112	0.1939	0.1605
Intercept	113	0.2960	0.1544
Intercept	114	−0.00688	0.1476
Intercept	115	−0.5682	0.1755
Intercept	116	0.007344	0.1617
Intercept	117	−0.2708	0.1768
Intercept	118	0.08075	0.1763
Intercept	119	−0.02979	0.1439
Intercept	120	−0.1112	0.1642

CHAPTER 8

Investigating grammatical gender agreement in Spanish

A methodological exploration of eye tracking

LeAnne Spino
University of Rhode Island

The present eye-tracking study tests the assumption that online techniques can tap implicit knowledge and reduce access to explicit knowledge by triangulating eye movements with a post-reading questionnaire. L1 English learners of L2 Spanish and Spanish native speakers (NSs) read sentences embedded with violations of determiner-noun and noun-adjective agreement followed by comprehension questions. They then completed a post-reading questionnaire that measured to what extent they were aware of the violations in the experimental stimuli. In the eye-tracking experiment, both L2 learners and Spanish NSs were sensitive to violations of determiner-noun agreement, but only Spanish NSs were sensitive to violations of noun-adjective agreement. On the post-reading questionnaire, most participants were also aware of the grammatical gender agreement violations, with determiner-noun agreement violations generally being more salient than noun-adjective violations. These results indicate that the violation detection paradigm implemented in this experiment did not entirely obviate explicit knowledge. Results are discussed in terms of methodological considerations researchers should keep in mind when designing and implementing studies that investigate grammatical gender agreement and other linguistic phenomena.

Keywords: online processing, eye tracking, implicit knowledge, explicit knowledge

1. Introduction

The advancement of knowledge on how grammatical gender agreement is processed and represented in the mind hinges on researchers' ability to tap first (L1) and second (L2) language learners' linguistic system. The linguistic system, however, can only be accessed in an indirect manner, an inconvenience that has spurred researchers' creativity when devising data collection methods. For instance, grammatical gender agreement has been studied with data gathered from language production

https://doi.org/10.1075/sibil.63.08spi

(e.g. Mariscal, 2009; Unsworth, 2008), a variety of offline comprehension tasks (e.g. López Prego & Gabriele, 2014; White et al., 2004), and online techniques such as eye tracking (e.g. Cholewa et al., 2019; Keating, 2009; Morales et al., 2016), self-paced reading (SPR, e.g. Renaud, 2014; Sagarra & Herschensohn, 2010, 2011), self-paced listening (SPL, e.g. Blom & Vasić, 2011; De Jong, 2005), ERPs (e.g. Alemán Bañon, Fiorentino, & Gabriele, 2012, 2018; Barber & Carreiras, 2005; Sabourin & Stowe, 2008) and fMRI (e.g. Miceli et al., 2002; Quiñones et al., 2018).

The data collection technique researchers ultimately choose likely depends on a host of variables such as participant characteristics, familiarity with and access to technology, and the specific research questions under investigation. For example, when investigating L1 acquisition in children, researchers generally opt for an evaluation of production data (e.g. Mariscal, 2009) and/or comprehension tasks that do not involve reading, such as the visual world paradigm (e.g. Lemmerth & Hopp, 2019) or SPL (e.g. Blom & Vasić, 2011). Researchers without access to expensive online technologies such as eye tracking, ERPs and fMRI may rely on more cost-effective data collection techniques, like SPR. So long as the data collection tool is appropriate to answer the research questions of interest, then a simpler tool is often all that is needed (for discussion see Mitchell, 2004).

While all data collection techniques can help shed light on the phenomena under investigation, they also inevitably filter that light, providing imprecise or at least incomplete measurements of the linguistic system. Such concerns can be attenuated through triangulation, or "the combination of methodologies in the study of the same phenomena, …[which helps to]… overcome the inherent weaknesses of single measurement instruments" (Denzin, 1989, p. 234). This methodological triangulation[1] can be within-method triangulation, where one method is employed, but the data is analyzed with various strategies. For example, researchers exploring the acquisition of grammatical gender agreement through production data may code for different types of agreement, such as determiner-noun and noun-adjective agreement (e.g. Bruhn de Garavito & White, 2002), and researchers using the eye-tracking technique generally analyze multiple reading measures (e.g. Keating, 2009). The danger in implementing only within-method triangulation is that "the flaws that arise from one method remain, no matter how many internal variations are devised" (Denzin, 1989, p. 244). Researchers therefore oftentimes engage in between-method triangulation, which combines different methodologies to explore a single phenomenon. For example, researchers may implement both production and comprehension measures

1. Methodological triangulation is but one form of triangulation, according to Denzin (1989). Other types of triangulation include data triangulation (examining multiple data sources together), investigator triangulation (bringing together multiple observers), and theory triangulation (entertaining various theoretical underpinnings simultaneously).

(e.g. Grüter, Lew-Williams, & Fernald, 2012; McCarthy, 2008; White et al., 2004) or multiple types of online comprehension measures (Foucart & Frenck-Mestre, 2012). Converging results from different methodologies may give researchers greater confidence in their findings. Results that do not converge should lead researchers to examine how or why the methodologies are capturing different aspects of the phenomena under investigation.

When attempting to answer any theoretical question, it is important to consider to what degree the methodology or methodologies that are being implemented are appropriate, and this "fit" should be empirically evaluated. The current study investigates how L2 learners process sentences embedded with grammatical gender agreement violations. To do so, I employ both within-method triangulation, by examining multiple types of grammatical gender agreement (i.e. determiner-noun and noun-adjective agreement) and multiple reading measures, as well as between-method triangulation, by pairing eye tracking with an additional data collection technique: a post-reading questionnaire.

2. Literature review

2.1 Eye tracking and grammatical gender agreement

Researchers implementing eye tracking to investigate matters related to grammatical gender agreement can either opt to implement the visual world paradigm (e.g. Dussias et al., 2013; Grüter et al., 2012; Hopp, 2012; Hopp & Lemmerth, 2018; Morales et al., 2016) or a text-based experimental design (e.g. Keating, 2009, 2010; Spinner, Gass, & Behney, 2013).

The visual world paradigm exploits the link that exists between auditory cues and eye movements by recording participants' eye movements while they look at a visual display and hear an auditory cue (for reviews see Godfroid, 2020; Huettig, Rommers, & Meyer, 2011). Studies that investigate grammatical gender agreement with the visual world paradigm take advantage of the anticipatory eye movements that occur when participants use morphosyntactic stimuli to predict linguistic input. This paradigm was first introduced to the field of second language acquisition (SLA) to study grammatical gender agreement by Lew-Williams & Fernald (2010).[2] In a series of three experiments, L1 English learners of L2 Spanish and Spanish NSs

2. Lew-Williams and Fernald (2010) employed the "looking-while-listening" procedure, which is similar to the visual world paradigm in terms of presentation of stimuli, but differs in that eye movements are manually coded offline instead of automatically recorded with an eye tracker (for a description of the procedure see Fernald et al. 2008).

viewed two objects on a computer screen that were of either the same or different grammatical gender. They then heard the prompt *Encuentra…* 'find…' or *¿Dónde está…?* 'Where is…?' followed by a definite article marked for gender and then the name of the object (e.g. *el pájaro* 'the-MASC bird-MASC). Spanish NSs looked more quickly at familiar and novel objects when the grammatical gender of the article was informative (i.e. when the objects were of two different genders) than when it was not informative (i.e. when the objects were of the same grammatical gender). The L2 Spanish learners, on the other hand, did not use the gender of the article to predict the object that would be named.

In a similarly designed experiment, Grüter et al. (2012) found that highly proficient L2 learners could indeed use gender as a predictive cue for familiar (not novel) nouns, providing limited evidence of L2 predictive ability. Proficiency seems to play a role in the ability of L2 learners to employ gender-based prediction, with highly proficient learners demonstrating evidence of prediction (e.g., Dussias et al., 2013; Hopp, 2012), but lower-proficiency learners not (e.g., Dussias et al., 2013; Hopp, 2012; Lew-Williams & Fernald, 2010). L1 background may also modulate the predictive use of grammatical gender (e.g., Dussias et al., 2013; Morales et al., 2016).

Text-based experimental designs are almost twice as common as visual world studies in the field of SLA (Godfroid, 2020, p. 64). To investigate the representation or processing of morphosyntactic phenomena, such as grammatical gender agreement, these text-based studies often utilize an anomaly or violation detection paradigm. In this experimental paradigm, participants' reading times on regions of sentences that contain violations are compared to matched, grammatical regions to determine sensitivity to violations. A longer reading time on regions with a violation is taken to be evidence of a processing cost. Keating (2009) implemented the violation detection paradigm with eye tracking to determine to what extent Spanish NSs and advanced, intermediate and beginning L2 Spanish learners (L1 = English) were sensitive to violations of noun-adjective grammatical gender agreement in three syntactic domains: in the DP, in the VP and in a subordinate clause. He found that beginning and intermediate students were not sensitive to violations in any condition, but the advanced learners were only sensitive to violations in the DP. He interpreted these results as indicating that advanced learners had acquired grammatical gender agreement, but that distance affected the detection of violations outside of the DP.

The violation detection paradigm has been used to investigate grammatical gender agreement not only with eye tracking, but also with other online techniques such as SPR (e.g. Renaud, 2014; Sagarra & Herschensohn, 2010, 2011) and ERPs (e.g. Barber & Carreiras, 2005). It has also been used to investigate a variety of other morphosyntactic phenomena such as subject-verb agreement (e.g. Jiang, 2004; Lim & Christianson, 2014) and tense-aspect agreement (e.g. Roberts & Liszka, 2013).

Unlike the visual world paradigm, the violation detection paradigm exposes participants to ungrammatical stimuli, which decreases the ecological validity of the experiment (Dussias et al., 2013; Godfroid, 2020). The presence of the violation may also increase participants' explicit awareness of the linguistic phenomena under investigation. Given that language learners' linguistic competence is, by definition, abstract and implicit, the efficacy of the methodology researchers select depends on its ability to circumvent the activation of explicitly learned knowledge during language processing (Keating & Jegerski, 2015). The current study therefore explores whether explicit knowledge is activated when learners are exposed to violations of grammatical gender agreement during an eye-tracking study.

2.2 Knowledge implemented in the violation detection paradigm

Online techniques are often deemed appropriate to tap linguistic competence because they are tacitly assumed to measure implicit or unconscious knowledge. For example, in a methodological review of 64 studies implementing 74 SPR tests, Marsden et al. (2018, p. 870) found "no challenges to the notion that SPRs in L2 research are a measure of implicit knowledge, and no discussion of a potential role for awareness or attention". They noted that although the field seems to have reached a consensus, this matter should still be validated empirically.

There have been cautionary warnings, though, about the type of knowledge elicited during online measures. For example, in a review of online methodologies employed in sentence processing research, Keating and Jegerski (2015) noted that "the simplest of design flaws, such as improper randomizations of test sentences or poorly designed comprehension questions, can unintentionally reveal the purpose of a study to participants or invite the use of explicit knowledge or strategic processing (i.e. unusually slow or careful reading), all of which can render the otherwise valuable moment-by-moment processing data uninformative".[3]

Godfroid and Winke (2015) echo this sentiment for the eye tracking in particular, urging that "eye-movement behavior must be interpreted with regard to the experimental treatment that elicited it (which could be more or less explicit)".

Within the violation detection paradigm, there are several methodological decisions that could increase participants' use of explicit knowledge, such as the number of experimental items, the randomization of stimuli, and the nature of distractor tasks (see Keating & Jegerski, 2015 for review). For example, Leeser, Brandl,

3. For other reviews of online processing methods see Conklin, Pellicer-Sánchez, & Carrol (2018); Dussias (2010); Godfroid (2020) Jegerski & VanPatten (2014); Marinis (2010), Roberts (2012, 2016), Roberts & Siyanova-Chanturia (2013), and Witzel, Witzel, & Forster (2012).

and Weissglass (2011) conducted a study that explored the effect post-stimulus distractor tasks could have on how participants process experimental sentences in an SPR experiment. Forty L2 Spanish learners completed two SPR tasks: one in which they answered a yes/no comprehension probe after each experimental sentence and another in which they assessed the grammaticality of each experimental sentence. The sentences in both tasks contained violations of noun-adjective agreement and subject-verb inversion in wh-questions. The participants were not sensitive to subject-verb inversion violations on either task; however, they were sensitive to grammatical gender agreement violations, but only when they were asked to make grammaticality judgments.

Leeser et al. (2011, p. 19) concluded that "if researchers are using on-line methods in order to investigate questions related to implicit L2 knowledge, incorporating a grammaticality judgement task into the experiment may lead to unwarranted conclusions regarding the learners' underlying competence, particularly if learners are likely to possess metalinguistic knowledge about the structure under investigation". For this reason, researchers often select post-stimulus distractor tasks that focus participants' attention on meaning instead of grammaticality, such as asking comprehension questions (e.g. VanPatten, Keating, & Leeser, 2012), making plausibility judgments (e.g. Wen et al., 2010), or judging whether a sentence expressed the same meaning as the experimental stimulus (e.g. Keating, 2009).

Carefully designing experimental stimuli is critical for curtailing the activation of explicit knowledge in this paradigm; however, there is no guarantee that it will be enough to ensure that explicit knowledge is not activated. For this reason, violation detection studies should be triangulated with other independent measures to determine to what extent explicit knowledge has been successfully obviated.

The current study explores the efficacy of one online methodology, eye tracking, in avoiding explicit knowledge by triangulating participants' reading times on sentences containing violations of determiner-noun and noun-adjective agreement with reports of their explicit awareness of the violations from a post-reading questionnaire. The research question explored in this study is whether increased reading times on grammatical violations are also accompanied by explicit awareness of the violations in the post-reading questionnaire. If only implicit knowledge is activated during the experiment (and if participants have a representation for grammatical gender agreement), then we would expect to see sensitivity to the violations in terms of reading times, but no explicit awareness of the violations on the post-reading questionnaire.

3. Method

3.1 Participants

Participants were L2 Spanish learners (n = 29) or Spanish NSs (n = 31). The L2 learners were all Spanish language majors pursuing a Bachelor's degree in Spanish. To be retained for analyses, L2 learners had to be close to graduating, which was operationalized as having three or fewer classes left to complete their Spanish degree.[4] L2 learners were on average 21.56 years old (SD = 1.23), had begun studying Spanish on average at 12.32 years old (SD = 3.11), and had studied Spanish for an average of 8.84 years (SD = 3.12). A total of 21 L2 learners had studied abroad for an average of 13.06 weeks (SD = 4.53). The Spanish NSs were all born in a Spanish-speaking country and immigrated to the United States at or after age 16. They immigrated to the United States at an average age of 22.41 (SD = 6.51) and had been in the United States for an average of 3.02 years (SD = 3.05). L2 learners and NSs were paid $20 for their participation.

Four L2 learners were eliminated from the analyses for reporting a native language other than English (n = 1), for studying another language with grammatical gender for more than two years (n = 1), or for not being close enough to degree completion (n = 2). Four Spanish NSs were eliminated for reporting English as their dominant language (n = 2), not having normal or corrected vision (n = 1), or not evidencing at least 80% comprehension of the experimental stimuli on the comprehension questions (n = 1). The final participant pool therefore consisted of 25 L2 learners and 27 Spanish NSs.

As an external measure of linguistic ability, all participants took a test from the grammar portion of the *Diploma de Español como Lengua Extranjera* for intermediate learners (Instituto Cervantes, 2008). This test is a 20-item fill-in-the-blank grammar test, with three multiple-choice options for each item. Participants were awarded 1 point for each correct response for a maximum possible score of 20. The L2 learners scored an average of 10.84 points (SD = 3.24), and the Spanish NSs scored an average of 18.30 points (SD = 1.24).

4. This operationalization ensured that the results of the study could shed light on the extent to which soon-to-be graduating Spanish majors were sensitive to grammatical gender agreement violations.

3.2 Eye-tracking materials

There was a total of 40 experimental items, 20 testing determiner-noun agreement and another 20 testing noun-adjective agreement. Experimental sentences were between 13 and 16 words in length. The same 20 nouns (10 masculine and 10 feminine) with transparent endings were used to test both determiner-noun and noun-adjective agreement. These nouns were piloted extensively to ensure participants would know them well (see Spino-Seijas, 2017 for details). The target nouns were all three syllables and between six and eight letters in length: *almuerzo* 'lunch', *dibujo* 'drawing', *ensayo* 'essay', *mercado* 'market', *momento* 'moment', *proyecto* 'project', *refresco* 'soft drink', *sombrero* 'hat', *trabajo* 'work', *zapato* 'shoe', *bebida* 'drink', *cámara* 'camera', *camisa* 'shirt', *comida* 'food', *escuela* 'school', *guitarra* 'guitar', *iglesia* 'church', *manzana* 'apple', *pregunta* 'question', and *ventana* 'window'.

In the determiner-noun condition, the nouns were paired with definite articles (e.g. *el almuerzo*, 'the lunch'). In the noun-adjective condition, the nouns were paired with adjectives with transparent /-o/ and /-a/ endings. The adjectives were all between four and seven letters in length: *barato* 'cheap', *bello* 'lovely', *blanco* 'white', *bonito* 'pretty', *bueno* 'good', *caro* 'expensive', *corto* 'short', *frío* 'cold', *largo* 'long', *limpio* 'clean', *lindo* 'beautiful', *malo* 'bad', *negro* 'black', *nuevo* 'new', *pequeño* 'small', *rojo* 'red', *rosado* 'pink', *sucio* 'dirty', *tonto* 'silly', and *viejo* 'old'.

Examples of the critical sentences are below:

(1) Determiner-noun
La madre prepara ***el/*la almuerzo*** *durante la mañana para su hija y su esposo.*
'The mother prepares **the-MASC/*the-FEM lunch-MASC** in the morning for her daughter and her husband.'

(2) Noun-adjective
El profesor lee el ensayo ***largo/*larga*** *cuando tiene tiempo libre por la noche.*
'The professor reads the-MASC essay-MASC **long-MASC/*long-FEM** when he has time at night.'

The critical regions for both conditions are bolded in the examples above.[5] For the determiner-noun condition, the critical region consisted of both the determiner and noun because readers often skip over short words (Brysbaert, Drieghe, & Vitu, 2005; Vitu et al., 1995), which can present special challenges for measuring fixations on determiners (see Spinner, Gass, & Behney, 2013). For the noun-adjective condition, the critical region was the adjective. Each critical region was followed by the words *durante* 'during' or *cuando* 'when'.

5. These regions are bolded for illustrative purposes here, but were not bolded in the experiment.

Each experimental sentence was comprised of a grammatical and ungrammatical doublet that was distributed across two different lists so that one participant would see the grammatical version of the doublet and another would see the ungrammatical version. The 40 experimental items were interspersed between 108 distractor items testing sensitivity to a variety of other violations.[6] Both lists were pseudo-randomized to ensure the target nouns did not appear within 15 sentences of each other to limit repetition effects (Reichle et al., 1998). The two lists were both then pseudo-randomized once more for a total of four lists. Each of the four lists contained the same number of target stimuli, grammatical and ungrammatical items and masculine and feminine tokens. Participants were randomly assigned to one of the four lists.

Each experimental sentence was followed by a comprehension question to ensure that participants attended to meaning while processing the experimental stimuli. An example of an experimental sentence with its corresponding comprehension question is presented in (3):

(3) *La madre prepara el almuerzo durante la mañana para su hija y su esposo.*
'The mother prepares the lunch in the morning for her daughter and her husband.'
¿La madre tiene hija?
'Does the mother have a daughter?'
A: *Sí* B: *No*
A: Yes B: No

The comprehension questions did not direct participants' attention to the portion of the sentence in which the grammatical violation occurred. Half of the comprehension questions required an affirmative answer. To be retained for analyses, participants had to respond correctly to 80% of the comprehension questions. L2 learners answered an average of 91.51% of the comprehension questions correctly (*SD* = 2.97) and Spanish NSs accurately responded to 91.62% of questions (*SD* = 3.46).

3.3 Vocabulary posttest

As is common in studies examining grammatical gender agreement (e.g. McCarthy, 2008; White et al. 2004), L2 learners were given a vocabulary posttest to ensure that they were familiar with the target nouns utilized in this study. For each of the 20

6. The distractors tested adverb placement, tense morphology, subject-verb agreement, adjective placement and N-Drop agreement.

nouns, L2 learners had to indicate whether the grammatical gender of the noun was masculine or feminine and then translate the word into English. The gender and translations of the target items were well-known to the L2 learners: they identified the correct gender for an average of 19.68 (*SD* = 0.48) of the 20 items, and were able to correctly translate an average of 19.84 (*SD* = 0.47). No participant assigned the incorrect gender to a noun more than once or provided an incorrect translation more than twice.

3.4 Post-reading questionnaire

Participants' awareness of the grammatical violations was measured with a post-reading questionnaire, administered immediately after the reading portion of the experiment. The questionnaire contained five questions of increasing specificity designed to gauge participants' awareness of the violations, and by extension, the likelihood that they used explicit knowledge when processing the experimental sentences. These questions are found in Table 5 in Section 4.3. below.

The second question asked whether participants had seen any grammatical errors while reading. If participants selected "no" for this question, the survey automatically skipped questions three through five, which inquired about the nature of those violations. Each of the five questions was presented on a different page of the survey and participants could not navigate backwards to ensure subsequent questions did not influence responses on previous questions.

3.5 Procedure and apparatus

Data collection transpired on an individual basis, and each data collection session lasted approximately 90 minutes. The researcher is bilingual in English and Spanish and spoke to participants in whichever language they preferred. Participants first completed the eye-tracking portion of the experiment. Participants' eye movements were recorded with an SR Research Ltd EyeLink 1000 eye tracker at 1000 Hz. To improve recording quality, participants placed their head on a chin and forehead rest while their eye movements were recorded. Participants were seated approximately 60 cm from the screen. Stimuli were presented on a 20-inch screen in size 18 Calibri font on a single line of text. The text appeared in black font on a white background. Participants began the reading portion of the experiment with five grammatical practice sentences, each followed by a comprehension question. They then read the 148 experimental sentences and answered a comprehension question after each one. The experimental sentences were divided as evenly as possible into 6 blocks. After each block, participants could remove their head from the tower mount and

take a break for as long as they desired to reduce fatigue. The length of break varied by participant and block from a few seconds to a few minutes, depending on the participant. A 9-point calibration was performed before participants commenced each block. Drift calibration was also performed before each experimental sentence. Participants pressed buttons on a hand-held controller to proceed through the experimental sentences and respond to comprehension questions.

After the eye-tracking experiment, participants immediately completed the post-reading questionnaire. They then completed the vocabulary posttest, a background questionnaire and the proficiency test. The background questionnaire consisted of a modified version of the Language Experience and Proficiency Questionnaire (LEAP-Q) (see Marian, Blumenfeld, & Kaushanskaya, 2007). All questionnaires and tests were completed through Survey Gizmo, an online survey tool.

3.6 Data analysis

Dependent variables were reading times on the critical regions. Four fixation times were measured: *first fixation duration* (the length of the duration of the first time a participant fixates on the area of interest), *first-pass time* (the sum of all fixations on the area of interest until the participant fixates outside of that area), *go-past time* (the sum of all fixations on and outside of the area of interest until exiting the region to the right) and *total time* (sum of all fixations on the area of interest). These measures were selected because of their prevalence in previous literature (for discussion see Godfroid, 2020).

Trials in which fixation times were under 80 ms were removed from the data set. Outliers were identified as any value that was over or under 2 standard deviations from the mean fixation time for each condition (determiner-noun or noun-adjective) and sentence type (grammatical or ungrammatical). These values were replaced with a value of two standard deviations from the mean.[7] Items were also excluded from analyses on an individual basis if the L2 learner could not correctly translate the target word on the vocabulary posttest or if they could not correctly identify the gender. This affected 0.89% and 1.43% of the dataset, respectively.

These data were analyzed with linear mixed-effects models (Baayen, Davidson, & Bates, 2008) in R with the lme4 package (Bates et al., 2015) and post hoc analyses were conducted with the lmerTest package (Kuznetsova, Brockhoff, & Christensen, 2017). Group (L2 learners, Spanish NSs) and grammaticality (grammatical,

7. An anonymous reviewer noted that trimming the data was an unnecessary step because ultimately a log transformation was applied to the data, which also helps attenuate outlier concerns. Another reason why data trimming may not have been necessary is that mixed-effects models are generally robust in the face of outliers (Keating & Jegerski, 2015).

ungrammatical) were set as fixed effects while subject and item were included as crossed random effects. Fixed effects were not contrast coded, and all models converged with the specified random effects structure. Separate analyses were run for each of the four fixation measures (first fixation duration, first-pass time, go-past time and total time). The determiner-noun and noun-adjective conditions were analyzed separately.

A visual inspection of the data indicated the reading times were not normally distributed, thus violating the assumptions of the model. Therefore, a log transformation was performed on the outcome variables. Model comparison tests were run between models with and without an interaction between group and grammaticality, and interactions are reported only when the model comparison test was statistically significant.

4. Results

4.1 Reading times on determiner-noun agreement

Table 1 presents the mean fixation times for the four reading measures (first fixation duration, first-pass time, go-past time and total time*)* for both the L2 learners and Spanish NSs on critical regions containing determiner-noun agreement.

Table 1. Mean (Standard Deviation) fixation times for determiner-noun condition

	First Fixation (ms)	First-Pass (ms)	Go-Past (ms)	Total (ms)
L2 Learners				
Grammatical	274 (106)	482 (284)	627 (413)	782 (454)
Ungrammatical	255 (100)	525 (285)	713 (464)	938 (508)
Difference	−19	43	87	156
Spanish NSs				
Grammatical	242 (99)	342 (187)	461 (308)	592 (385)
Ungrammatical	244 (116)	411 (257)	592 (429)	755 (431)
Difference	2	69	130	163

The results of the linear mixed-effects models for each of the four reading measures at the determiner-noun critical region are presented in Table 2; group (Spanish NSs', L2 learner) and grammaticality (grammatical, ungrammatical) are categorical variables. The reference category is Spanish NSs' reading times on grammatical sentences (Intercept).

Table 2. Model results for determiner-noun condition

Description	Predictor	β	SE	*t*	*p*
First Fixation	Intercept	5.43	0.03	179.64	< .001
	Group	0.08	0.04	2.04	.046
	Grammaticality	−0.02	0.02	−1.01	.311
First-Pass	Intercept	5.72	0.05	125.09	< .001
	Group	0.29	0.06	5.26	< .001
	Grammaticality	0.11	0.03	3.56	< .001
Go-Past	Intercept	5.98	0.05	113.82	< .001
	Group	0.27	0.06	4.24	< .001
	Grammaticality	0.17	0.04	4.90	< .001
Total Time	Intercept	6.22	0.05	114.63	< .001
	Group	0.27	0.07	3.83	< .001
	Grammaticality	0.22	0.03	6.38	< .001

At the critical region, first fixation durations showed a main effect of group ($t = 2.04$, $p = .046$), with the L2 learners reading more slowly than Spanish NSs, but no main effect of grammaticality ($t = -1.01$, $p = .311$). There were statistically significant main effects of group and grammaticality for first-pass (Group: $t = 5.26$, $p < .001$; Grammaticality: $t = 3.56$, $p < .001$), go-past (Group: $t = 4.24$, $p < .001$; Grammaticality: $t = 4.90$, $p < .001$) and total time (Group: $t = 3.83$, $p < .001$; Grammaticality: $t = 6.38$, $p < .001$). For these three measures, the L2 learners evidenced longer reading times than the Spanish NSs. Both groups also evidenced longer reading times on the ungrammatical regions. According to the models, first-pass showed a .11 increase in log fixation time on ungrammatical regions, go-past showed a .17 increase, and total time showed a .22 increase. There were no interactions between group and grammaticality for these three reading measures, indicating that both the L2 learners and the Spanish NSs slowed down while processing the violations.

4.2 Reading times on noun-adjective agreement

Table 3 presents the mean fixation times for the four reading measures (first fixation duration, first-pass time, go-past time and total time) for both the L2 learners and Spanish NSs on critical regions containing noun-adjective agreement.

Table 3. Mean (Standard Deviation) fixation times for noun-adjective condition

	First Fixation (ms)	First-Pass (ms)	Go-Past (ms)	Total (ms)
L2 Learners				
Grammatical	299 (115)	366 (144)	461 (294)	531 (289)
Ungrammatical	291 (107)	353 (184)	465 (329)	570 (359)
Difference	−8	−13	3	39
Spanish NSs				
Grammatical	275 (99)	301 (114)	372 (220)	425 (236)
Ungrammatical	290 (121)	331 (144)	457 (316)	592 (393)
Difference	14	31	86	167

The results of the linear mixed-effect models for each of the four reading measures at the noun-adjective critical region are presented in Table 4. Group (Spanish NS, L2 learner) and grammaticality (grammatical, ungrammatical) are categorical variables. The reference category is Spanish NSs' reading times on grammatical sentences (Intercept).

Table 4. Model results for noun-adjective condition

Description	Predictor	β	SE	*t*	*p*
First Fixation	Intercept	5.56	0.03	195.52	< .001
	Group	0.05	0.04	1.41	= .166
	Grammaticality	0.01	0.02	0.39	= .694
First-Pass	Intercept	5.62	0.04	141.79	< .001
	Group	0.20	0.05	4.34	< .001
	Grammaticality	0.09	0.04	2.78	< .001
	Group*Grammaticality	−0.16	0.05	−3.15	< .001
Go-Past	Intercept	5.78	0.05	117.85	< .001
	Group	0.21	0.06	3.40	< .001
	Grammaticality	0.17	0.05	3.77	< .001
	Group*Grammaticality	−0.20	0.07	−3.07	< .001
Total time	Intercept	5.91	0.06	95.77	< .001
	Group	0.23	0.07	3.40	= .001
	Grammaticality	0.29	0.04	6.62	< .001
	Group*Grammaticality	−0.26	0.06	−4.11	< .001

At the critical region, first fixation duration did not show a main effect of group ($t = 1.41$, $p = .166$) or grammaticality ($t = 0.39$, $p = .694$).

First-pass showed a main effect of group ($t = 4.34$, $p < .001$) and grammaticality ($t = 2.78$, $p < .001$), and a group*grammaticality interaction ($t = -3.15$,

$p < .001$). Spanish NSs evidenced .9 increase in log reading times on ungrammatical regions, and post hoc analyses revealed this difference was statistically significant ($\beta = -.10$, t = −2.78, $p = .029$). Post hoc analyses revealed that L2 learners' reading times were not statistically different on grammatical and ungrammatical sentences ($\beta = .06$, $t = 1.67$, $p = .343$).

Go-past showed a main effect of group ($t = 3.40$, $p < .001$) and grammaticality ($t = 3.77$, $p < .001$), and a group*grammaticality interaction ($t = -3.07$, $p < .001$). Spanish NSs evidenced .17 increase in log reading times in the ungrammatical condition and post hoc analyses revealed this difference was statistically significant ($\beta = -.17$, $t = -3.77$, $p = .001$). Post hoc analyses revealed that L2 learners' reading times were not statistically different on grammatical and ungrammatical sentences ($\beta = .03$, $t = 0.56$, $p = .943$).

Total time showed a main effect of group ($t = 3.40$, $p = .001$) and grammaticality ($t = 6.62$, $p < .001$), and a group*grammaticality interaction ($t = -4.11$, $p < .001$). Spanish NSs evidenced a .29 increase in log reading times in the ungrammatical condition, and post hoc analyses revealed this difference was statistically significant ($\beta = -.29$, $t = -6.62$, $p = < .001$). Post hoc analyses revealed that L2 learners' reading times were not statistically different on grammatical and ungrammatical sentences ($\beta = -.03$, $t = -0.70$, $p = .896$).

4.3 Post-reading questionnaire

The post-reading questionnaire was designed to measure participants' explicit knowledge of the violations of grammatical gender agreement in this study. The five questions as well as a summary of participants' responses are displayed in Table 5. The responses to each question are discussed below.

Table 5. Responses on the post-reading questionnaire

Questions		
Q 1: ***Did you notice anything strange about the sentences you read during the eye-tracking experiment? If so, what?***	**L2 Learners (*n* = 25)**	**Spanish NSs (*n* = 27)**
Percent reporting any grammatical violations (including but not limited to gender)	60%	96%
Percent mentioning gender violations specifically	40%	41%
Percent specifying determiner-noun violations	28%	19%
Percent specifying noun-adjective violations	0%	4%
Q 2. ***Where there any grammatical errors in the sentences you read during the eye-tracking experiment?***	**L2 Learners (*n* = 25)**	**Spanish NSs (*n* = 27)**
Percent reporting noticing errors	76%	100%

(*continued*)

Table 5. (*continued*)

Questions		
Q 3. ***What types of grammatical errors did you notice? Please list all the errors you remember and provide examples when possible.***	**L2 Learners (*n* = 25)**	**Spanish NSs (*n* = 27)**
Percent mentioning gender violations specifically	56%	59%
Percent specifying determiner-noun violations	48%	41%
Percent specifying noun-adjective violations	4%	7%
Q 4* ***Please check off all the types of errors you noticed in the sentences. If you are unsure as to what something is, please ask the researcher.***	**L2 Learners (*n* = 25)**	**Spanish NSs (*n* = 27)**
Percent who checked off noticing…		
determiner-noun violations	64%	93%
noun-adjective violations	52%	93%
incorrect adverb placement	12%	78%
incorrect adjective placement	32%	82%
tense violations	48%	78%
subject-verb agreement violations	40%	85%
noun-null nominal violations	44%	96%
#incorrect subjunctive use	8%	41%
#incorrect *por* and *para* use	12%	33%
#incorrect *ser* and *estar* use	20%	41%
Q 5.* ***What percentage of the experimental sentences (EXCLUDING comprehension questions) do you think contained grammatical errors?***	**L2 Learners**	**Spanish NSs**
Average percent of stimuli containing errors	29%	71%

Note. Participants who did not respond affirmatively to question 2 were not shown questions 3–5. Consequently, a total of 25 L2 learners answered questions 1 and 2, but only 19 L2 learners responded to questions 3–5. Percentages for questions 1 through 4 are calculated using the total number of L2 (*n* = 25) and Spanish NS (*n* = 27) participants.

The following options were provided: (a) incorrect gender agreement between articles (e.g. *el, la*) and nouns, (b.) incorrect gender agreement between nouns and adjectives, (c.) adverbs appeared in the wrong place in the sentence, (d.) adjectives appeared in the wrong place in the sentence, (e.) incorrect tense (present, past, etc.) was used, (f.) incorrect agreement between subjects and verbs, (g.) incorrect gender agreement between nouns and null nominals (e.g., *el rico, la cómica*), (h.) subjunctive was used incorrectly, (i.) *por* and *para* were used incorrectly, (j.) *ser* and *estar* were used incorrectly. Violations h through j (indicated with a pound sign) were not present in the stimuli and served to ensure participants were not overestimating the errors they saw in the experiment.

Question 1

Participants' responses to question 1 were coded to indicate whether they mentioned any type of grammatical violation, whether they noticed violations containing grammatical gender agreement, and if so, whether they specified the type of violation they saw (i.e. determiner-noun, noun-adjective). About 40% of L2 learners did not report noticing any grammatical violations. For example, one participant stated that there was "nothing too strange only that there was not very much variation in tense". The majority, however, did specify noticing some type of violation. Responses varied in specificity with regards to the type of error. Some responses were quite general, mentioning grammatical errors, but not specifying the type(s) of errors. For example, one Spanish NS stated, "There were fairly obvious grammatical errors that tripped me up at times". Another NS stated, "I noticed that a lot of the sentences were grammatically incorrect". Some responses to this question, however, were quite specific. For example, one L2 learner noted that "there were errors in the use of gender with the articles and nouns". Another stated: "There were problems with gender agreement between nouns and their articles. There were also some verbs that did not seem to make sense in the context of the sentence, but the questions for those sentences were not related specifically to the content, so I was able to answer them confidently". Both L2 learners and Spanish NSs reported noticing determiner-noun errors more often than noun-adjective errors. Their identification rates of both types of errors were also somewhat similar (28% vs. 19% for determiner-noun errors compared to 0% vs. 4% for noun-adjective errors).

Question 2

Question 1 was slightly ambiguous, as it only asked participants if they noticed anything "strange" while reading. Question 2, however, was more pointed and asked participants whether they noticed "grammatical errors". A total of 76% of L2 learners and 100% of Spanish NSs reported noticing grammatical errors in the stimuli.[8]

Question 3

Question 3 requested that participants provide a free recall of the errors they remembered noticing. A relatively equal percentage of L2 Spanish learners and Spanish NSs reported noticing grammatical gender agreement violations (56% and 59%, respectively). Determiner-noun violations were again more salient than noun-adjective violations for both L2 learners and Spanish NSs.

---

8. A reviewer noted that Question 2 may have had a strong "yes" bias. This is indeed potentially true. As participants only saw Questions 3–5 if they answered affirmatively to Question 2, this potential wording bias could artificially increase the number of participants responding to Questions 3–5.

Question 4

Question 4 asked participants to check off the errors they noticed from a list. Included in that list were errors that were indeed present in the stimuli, but also errors that were not included to ensure that participants were not overestimating the errors they saw. Most participants noticed grammatical gender violations. For the L2 learners, the determiner-noun agreement violations (64%) were slightly more salient than the noun-adjective violations (52%). For the Spanish NSs, both error types were equally salient (93%). Some L2 learners and Spanish NSs reported noticing errors that were not present in the stimuli (i.e., incorrect uses of the subjunctive, *por* and *para*, and *ser* and *estar*). The percentage of participants incorrectly reporting that they had noticed these errors were higher for the Spanish NSs than L2 learners.

Question 5

For Question 5, participants estimated the total percentage of stimuli that contained errors. In total, 50% of the stimuli contained grammatical errors. The L2 learners underestimated the number of errors (29%) and the Spanish NSs overestimated the number of errors (71%). The saliency of the grammatical errors was therefore larger for the Spanish NSs. This also may be why Spanish NSs reported noticing errors that were not present in the stimuli at a higher rate than L2 learners in Question 4.

5. Discussion

Online sentence processing experiments are often assumed to measure implicit knowledge and circumvent the activation of explicit knowledge. The purpose of this experiment was to determine whether this claim could be empirically validated. To that end, L2 learner and Spanish NSs' sensitivity to determiner-noun and noun-adjective violations of grammatical gender agreement was measured through eye movements and a post-reading questionnaire. While eye tracking can potentially measure both implicit and explicit knowledge, the post-reading questionnaire only measured explicit, verbalizable knowledge (for similar reasoning see Godfroid & Winke, 2015). Results from the eye-tracking experiment indicated that both L2 learners and Spanish NSs evidenced sensitivity to violations of determiner-noun agreement in their reading times, but only the Spanish NSs were sensitive to violations of noun-adjective agreement. Many L2 learners and Spanish NSs reported explicit awareness of the grammatical gender agreement violations on the post-reading questionnaire, especially violations of determiner-noun agreement. These findings indicate that the violation detection paradigm implemented in this study did not always circumvent the activation of explicit knowledge.

Sensitivity to determiner-noun agreement violations was evidenced by L2 learners and Spanish NSs on first-pass, go-past and total reading times, but not on first fixation reading times. Eye movement measures can be divided into early and late processing measures. First fixation duration and first-pass time are often considered to be early processing measures, related primarily to word identification. Total duration, by contrast, is considered a late processing measure, which can often illustrate processing difficulty (see Pickering et al., 2004). Go-past time can be considered both an early and late processing measure, because it combines both word integration and the time it takes to overcome processing difficulties to move forward in the sentence (Clifton, Staub, & Rayner, 2007). It is unsurprising that sensitivity was not evidenced during first fixation, as this reading time is thought to reflect lexical access and often does not show any effects of the phenomena under investigation within this paradigm (Godfroid, 2020). Notably, though, sensitivity to the determiner-noun violations occurred during both early (first-pass, go-past) and late (go-past and total time) processing measures.

The same was true for noun-adjective agreement violations, but only for the Spanish NSs. Therefore, when sensitivity occurred, it was robust and transpired during early and late processing measures. It may be tempting to attempt to link reading measures with different knowledge types; however, eye-movement measures have not been reliably mapped to cognitive events (Pickering et al., 2004; Rayner & Liversedge, 2011). That said, Godfroid and Winke (2015) note that late processing measures "seem to be the most likely places for conscious processing to surface in eye-movement behavior" (p. 331), especially when triangulated with an external measure to help confirm that an increase in fixation times signals a shift from unconscious to conscious processing. Therefore, it is possible that in the present study, the longer fixation times in the late processing measures are reflective of more explicit processing; however, this would benefit from empirical validation.

A better measure of explicit awareness can be found in participants' responses to the post-reading questionnaire where the majority of L2 learners and Spanish NSs indicated awareness of grammatical violations in the experiment stimuli. When engaging in free recall of errors, both L2 learners and Spanish NSs were more likely to report having seen determiner-noun than noun-adjective agreement violations. These results parallel the results found in production studies, where L2 learners are generally more accurate on determiner-noun than noun-adjective agreement (Bruhn de Garavito & White, 2002; Franceschina, 2001; White et al., 2004). When survey questions required participants to check off errors they had noticed from a list, determiner-noun violations were only slightly more salient for L2 learners and were equally salient for Spanish NSs.

While it seems that participants activated explicit knowledge, this does not necessarily mean that they did not use implicit knowledge as well when processing

the experimental stimuli. It is entirely possible that implicit knowledge gave rise to explicit knowledge during sentence processing. This is what likely occurred with the Spanish NSs, who have a robust implicit representation for grammatical gender agreement, yet still reported awareness of the violations. The same could also be true for the L2 Spanish learners.

It is an empirical question, however, whether the violation detection paradigm can be implemented in a way that does not trigger explicit knowledge. In all likelihood, a researcher's success in tapping only implicit knowledge within this paradigm depends on the design of the experiment (Keating & Jegerski, 2015). Researchers must make many methodological decisions and there is unfortunately relatively little empirical research to guide them. Possible experimental decisions that could impact to what extent explicit knowledge is activated in the violation detection paradigm include (but are not limited to) the selection of a poststimulus director task (see Leeser, Brandl, & Weissglass, 2011; Lim & Christianson, 2014), the randomization of experimental sentences, the number of experimental sentences, the proficiency of the participants, the context of acquisition, the ratio of critical to noncritical items, the number of fillers vs. distractors, the nature of the linguistic phenomena under investigation, and the online methodology selected. The last four factors listed will be explored in more detail below.

Best practices in stimuli design involve having relatively few critical stimuli compared to noncritical stimuli. This is because it is possible that having too many critical items could induce explicit awareness of the target stimuli. Determining this ratio, however, is a difficult balancing act: researchers need multiple stimuli testing the target structure(s), but do not want to make the experiment too long and induce fatigue. Keating and Jegerski (2015) recommend that researchers include at least 50% (but ideally 75%) noncritical items in their stimuli. For SPR, researchers generally tend to follow the more liberal recommendation (Marsden, Thompson & Plonsky, 2018). In this study, noncritical items represented 73% of experimental stimuli. It is possible that the outcomes would have been different with a different ratio of critical to noncritical items.

Whether the noncritical items are fillers or distractors may also matter. Keating and Jegerski (2015, p. 16) define distractors as "intentionally designed to contain a specific linguistic form or structure, either as critical items for another experiment or to counterbalance some characteristic of the critical stimuli that might otherwise make them stand out to the participant" and fillers as "unrelated sentences that are not intended to elicit any specific type of processing effects". Unfortunately, comparing ratios of distractors and fillers across studies is difficult because researchers often use the terms fillers and distractors interchangeably (Keating & Jegerski, 2015; Marsden et al., 2018). In the current study, all noncritical stimuli were distractors, and half of them were ungrammatical. This practice is not uncommon (e.g.,

Coughlin & Tremblay, 2013) and is often done so that the distractors can be used for another experiment (see Keating & Jegerski, 2015, p. 16; VanPatten, Keating, & Leeser, 2012, p. 121).

It is unclear whether the ungrammatical nature of the distractors makes the violations in the critical stimuli more or less salient: while it is possible that including violations in other structures heightens participants' explicit awareness that something is amiss and encourages them to look for errors, it is also possible that including no other violations serves to make the grammatical gender agreement violations even more salient when they occur. The ideal ratio of critical to noncritical items, as well as the influence of the nature of the noncritical items (distractors vs. fillers) still needs to be empirically validated.

The present study investigated explicit awareness of grammatical gender agreement, something that receives much metalinguistic discussion and error correction in L2 classrooms. It is possible that this type of linguistic phenomena may therefore be more prone to the activation of explicit awareness during sentence processing than phenomena that receive little to no explicit instruction, such as adverb placement. The results of the post-reading questionnaire could be interpreted to support this assertion, as L2 learners were more likely to identify grammatical gender agreement than adverb placement violations. The benefit of explicitly taught linguistic phenomena such as grammatical gender agreement, however, is that they are more easily identifiable on the post-reading questionnaire because participants generally have the metalinguistic knowledge necessary to discuss it. Researchers are therefore less likely to run the risk that participants are explicitly aware of the violation during reading, but unable to verbalize that awareness on the post-reading questionnaire.

Finally, the online technique researchers select could also impact to what degree participants access explicit knowledge during processing. For example, the violation detection paradigm may be less prone to activating explicit knowledge in non-cumulative SPR experiments because the transient nature of the stimuli does not allow for reanalysis (Jiang, 2004). SPL could also be a viable option because the nature of the stimuli is transient as well. This hypothesis, however, should be empirically validated through triangulating an SPR or SPL task with a post-reading questionnaire.

Unfortunately, it may be difficult to avoid activating explicit knowledge in the violation detection paradigm, as seeing or hearing a violation may trigger explicit knowledge. Researchers concerned about the activation of explicit knowledge may also opt for different paradigms that do not involve exposing participants to grammatical violations. This could be achieved by implementing a design that focuses on anticipatory processing, which has been demonstrated in the visual world paradigm. When implementing this paradigm, however, researchers still should

triangulate methodologies to verify how aware participants are of the linguistic phenomena under investigation. Andringa (2020), for example, triangulated the visual world paradigm with participant interviews to determine participants' awareness of grammatical rules governing the determiner system of an unknown miniature language. Andringa found that participants that did not develop awareness of the target stimuli were unable to use them as predictive cues.

6. Conclusion

The data from the current study suggest that the violation detection paradigm did not obviate activating explicit knowledge, as many researchers often tacitly assume. More research is needed, however, to determine whether this paradigm could avoid awakening explicit knowledge if the right methodological conditions were met. In the interim, researchers should either opt for a paradigm that does not expose participants to grammatical violations, or include an independent measure of knowledge, like the post-reading questionnaire implemented in the current study, to determine to what extent explicit knowledge was employed. The findings of this study therefore highlight the need for researchers to not only be forthcoming about their assumptions regarding the type(s) knowledge they are tapping, but also empirically test that assumption through methodological triangulation. Doing so will help the field continue to disentangle the types of knowledge implemented during sentence processing.

Acknowledgements

This work was funded by a *Language Learning* Dissertation grant and an internal grant from the Second Language Studies Program at Michigan State University.

References

Alemán Bañon, J., Fiorentino, R., & Gabriele, A. (2012). The processing of number and gender agreement in Spanish: An event-related potential investigation of the effects of structural distance. *Brain Research*, 1456, 49–63. https://doi.org/10.1016/j.brainres.2012.03.057

Alemán Bañon, J., Fiorentino, R., & Gabriele, A. (2018). Using event-related potentials to track morphosyntactic development in second language learners: The processing of number and gender agreement in Spanish. *PLoS One*, 13(7), 1–35. https://doi.org/10.1371/journal.pone.0200791

Andringa, S. (2020). The emergence of awareness in uninstructed L2 learning: A visual world eye tracking study. *Second Language Research*, 36(3), 335–357. https://doi.org/10.1177/0267658320915502

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

Barber, H., & Carreiras, M. (2005). Grammatical gender and number agreement in Spanish: An ERP comparison. *Journal of Cognitive Neuroscience*, 17(1), 137–153. https://doi.org/10.1162/0898929052880101

Bates, D. M., Mächler, M., Bolker, B. M., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67(1), 1–48. https://doi.org/10.18637/jss.v067.i01

Blom, E., & Vasić, N. (2011). The production and processing of determiner–noun agreement in child L2 Dutch. *Linguistic Approaches to Bilingualism*, 1(3), 265–290. https://doi.org/10.1075/lab.1.3.03blo

Bruhn de Garavito, J., & White, L. (2002). The second language acquisition of Spanish DPs: The status of grammatical features. In A. T. Pérez-Leroux & J. Muñoz Liceras (Eds), *The acquisition of Spanish morphosyntax: The L1/L2 connection* (pp. 153–178). Kluwer. https://doi.org/10.1007/978-94-010-0291-2_6

Brysbaert, M., Drieghe, D., & Vitu, F. (2005). Word skipping: Implications for theories of eye movement control in reading. In G. Underwood (Ed.), *Cognitive Processes in Eye Guidance* (pp. 53–77). Oxford University Press. https://doi.org/10.1093/acprof:oso/9780198566816.003.0003

Cholewa, J., Neitzel, I., Bürsgens, A., & Günther, T. (2019). Online-processing of grammatical gender in noun-phrase decoding: An eye-tracking study with monolingual German 3rd and 4th Graders. *Frontiers in Psychology*, 10, 1–16. https://doi.org/10.3389/fpsyg.2019.02586

Clifton, C., Staub, A., & Rayner, K. (2007). Eye movements in reading words and sentences. In R. P. G. van Gompel, M. H. Fischer, W. S. Murray, & R. L. Hill (Eds.), *Eye movements: A window on the mind and brain* (pp. 341–371). Elsevier. https://doi.org/10.1016/B978-008044980-7/50017-3

Conklin, K., Pellicer-Sánchez, A., & Carrol, G. (2018). *Eye-tracking: A guide for applied linguistics research*. Cambridge University Press. https://doi.org/10.1017/9781108233279

Coughlin, C. E., & Tremblay, A. (2013). Proficiency and working memory based explanations for nonnative speakers' sensitivity to agreement in sentence processing. *Applied Psycholinguistics*, 34(3), 615–646. https://doi.org/10.1017/S0142716411000890

De Jong, N. (2005). Can second language grammar de learned through listening? *Studies in Second Language Acquisition*, 27, 205–234. https://doi.org/10.1017/S0272263105050114

Denzin, N. K. (1989). *The research act: A theoretical introduction to sociological methods* (3rd ed.). Prentice Hall.

Dussias, P. E. (2010). Uses of eye-tracking data in second language sentence processing research. *Annual Review of Applied Linguistics*, 30, 149–166. https://doi.org/10.1017/S026719051000005X

Dussias, P. E., Valdés Kroff, J. R., Guzzardo Tamargo, R. E., & Gerfen, C. (2013). When gender and looking go hand in hand: Grammatical gender processing in L2 Spanish. *Studies in Second Language Acquisition*, 35, 353–387. https://doi.org/10.1017/S0272263112000915

Fernald, A., Zangl, R., Portillo, A. L., & Marchman, V. A. (2008). Looking while listening: using eye movements to monitor spoken language comprehension by infants and young children. In I. Sekerina, E. M. Fernández, & H. Clahsen (Eds.), *Developmental psycholinguistics: Online methods in children's language processing* (pp. 97–135). John Benjamins. https://doi.org/10.1075/lald.44.06fer

Foucart, A., & Frenck-Mestre, C. (2012). Can late L2 learners acquire new grammatical features? Evidence from ERPs and eye-tracking. *Journal of Memory and Language*, 66(1), 226–248. https://doi.org/10.1016/j.jml.2011.07.007

Franceschina, F. (2001). Morphological or syntactic deficits in near-native speakers? An assessment of some current proposals. *Second Language Research*, 17(3), 213–247. https://doi.org/10.1177/026765830101700301

Godfroid, A. (2020). *Eye tracking in second language acquisition and bilingualism: A research synthesis and methodological guide*. Routledge.

Godfroid, A., & Winke, P. (2015). Investigating implicit and explicit processing using L2 learners' eye-movement data. In P. Rebuschat (Ed.), *Implicit and explicit learning of languages*, (pp. 325–348). John Benjamins. https://doi.org/10.1075/sibil.48.14god

Godfroid, A., Winke, P., & Conklin, K. (2020). Exploring the depths of second language processing with eye tracking: An introduction. *Second Language Research*, 36(3), 243–255. https://doi.org/10.1177/0267658320922578

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

Hopp, H. (2012). Grammatical gender in adult L2 acquisition: Relations between lexical and syntactic variability. *Second Language Research*, 29(1), 33–56. https://doi.org/10.1177/0267658312461803

Hopp, H., & Lemmerth, N. (2018). Lexical and syntactic congruency in L2 predictive gender processing. *Studies in Second Language Acquisition*, 40(1), 171–199. https://doi.org/10.1017/S0272263116000437

Huettig, F., Rommers, J., & Meyer, A. S. (2011). Using the visual world paradigm to study language processing: A review and critical evaluation. *Acta Psychologica*, 137(2), 151–171. https://doi.org/10.1016/j.actpsy.2010.11.003

Instituto Cervantes. (2008). *Diploma de español. Nivel intermedio*. Retrieved on 10 June from http://www.dele.org/uploads/test/intermedio/textos_orales.pdf

Jegerski, J., & VanPatten, B. (2014). *Research methods in second language psycholinguistics*. Routledge.

Jiang, N. (2004). Morphological insensitivity in second language processing. *Applied Psycholinguistics*, 25, 603–634. https://doi.org/10.1017/S0142716404001298

Keating, G. D. (2009). Sensitivity to violations of gender in native and nonnative Spanish: An eye-movement investigation. *Language Learning*, 59(3), 503–535. https://doi.org/10.1111/j.1467-9922.2009.00516.x

Keating, G. D. (2010). The effects of linear distance and working memory on the processing of gender agreement in Spanish. In B. VanPatten & J. Jegerski (Eds), *Research in second language processing and parsing* (pp. 113–134). John Benjamins. https://doi.org/10.1075/lald.53.05kea

Keating, G. D., & Jegerski, J. (2015). Experimental designs in sentence processing research: A methodological review and user's guide. *Studies in Second Language Acquisition*, 37(1), 1–32. https://doi.org/10.1017/S0272263114000187

Kuznetsova, A., Brockhoff, P., & Christensen, R. (2017). lmerTest Package: Tests in linear mixed effects models. *Journal of Statistical Software*, 82, 1–26. https://doi.org/10.18637/jss.v082.i13

Leeser, M. J., Brandl, A., & Weissglass, C. (2011). Task effects in second language sentence processing research. In P. Trofimovich & K. McDonough (Eds.), *Applying priming methods to L2 learning, teaching and research* (pp. 179–198). John Benjamins. https://doi.org/10.1075/lllt.30.13lee

Lemmerth, N., & Hopp, H. (2019). Gender processing in simultaneous and successive bilingual children: Cross-linguistic lexical and syntactic influences. *Language Acquisition*, 26(1), 21–45. https://doi.org/10.1080/10489223.2017.1391815

Lew-Williams, C., & Fernald, A. (2010). Real-time processing of gender-marked articles by native and non-native Spanish speakers. *Journal of Memory and Language*, 63, 447–464. https://doi.org/10.1016/j.jml.2010.07.003

Lim, J. H., & Christianson, K. (2014). Second language sensitivity to agreement errors: Evidence from eye movements during comprehension and translation. *Applied Psycholinguistics*, 36(6), 1–33. https://doi.org/10.1017/s0142716414000290

López Prego, B., & Gabriele, A. (2014). Examining the impact of task demands on morphological variability in native and non-native Spanish. *Linguistic Approaches to Bilingualism*, 4(2), 192–221. https://doi.org/10.1075/lab.4.2.03lop

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. https://doi.org/10.1044/1092-4388(2007/067)

Marinis, T. (2010). Using on-line processing methods in language acquisition research. In E. Blom & S. Unsworth (Eds.), *Experimental methods in language acquisition research* (pp. 139–162). John Benjamins. https://doi.org/10.1075/lllt.27.09mar

Mariscal, S. (2009). Early acquisition of gender agreement in the Spanish noun phrase: Starting small. *Journal of Child Language*, 36, 143–171. https://doi.org/10.1017/S0305000908008908

Marsden, E., Thompson, S., & Plonsky, L. (2018). A methodological synthesis of self-paced reading in second language research. *Applied Psycholinguistics*, 39(5), 861–904. https://doi.org/10.1017/S0142716418000036

McCarthy, C. (2008). Morphological variability in the comprehension of agreement: An argument for representation over computation. *Second Language Research*, 24(4), 459–486. https://doi.org/10.1177/0267658308095737

Miceli, G., Turriziani, P., Caltagirone, C., Capasso, R., Tomaiuolo, F., & Caramazza, A. (2002). The neural correlates of grammatical gender: An fMRI investigation. *Journal of Cognitive Neuroscience*, 14(3), 618–628. https://doi.org/10.1162/08989290260045855

Mitchell, D. C. (2004). On-line methods in language processing: Introduction and historical review. In M. Carreiras & C. Clifton (Eds.), *The online study of sentence comprehension: Eyetracking, ERPs and beyond* (pp. 15–32). Psychology Press.

Morales, L., Paolieri, D., Dussias, P. E., Valdes Kroff, J. R., Gerfen, C., & Bajo, M. T. (2016). The gender congruency effect during bilingual spoken-word recognition. *Biling (Camb Engl)*, 19(2), 294–310. https://doi.org/10.1017/S1366728915000176

Pickering, M. J., Frisson, S., McElree, B., & Traxler, M. J. (2004). Eye movements and semantic composition. In M. Carreiras & C. Clifton (Eds.), *The online study of sentence comprehension: Eyetracking, ERPs and beyond* (pp. 33–50). Psychology Press.

Quiñones, I., Molinaro, N., Mancini, S., Hernández-Cabrera, J. A., Barber, H., & Carreiras, M. (2018). Tracing the interplay between syntactic and lexical features: fMRI evidence from agreement comprehension. *Neuroimage*, 175, 259–271. https://doi.org/10.1016/j.neuroimage.2018.03.069

Rayner, K., & Liversedge, S. P. (2011). Linguistic and cognitive influences on eye movements during reading. In S. P. Liversedge, I. D. Gilchrist, & S. Everling (Eds.), *The Oxford handbook of eye movements* (pp. 751–766). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199539789.013.0041

Reichle, E. D., Pollatsek, A., Fisher, D. L., & Rayner, K. (1998). Toward a model of eye movement control in reading. *Psychological Review*, 105(1), 125–157. https://doi.org/10.1037/0033-295X.105.1.125

Renaud, C. (2014). A processing investigation of the accessibility of the uninterpretable gender feature in L2 French and L2 Spanish adjective agreement. *Linguistic Approaches to Bilingualism*, 4(2), 222–255. https://doi.org/10.1075/lab.4.2.04ren

Roberts, L. (2012). Psycholinguistic techniques and resources in second language acquisition research. *Second Language Research*, 28(1), 113–127. https://doi.org/10.1177/0267658311418416

Roberts, L. (2016). Self-paced reading and L2 grammatical processing. In A. Mackey & E. Marsden (Eds.), *Advancing methodology and practice: The IRIS repository of instruments for research into second languages* (pp. 58–73). Routledge.

Roberts, L., & Liszka, S. A. (2013). Processing tense/aspect-agreement violations on-line in the second language: A self-paced reading study with French and German L2 learners of English. *Second Language Research*, 29(4), 413–439. https://doi.org/10.1177/0267658313503171

Roberts, L., & Siyanova-Chanturia, A. (2013). Using eye-tracking to investigate topics in L2 acquisition and L2 processing. *Studies in Second Language Acquisition*, 35(2), 213–235. https://doi.org/10.1017/S0272263112000861

Sabourin, L., & Stowe, L. A. (2008). Second language processing: When are first and second languages processed similarly. *Second Language Research*, 24(3), 397–430. https://doi.org/10.1177/0267658308090186

Sagarra, N., & Herschensohn, J. (2010). The role of proficiency and working memory in gender and number agreement processing in L1 and L2 Spanish. *Lingua*, 120(8), 2022–2039. https://doi.org/10.1016/j.lingua.2010.02.004

Sagarra, N., & Herschensohn, J. (2011). Proficiency and animacy effects on L2 gender agreement processes during comprehension. *Language Learning*, 61(1), 80–116. https://doi.org/10.1111/j.1467-9922.2010.00588.x

Spinner, P., Gass, S. M., & Behney, J. (2013). Ecological validity in eye-tracking: An empirical study. *Studies in Second Language Acquisition*, 35(2), 389–415. https://doi.org/10.1017/S0272263112000927

Spino-Seijas, L. (2017). Grammatical gender agreement in L2 Spanish: The role of syntactic context (Unpublished PhD dissertation). Michigan State University.

Unsworth, S. (2008). Age and input in the acquisition of grammatical gender in Dutch. *Second Language Research*, 24(3), 365–395. https://doi.org/10.1177/0267658308090185

VanPatten, B., Keating, G. D., & Leeser, M. J. (2012). Missing verbal inflections as a representational problem: Evidence from self-paced reading. *Linguistic Approaches to Bilingualism*, 2(2), 109–140. https://doi.org/10.1075/lab.2.2.01pat

Vitu, F., O'Regan, K., Inhoff, A. W., & Topolski, R. (1995). Mindless reading: Eye-movement characteristics are similar in scanning letter strings and reading texts. *Perception and Psychophysics*, 57(3), 352–364. https://doi.org/10.3758/BF03213060

Wen, S., Miyao, M., Takeda, A., Chu, W., & Schwartz, B. (2010). Proficiency effects and distance effects in nonnative processing of English number agreement. In K. Franich, K. Iserman & L. Keil (Eds.), *Proceedings of the 34th annual Boston University Conference on Language Development* (pp. 445–456). Cascadilla Press.

White, L., Valenzuela, E., Kozlowska-Macgregor, M., & Leung, Y.-K. I. (2004). Gender and number agreement in nonnative Spanish. *Applied Psycholinguistics*, 25, 105–133. https://doi.org/10.1017/S0142716404001067

Witzel, N., Witzel, J., & Forster, K. (2012). Comparisons of online reading paradigms: Eye tracking, moving-window, and maze [Research Support, Non-U.S. Gov't]. *Journal of Psycholinguistic Research*, 41(2), 105–128. https://doi.org/10.1007/s10936-011-9179-x

CHAPTER 9

Cross-linguistic influence in bilingual grammars

Evidence from gender assignment in unilingual Dutch and mixed speech

Brechje van Osch[1,3], Ivo Boers[1,2], Janet Grijzenhout[1,2], M. Carmen Parafita Couto[1,2,5], Bo Sterken[1,2] and Deniz Tat[1,3,4]
[1]Heritage Linguistics Lab, HERLING / [2]Leiden University, Centre for Linguistics, LUCL / [3]UiT The Arctic University of Norway / [4]Leiden University, Institute for Area Studies, LIAS / [5]Leiden Institute for Brain and Cognition, LIBC

This study reports on grammatical gender assignment in elicited production data from heritage speakers of Turkish, Papiamento, and Spanish in the Netherlands. We investigate the role of cross-linguistic influence from the heritage language onto the societal language by comparing three heritage languages that differ in terms of the properties of the nominal domain, including gender. Determiner-adjective-noun constructions were elicited by means of a Director-Matcher task (Gullberg, Indefrey, & Muysken, 2009), which was performed both in a unilingual Dutch mode, and in a code-switching mode from Dutch to the heritage language. The results show that all groups tend to overgeneralize the common gender in the Dutch unilingual mode. Strikingly, the performance of heritage speakers of Spanish was more target-like than the Papiamento and Turkish speakers, which may be due to the fact that Spanish is the only language that has a grammatical gender system. In code-switching mode, most speakers tend to assign common gender to inserted nouns, but some speakers also apply a gender assignment strategy based on the translation equivalent of the noun in Dutch, or produce a postnominal adjective construction with an uninflected adjective. An analysis of extra-linguistic variables demonstrated that gender assignment strategies seem to be determined to some extent by the degree of dominance in the societal language.

Keywords: cross-linguistic influence, bilingualism, code switching, grammatical gender

https://doi.org/10.1075/sibil.63.09van

1. Introduction

Children acquire gender agreement in the noun phrase of their first language along different timelines. A complex system with three genders (masculine, feminine, neuter) as in German is acquired relatively fast with use of gender-marked articles by age 1;5 and nearly 90% accuracy rates by age 3;0 (Szagun et al., 2007). Two-way gender systems are acquired at different rates. A system with phonological regularities such as Italian or Spanish is acquired with high accuracy rates at an early age for both monolinguals and bilinguals (e.g. Paolieri et al., 2010; Montrul, 2004). A less transparent system such as French is acquired slightly later (e.g. Kupisch, Müller, & Cantone, 2002) while a highly opaque system such as Dutch may take seven to nine years to be fully acquired (e.g. Cornips & Hulk, 2006).

This study considers strategies for gender marking in Dutch by (pre)adolescent and adult heritage speakers (HSs) of three different languages. Our main aim is to explore the role of cross-linguistic influence (CLI) from the heritage language (HL) onto the majority language. CLI is often discussed as a source of non-target-like behaviour in bilingual research. However, since most research focuses on a single language pair, it is generally impossible to distinguish CLI from issues related to bilingualism itself such as reduced input or cognitive issues. The present study offers a valuable contribution by comparing HSs of three different languages in the Netherlands regarding their strategies for gender marking in Dutch. Specifically, we look at two languages that do not exhibit grammatical gender marking (Turkish, Papiamento) and one that does (Spanish). We consider CLI to be any type of influence – whether facilitative or not – from the structure of one language onto the other. In addition to gender assignment, we also touch upon two other phenomena related to the nominal domain that constitute potential sources of CLI between the languages of interest, namely adjective noun order and definiteness.

By including HSs of different ages (children, teens and adults), we aim to bridge the gap between HL studies and early child bilingualism studies (see also Aalberse & Hulk, 2018). HL studies tend to include only adults, and focus exclusively on the HL (e.g. Benmamoun, Montrul, & Polinksy, 2013), often (implicitly) assuming dominance in the societal language. On the other hand, (early) child bilingualism studies take both languages into consideration, and generally conclude that both are acquired in a monolingual-like manner (e.g. De Houwer, 1990; Meisel, 2004, 2009). However, they typically do not follow children beyond the (pre)-school period. In our study, we demonstrate that HSs show a considerable degree of non-target-like behaviour in Dutch, their (supposedly dominant) societal language. Moreover, we include bilinguals of a wide range of ages, and instead of a rigid division between child and adult participants, we treat age as a continuous variable that is included as one of the extra-linguistic predictor variable in our analysis.

Finally, our study provides an important contribution to the field by testing bilingual speakers both in unilingual mode, that is, when they only speak Dutch and in code-switching mode, that is, when they embed nouns from the HL into Dutch (the matrix language), resulting into mixed determiner phrases (DPs).

In the following section, we describe the differences between the four languages under consideration with respect to the properties of the nominal domain.

2. The nominal domain in Dutch, Spanish, Turkish and Papiamento

Standard Dutch exhibits a binary gender system consisting of common and neuter nouns. Gender is marked on determiners, adjectives, and pronouns. Common gender, which used to incorporate masculine and feminine, is the larger class, both in terms of simplex dictionary entries (75%; Hulk & Cornips, 2006) and in terms of usage (67%; van Berkum, 1996). The canonical word order in the Dutch DP is D(Adj)N. Nouns in the common gender select the definite article **de**,[1] whereas neuter nouns select the definite form **het**. Plural nouns of both genders are assigned the article **de.** Other elements in the DP modified by the gender of the noun are demonstratives (common gender **deze** 'this' and **die** 'that', neuter **dit** 'this' and **dat** 'that') and the first plural possessive pronoun (common gender **onze**, neuter **ons**). Adjectives are prenominal and are inflected with -e /ə/ at all times, except in indefinites with a neuter noun, as exemplified in (1):

(1)	a.	common gender DP with definite article:	**de grote boom**	'the big tree'
	b.	common gender DP with indefinite article:	**een grote boom**	'a big tree'
	c.	neuter gender DP with definite article:	**het grote huis**	'the big house'
	d.	neuter gender DP with indefinite article:	**een groot huis**	'a big house'

The pronominal system has traditionally been analysed as a mixed system where for animate objects the pronoun agrees with the natural (semantic) gender (**hij** for masculine, **zij** for feminine and **het** for neuter), while for inanimate objects, syntactic agreement occurs: **hij** for common gender nouns and **het** for neuter gender nouns (e.g. Geeraerts, 1992).

1. Throughout this chapter, bold text is used for Dutch, while italics are used for the heritage languages and other languages.

However, in spoken Dutch, the distribution of personal pronouns referring to inanimate objects is subject to a process of resemanticization, whereby **hij** is sometimes used for count nouns and **het** for mass nouns, regardless of the syntactic gender (Audring, 2006; Kraaikamp, 2017).

Although the Dutch gender system is generally opaque, some derivational suffixes function as cues, the most important one being the diminutive suffix **-je** or one of its allomorphs, which are always neuter (e.g. **het**.N **boek**-**je**.N – 'the little book')

The Turkish nominal domain differs from Dutch in that it does not have a grammatical gender distinction, and it does not have a definite article (2a). Indefinite singulars (and some plural and mass nouns) are introduced by the determiner *bir* (2b), which is generally assumed to differ from the numeral *bir* due to distributional differences (e.g. Kornfilt, 1997: 106; Göksel & Kerslake, 2004: 179; Von Heusinger & Kornfilt, 2005; Tat, 2010; Kornfilt, 2017).

(2) a. *Büyük ev amca-m-ın* definite DP
big house uncle-1SG.POSS-GEN
'The big house is my uncle's.'
b. *Büyük bir ev hayal-im* indefinite DP
big a house dream-1SG.POSS
'A big house is my dream.'

Papiamento is a Portuguese creole spoken on the ABC islands (Aruba, Curaçao, Bonaire), relexified by Spanish and influenced by Dutch due to an extended period of contact (Jacobs, 2012). Like Turkish, Papiamento lacks a gender distinction, but unlike Turkish, it exhibits a definite article and postnominal adjectives as in (3a, b).[2]

(3) a. DP with definite article: *e kas grandi*[2] 'the big house'
b. DP with indefinite article: *un kas grandi* 'a big house'

Spanish displays the same DN(Adj) order as Papiamento, although prenominal adjectives are possible for some adjectives in certain contexts in both languages. In contrast to Papiamento and Turkish, Spanish has a gender system, but unlike Dutch, it distinguishes between masculine and feminine gender. Gender is marked on all determiners (*el* and *la* for the definite masculine and feminine articles respectively, and *un* and *una* for the indefinite counterparts), pronouns, and on most adjectives (Caroll, 1989). Most noun endings are a cue to gender, for instance, masculine nouns tend to end in *-o,* while feminine nouns tend to end in *-a*. These are so-called canonical or transparent nouns, while nouns ending in different vowels or consonants are referred to as non-canonical or non-transparent nouns. Table 1 provides examples of Spanish nominal constructions.

2. Papiamento words follow the Curaçaoan spelling used by Dijkhoff (2016).

Table 1. Agreement in Spanish nominal constructions

	Canonical noun	Non-canonical noun
Masculine	*el* / *un libro pequeño* 'the / a small book'	*el* / *un peine pequeño* 'the / a small comb'
Feminine	*la* / *una mesa pequeña* 'the small table'	*la* / *una flor pequeña* 'the small flower'

Note. The elements showing gender agreement have been underlined.

Table 2 illustrates the relevant differences and similarities between the four languages under discussion.

Table 2. Articles, gender marking, and Adj Noun order in four languages

	Definite article	Indefinite article	Gender marking	Gender inflection adjective	Word order within the DP
Dutch	*yes*	*yes*	*yes*	*yes*	Det Adj Noun
Turkish	no	*yes* (optional)	no	no	Adj (Det) Noun
Papiamento	*yes*	*yes*	no	no	Det Noun Adj
Spanish	*yes*	*yes*	*yes*	*yes*	Det Noun Adj

Apart from the structural differences between the three HLs of interest, there are also important differences between the Turkish, Antillean and Spanish-speaking communities in the Netherlands, especially in terms of length of language contact, as discussed next.

3. The Turkish-, Spanish- and Papiamento-speaking communities in the Netherlands

Turkish is spoken by one of the largest immigrant communities in the Netherlands (409,877 people according to the Central Bureau of Statistics (CBS 2019)[3] and consists mainly of migrant workers who arrived in the 1960s and early 1970s, their families who were reunited in the late 1970s, and their descendants. There is also a smaller community of political migrants who arrived in the 1980s.

Spanish is a smaller, yet still relatively large immigrant language in the Netherlands (with 130,160 people; CBS, 2019). About a third of them have origins in Spain (46,741), but the majority comes from a variety of Latin American countries.

3. These numbers only include speakers of the first and second generations, that is, immigrants and the children of at least one immigrant.

Among the Spanish-speaking migrants were contracted workers from Spain in the 1960s and 1970s, political refugees from Latin America during the 1970s and 1980s, and more migrants during the 1990s (mainly from the Dominican Republic and Colombia) (Irizarri van Suchtelen, 2016). Compared to Turkish HSs, Spanish HSs in the Netherlands are less inclined to form close-knit communities; they live more dispersed throughout the country (van Osch, 2019).

Whereas Turkish and Spanish have a relatively short history of contact, Papiamento and Dutch share more than three and a half centuries of history. Nowadays, 161,265 Dutch Antilleans live in the (European) Netherlands (CBS, 2019).[4] This sociologically very diverse group was already proficient in Dutch before their emigration, in contrast to the other two groups of HSs, although the Dutch spoken on the ABC islands and probably by migrants of the first generation as well is considerably different from European Dutch (Depuydt, 2010). Aalberse, Backus, & Muysken, (2019) consider Papiamento in the Netherlands as a post-colonial HL, comparable to Hindi in the UK, for instance.

4. Dutch gender in bilinguals

Several studies observed a differential acquisition of Dutch gender in bilinguals when compared to monolinguals. A frequently reported pattern is that bilinguals tend to overgeneralize the common gender. This has been found for Dutch in contact with English, a non-gendered language (Unsworth et al., 2014), as well as for Dutch in contact with Moroccan Arabic/Berber, which lack gender (Blom, Polišenská, & Weerman, 2008). Hulk and Cornips (2006) found an overextension of common gender for a group of child HSs of several different languages, some with gender (French, Moroccan Arabic/Berber) and others without (Turkish, Akan, Ewe, Sranan).

Some studies suggest that the acquisition of Dutch gender in bilinguals is less problematic or can even be accelerated if the other language has gender. For instance, Egger, Hulk, and Tsimpli (2018) found that Dutch gender acquisition in Greek child HSs (aged 4;4–13;3) is accelerated – at least in the initial stages – by CLI from Greek. Similarly, Hulk and van der Linden (2010) argue that having French and Spanish as one of the languages accelerates the acquisition of gender in Dutch. Similar claims of CLI in the form of acceleration or delays in the acquisition of gender systems have been made for bilingual children of other language pairings (e.g. Eichler, Jansen, & Müller, 2013; Schwartz et al., 2015; Kaltsa, Tsimpli,

4. This number includes people who do not necessarily speak Papiamento, such as migrants from Sint Maarten, Sint Eustatius and Saba.

& Argyri, 2019; but see Rodina et al., 2020). Furthermore, some studies on pronominal gender agreement in the (L2) acquisition of Dutch seem to suggest that the type of agreement strategy applied by learners (semantic vs. grammatical) may also be influenced by the specific type of gender system in the other language (see Aalberse & Weerman, in prep for an overview).

In previous work (Boers et al., 2020), we analysed elicited production data of Spanish HSs in the Netherlands (one of the groups that are also considered in the present chapter), and found that their command of Dutch gender assignment was considerably non-target-like, even more so than in Spanish, their HL. In line with other research, these speakers showed an overextension of common gender. In the present chapter, we complement these findings with data from two other (non-gendered) HLs in the Netherlands, in order to shed more light on the question of CLI.

The acquisition of gender in bilinguals also seems to depend on various extralinguistic factors related to the linguistic experience of the individual (e.g. Gathercole & Thomas, 2005; Rodina & Westergaard, 2017; Mitrofanova et al., 2018). For Dutch gender in particular, Unsworth et al. (2014) found that factors related to the quantity and the quality of the input as well as language use by the children were important predictors of bilingual children's accuracy in assigning common or neuter gender in Dutch. Cornips and Hulk (2008) report effects of age of onset and length and intensity of the input as important predictors for accuracy in Dutch grammatical gender in bilinguals. Boers et al. (2020) showed that age of arrival to the Netherlands, self-reported proficiency in Spanish, and the amount of 'other' exposure to Spanish (books, TV, social media, music) affected Spanish HSs' accuracy with Dutch gender.

5. Gender in code-switching

Several strategies for gender assignment in code-switching have been identified in the literature. Many studies report a default gender strategy by which means bilinguals assign assign a single gender to most nouns for instance masculine gender in Spanish for Spanish-English bilinguals (e.g. Otheguy & Lapidus, 2003; Balam, 2016; Valdés Kroff, 2016). For Dutch, a common gender default strategy is used by Dutch-English bilinguals (HSs and L2 speakers) living in Australia (Clyne, 1977; Clyne & Pauwels, 2013) and by Dutch-Arabic bilinguals in the Netherlands (Boumans, 1998). Treffers-Daller (1993), on the other hand, mentions a neuter gender default in Dutch-French bilinguals in Brussels, which may be related to the fact that Brussels Dutch is a three-way (masculine, feminine and neuter) gender system.

Another commonly reported gender assignment strategy is the analogical gender strategy, whereby bilinguals assign the gender of the translation equivalent of

the noun. Thus, Spanish-English bilinguals might say "*la*.FEM table", because its translation, *mesa*, is feminine in Spanish (e.g. Jake, Myers-Scotton & Gross, 2002). This strategy has also been attested for code-switching in Dutch (Clyne, 1977; Clyne & Pauwels, 2013).

French-Brabant Dutch bilinguals choose another analogical strategy by using the gender of the inserted noun itself, that is, masculine French nouns inserted in Dutch agree with Dutch masculine determiners (Treffers-Daller, 1993). Finally, some studies (e.g. Bellamy, Parafita Couto, & Stadthagen-González, 2018; Parafita Couto et al., 2015) report a gender assignment strategy based on the morpho-phonological ending of the noun. However, this strategy is not relevant for Dutch due to its opaque nature.

Interestingly, some studies attested different gender strategies for the same language combination depending on the sociolinguistic characteristics of the community or the individual speaker. Królikowska et al. (2019) found that Spanish-English communities where code-switching is highly frequent are more prone to apply a masculine default strategy when code-switching than communities where it is less common.

Liceras et al. (2008) looked at Spanish-English and Spanish-French bilingual children for whom Spanish was either the L1 or the L2 and found that the analogical gender strategy was used more often by the Spanish L1 speakers, while the masculine default strategy was preferred by the Spanish L2 speakers. Similarly, Munarriz-Ibarrola et al. (2021) demonstrated that Spanish-Basque bilinguals who acquired Spanish first were more likely to use the analogical gender strategy, while L1 Basque speakers tended to use phonological cues in gender assignment.

In Boers et al. (2020), we found that Spanish HSs use a mix of gender assignment strategies when they insert Spanish nouns into Dutch phrases. A default gender strategy with the common gender was the most frequently used strategy, for all but two of them. The analogical gender strategy was also adopted by several speakers, particularly by those more dominant in Dutch. Finally, some speakers used a construction containing a postnominal uninflected adjective (e.g. **een** *cama* **zwart** 'a bed black').

6. Research questions and hypotheses

To explore the linguistic and extralinguistic variables that play a role in gender assignment in a Dutch unilingual mode and in code-switched speech, we ask the following research questions: (a) does the presence or absence of a grammatical gender system in one's HL determine the degree to which the Dutch grammatical system is target-like?; (b) do code-switching strategies differ depending on the HL?;

(c) which extralinguistic factors are predictors for the linguistic outcomes of the acquisition of Dutch grammatical system and for gender assignment strategies in code-switching?

In the unilingual mode, we expect to find evidence for an overextension of the common gender (cf. Unsworth et al., 2014; Blom et al., 2008; Hulk & Cornips, 2006). However, we also expect to find potential differences between groups related to the particular properties of the nominal domain in their HL. First, we hypothesise that having a gender system in the additional language will facilitate the use of a two-way gender system in the other language (Spanish/Dutch), that is, we expect the Turkish and the Papiamento HSs to perform less-target-like than their Spanish-speaking peers (cf. Egger et al., 2018; Hulk & van der Linden, 2010). Moreover, we expect other differences between the languages in the nominal domain with respect to word order and definiteness to potentially result in differences between the three groups regarding gender in Dutch.

We also expect extralinguistic factors such as the age of arrival to the host country, and the amount of input and use of the language in several domains to have an effect on gender accuracy in Dutch (cf. Unsworth et al., 2014; Cornips & Hulk, 2008).

As for code-switching, we expect to observe a common default strategy (cf. Clyne, 1977; Clyne & Pauwels, 2013; Boumans, 1998; Boers et al., 2020), possibly in addition to other strategies. Moreover, we hypothesize that the correlation between language dominance and the use of the analogical gender strategy – demonstrated for HSs of Spanish in Boers et al. (2020) – will be applicable to the HSs of Turkish and Papiamento as well; in other words, those speakers who are more dominant in Dutch are expected to assign gender based on the gender of the translation equivalent more often.

7. Methodology

7.1 Procedure

A Director-Matcher task (Gullberg et al., 2009) was used to elicit nominal constructions consisting of a determiner, an adjective, and a noun. This task requires two participants, a director and a matcher sitting across from each other with a division between them, both with the same set of cards laid down before them. One participant (the Director) instructs the Matcher to arrange the cards in the same order as his or her own, thus eliciting phrases like "next to the black bed is a red house". The task was performed four times: first in two unilingual modes (Dutch and the respective HL) and then in two code-switching modes, in which they were instructed to perform the task in Dutch, but name only the object in the HL and vice

versa. In this chapter, we only report the results from the Dutch unilingual mode and the code-switching mode with HL noun insertions into Dutch. The results for the Spanish HSs (in Spanish, Dutch, and both code-switching modes) are reported in Boers et al., 2020).

Prior to the experiment, participants received oral instructions in the language of preference, as well as a consent form. After the completion of the task, the participants (or the participant's care-giver(s) for children younger than 12) were asked to complete the background questionnaire in their preferred language. Participants older than 12 received monetary compensation, while younger children received a small toy.

7.2 Materials

The task was designed to test gender assignment accuracy in the Dutch unilingual mode and gender assignment strategies in code-switching mode. Table 3 provides hypothetical code-switching strategies. If a default strategy is used, most nouns would be assigned to one gender category, either common or neuter. If the analogical strategy is applied, nouns would be assigned the gender of the translation equivalent.

Table 3. Examples of hypothetical gender assignment strategies with Papiamento nouns embedded in Dutch

Gender strategy	***bòter* 'bottle' (cf. de fles)**	***wowo* 'eye' (cf. het oog)**
Default gender (common)	**de** *bòter*	**de** *wowo*
Default gender (neuter)	**het** *bòter*	**het** *wowo*
Analogical gender	**de** *bòter*	**het** *wowo*

To avoid influence from morphological cues, the objects were selected so that they did not target any nouns containing suffixes that might indicate their gender. Since Papiamento and Turkish lack grammatical gender, the depicted objects were counterbalanced for their gender in Dutch and in Spanish. The objects were also counterbalanced for canonicity in Spanish. Cognates between Dutch and the three HLs were avoided. Finally, the objects were controlled for lexical variation in different Spanish dialects (cf. Balam et al., 2021). Because it was impossible to find 8 neuter nouns that fit all the required criteria, the set contained 8 common gender objects and 7 neuter gender objects.

The objects were depicted in color adjectives – black, white, green, and red – that inflect in Dutch. All but *verde* 'green' inflect in Spanish. Dutch colour adjectives that were phonologically similar to HL colour adjectives were avoided. Every object occurred twice in a different colour. (See Appendix A for all the objects and colours).

A background questionnaire was designed to obtain information about language history, education, and relative amount of Dutch and HL use and exposure. For participants younger than 12 years, the questionnaire was completed by (one of) the parents. and also included questions on the parents' language history, education and language usage within the family.

7.3 Participants

A total of 58 HSs (29 male, 29 female) who lived in the western part of the Netherlands (Randstad area) participated in the study. We tested a wide range of speakers of different ages, with different ages of arrival and language use patterns, with the aim of including such variables as predictors in our analyses.[5] As a consequence, the groups differed significantly on some of these variables, as illustrated in Tables 4 and 5 (light vs. dark-grey cells indicate a significant difference between two or more groups).

Table 4. Extralinguistic information according to HL group. Light vs. dark grey indicates a significant difference

	Spanish	Turkish	Papiamento
Number of participants	21	22	15
Gender	6 M, 15 F	15 M, 7 F	8 M, 7 F
Age at testing (range = 7 to 55)	17.19	24.82	27.27
Age of arrival (range = 0 to 21)	3.24	2.18	8.27
Years spent in the Netherlands	13.57	21.86	18.87
Years spent in heritage country	3.86	2.32	8.40
Self-reported HL skill (0–3)[a]	2.42	2.11	2.13
Self-reported Dutch usage[b]	60.97%	53.26%	58.28%
Self-reported HL usage[c]	35.12%	46.63%	39.97%
Other (media) HL input (hours/week) (range = 0 to 70)[c]	12.14	17.64	8.18
Participants who took HL classes	5/21	2/22	0/15
Heritage country visits (0–3)[d]	2.38	2.32	2.40

a. Average of self-reported reading, writing, listening, and speaking skill (0 = none, 3 = advanced).
b. Average of self-reported usage with different family members and five frequent non-family contacts.
c. Sum of self-reported input from different categories (books, television, music, social media).
d. Frequency of visit to the heritage country (0 = never, 3 = at least once a year).

5. We are aware that not all our participants would be considered HSs under all definitions. Most definitions use some cut-off point for age of arrival. However, this cut off point differs between studies (Ortega, 2020) and any particular cut-off point may be arbitrary. Instead, we included a wide range of ages of arrival to use as a predictor variable.

Table 5. Additional extralinguistic information for child participants according to HL group. Light vs. dark grey indicates a significant difference

	Spanish	Turkish	Papiamento
Dutch usage immediate family[e]	50.83%	40.75%	47.63%
HL usage immediate family[e]	44.17	59.25%	52.38%
Dutch usage non-immediate family[f]	73.75%	82.50%	94.50%
HL usage non-immediate family[f]	23.75%	17.50%	5.50%
Dutch input aged 0–4[g]	33.13%	25.00%	21.75%
HL input aged 0–4[g]	60.34%	75.00%	78.25%
Dutch input aged 4 till present[g]	38.44%	27.50%	49.00%
HL input aged 4 till present[g]	54.63%	72.50%	51.00%

e. Average of parent-reported usage with members of the immediate family (parents, siblings).
f. Average of parent-reported usage with five frequent non-immediate contacts.
g. Average of input from the parents, reported by the parents.

7.4 Coding

Participants sometimes produced a different type of speech from what was intended in a certain mode. Unilingual nominal constructions used in the code-switching mode were excluded from the analysis, since they did not represent regular unilingual speech (i.e. they had full nominal constructions from one language embedded in the other). Code-switched constructions that were not in the target directionality were included in the analysis.

If participants referred more than once to the same object (e.g. "next to A is B, and next to B is C"), both nominal constructions were included in the analysis. Nominal constructions that did not refer to the target stimuli (e.g. "the next card") were excluded.

8. Analysis and results

8.1 Unilingual Dutch mode

A total of 1958 cases were produced by all three groups of HSs combined. For each of these cases, the gender assigned to the DP was based on either the determiner (**de** for common, **het** for neuter), or, in the case of either an absent or indefinite determiner (1501 cases), on the adjective inflection (inflected indicating common,

uninflected indicating neuter[6]) (cf. Bellamy et al., 2018). The gender of the DP could not be unambiguously determined in 14 cases (e.g. some DPs contained an indefinite determiner and no adjective), so they were excluded.

In 36 cases (of which 28 were produced by the same Turkish HS participant), the adjective was postnominal, which is ungrammatical in Dutch. Most of these cases (n = 33) occurred with an indefinite determiner, and all but one had an uninflected adjective (e.g. **een hamer**.COM **rood**-Ø 'A hammer red'). These constructions may be cases of a predicative construction with a missing copula, rather than DP-internal agreement (cf. Cinque, 1994). Dutch adjectives are always uninflected when they are used in a predicative manner regardless of the gender of the noun (**de**.COM **boom**.COM **is groot**-Ø 'The tree is big' vs. **het**.NEUT **huis**.NEUT **is groot**-Ø 'The house is big').[7] These structures were therefore excluded from the analyses on gender assignment.

Apart from these 36 cases, the participants adhered to the grammatical word order in Dutch. For nouns for which the target was common (1042 cases), the common gender was assigned in 96.74% of the cases. In contrast, when the target was neuter (870 cases), it was assigned only 41.61% of the time.[8] This percentage was especially low with indefinite nouns (38.37%) compared to definite nouns (52.82%). These results are depicted in Figure 1.

---

6. An anonymous reviewer suggested that adjectival inflection may reflect gender agreement rather than assignment, and that bilinguals may have better mastery of assignment rules (selecting the correct determiner) than agreement rules (selecting the correct adjective). However, as mentioned below, definiteness was not a significant predictor for gender assignment. This means that target-like use of gender was similar for definite determiners and for adjectives. Moreover, in those cases in which both a definite determiner and an adjective were used, the gender of the definite determiner and the adjective matched (three exceptions were excluded from the analysis). We thus think that the inflection of the adjective can be used as an appropriate indicator of gender assignment.

7. An anonymous reviewer mentioned that postnominal adjectives were in fact possible in Middle Dutch (and are still present in certain poems and songs) and were always uninflected in that position.

8. An anonymous reviewer wondered about diminutives, which are a consistent cue for neuter gender. Participants used diminutives only 34 times, and of these 15 were assigned common gender, so the noun ending did not seem to function as a cue for gender assignment.

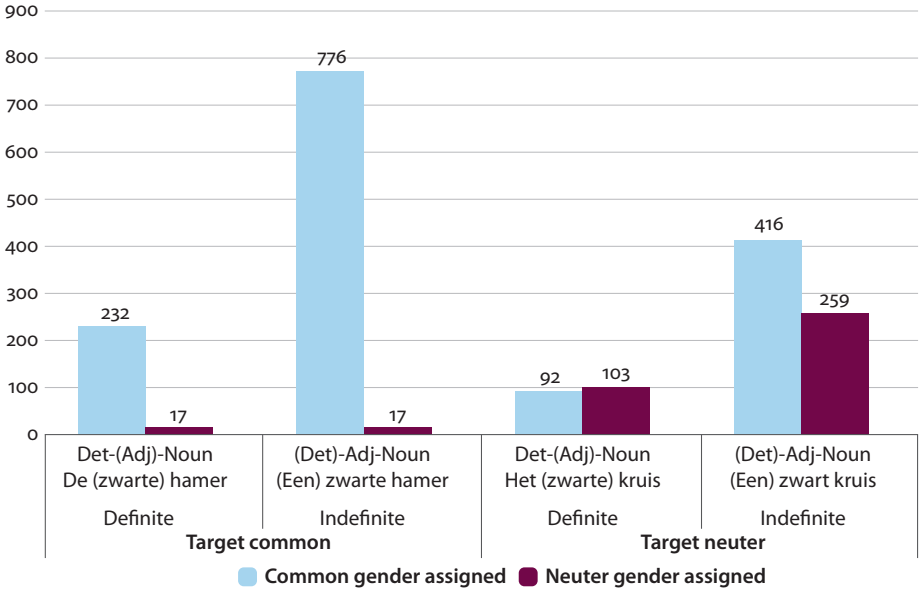


Figure 1. Gender assignment in (D)(Adj)N constructions in unilingual Dutch by target gender and definiteness

If we split the group into the three different HLs, some differences between the groups are revealed (Figure 2).

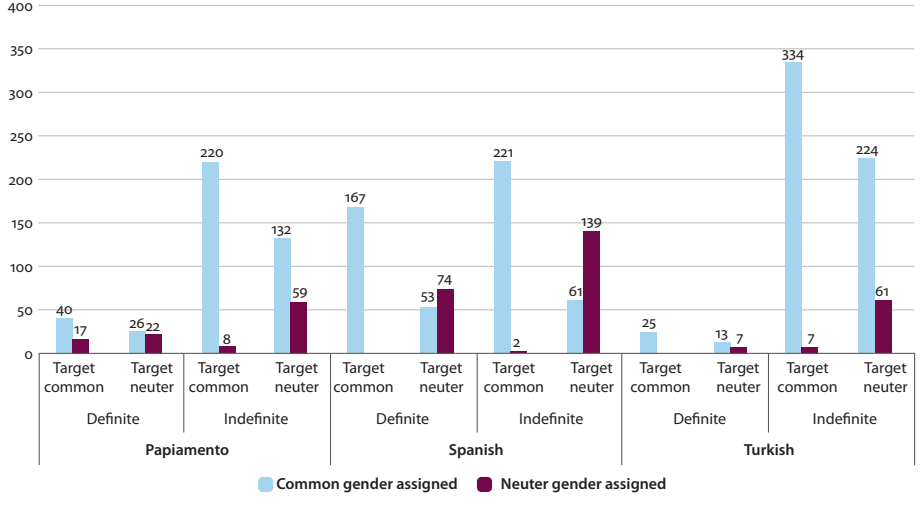


Figure 2. Gender assignment in (D)(Adj)N constructions in unilingual Dutch by group, definiteness, and target gender

The Papiamento and the Turkish HSs seem to behave similarly: they prefer to use an indefinite determiner, and they tend to prefer common gender irrespective of the target gender. The Spanish HSs, on the other hand, seem to produce definite and indefinite DPs more equally, and produce neuter gender more often when the target noun is neuter. In other words, they seem to be more target-like than the other two groups for neuter gender.

To check if any of these effects are significant, we ran a series of generalized linear mixed effects models using the lme4 package in R (R Development Core Team, 2019). In all models described in this section, independent variables and possible interactions were added to the model one by one in a forward regression manner, and nested models were compared by performing log likelihood ratio tests. Variables that did not improve the model significantly were excluded from the final model. Random intercepts and slopes for 'subject' and 'item' (the object that had to be described) were also included only if these significantly improved the model (following Baayen et al., 2008). All binary factors were coded as −0.5 vs 0.5. For the ternary variable HL, orthogonal sum-to-zero contrasts were used.

In the first model, the dependent variable was the definiteness of the determiner used, and only the HL (Papiamento, Spanish or Turkish) was considered a variable of interest. Adding this variable significantly improved the model. The effect of HL was significant both for contrast 1 (Turkish versus the mean of Papiamento and Spanish – $\beta = 5.17$, $SE = 1.37$, $z = 3.78$, $p < .001$) and contrast 2 (Papiamento versus Spanish – $\beta = 3.74$, $SE = 1.71$, $z = 2.19$, $p = 0.03$). This means that Turkish HSs used the indefinite determiner more than both other groups, and Papiamento HSs used the indefinite determiner more than the Spanish HSs. Turkish speakers were thus the most likely to use an indefinite determiner.

Another model was run to test which linguistic factors affected gender assignment. The dependent variable was target-likeness, that is, whether the gender assigned to the DP corresponded to the target gender. Predictor variables taken into consideration were the target gender (common vs. neuter), the definiteness of the DP (definite vs. indefinite) and the HL (Papiamento, Spanish or Turkish). The best fitted model included significant effects of both target gender ($\beta = 5.91$, $SE = 0.82$, $z = 7.24$, $p < .001$) and HL ($\beta = 3.34$, $SE = 0.75$, $z = 4.44$, $p < .001$), but not definiteness. The intercept of the model was significant as well ($\beta = 1.73$, $SE = 0.39$, $z = 4.48$, $p < .001$), which means that overall, participants produced more target-like than non-target-like constructions. The main effect of target gender indicated that participants were more target-like with common nouns than with neuter nouns, indicating an overextension of the common gender. The effect of HL ($\beta = 3.34$, $SE = 0.75$, $z = 4.44$, $p < .001$) was significant for the first contrast (Spanish versus the mean of Papiamento and Turkish), indicating that Spanish HSs were more target-like than the Papiamento and the Turkish speakers combined. This effect of

HL interacted with target gender, although not significantly ($\beta = -2.79$, $SE = 1.58$, $z = -1.76$, $p = .078$). The direction of the effect suggests that the advantage for the Spanish HSs compared to the other groups is more pronounced when the target is neuter than when the target is common gender. To explore this interaction further, we conducted Tukey post hoc comparisons which revealed that indeed, for neuter gender, the Spanish HSs were significantly more target-like than both the Turkish speakers ($z = 3.71$, $p = .002$) and the Papiamento speakers ($z = -2.73$, $p = .045$). For common gender, the only significant difference was between Spanish and Papiamento speakers ($z = -2.95$, $p = .02$).

8.1.1 *Extralinguistic variables*

To investigate the influence of extralinguistic variables on the HSs' performance with gender in Dutch, we carried out another statistical analysis with the extralinguistic information deduced from the background questionnaire. Apart from HL (Spanish, Papiamento and Turkish), the following independent variables were taken into consideration: age at testing, age of arrival to the Netherlands, length of residence in the Netherlands, years spent in the heritage country, number of visits to the heritage country, instruction in the HL, self-reported proficiency in the HL, usage of the HL and Dutch, both with immediate family[9] and outside of the family for children younger than 12, and reported hours of other types of input/use in the HL such as through TV, social media, books and/or music. As some of these variables (i.e. age at testing, age of arrival, length of residence in the Netherlands, length of residence in heritage country) correlated, we first ran a set of models including each of these factors as a single variable, to avoid issues with multicollinearity. Of this set, only the model containing age of arrival turned out to be significant. We therefore chose this variable to be included in the larger model. The optimal model included a significant main effect of the HL for contrast 1 (Spanish vs. the mean of Papiamento and Turkish) ($\beta = 1.97$, $SE = 0.50$, $z = 3.91$, $p < .001$) similar to the previous model. There was also a main effect of total usage of the HL ($\beta = -3.50$, $SE = 1.18$, $z = -2.98$, $p = .002$), which indicates that the less the HL was used, the more target-like gender assignment was in Dutch. Age of arrival was another significant predictor in the model ($\beta = -0.14$, $SE = 0.05$, $z = -2.73$, $p = .006$), meaning

9. An anonymous reviewer asked about the effect of having one or two non-Dutch speaking parents. While we have this information, we did not include it as a variable because we believe it is very coarse-grained. A native Turkish/Papiamento/Spanish parent may speak Dutch at home, just as a native Dutch parent may speak the HL if s/he learned it as an L2. Instead, our questionnaire asked about the input from different family members/in different contexts in both languages in percentages and we calculated average percentages based on those numbers. We believe this reflects amount of input more precisely than a categorical variable such as mixed vs. non-mixed.

that the younger a HS was when s/he arrived in the Netherlands, the higher the accuracy score with gender in Dutch. Finally, there was a significant main effect of the general proficiency level as reported by the participants ($\beta = 1.48$, $SE = 0.38$, $z = 3.88$, $p < .001$), which indicated that, perhaps surprisingly, better performance with gender in Dutch correlated with higher proficiency in the HL.

Given that the children's questionnaire was somewhat different (e.g. it contained an additional question about previous input, and a distinction between usage of the HL within the immediate family and outside the family), a separate analysis was carried out for this group. In this model, similar to the full model, there was a significant main effect of the general proficiency in the HL (as reported by the parents in this case) ($\beta = 3.40$, $SE = 0.88$, $z = 3.87$, $p < .001$), indicating that, the higher the proficiency in the HL, the higher the accuracy with Dutch gender. Furthermore, there was an effect of the usage of the HL with the immediate family ($\beta = -7.40$, $SE = 2.18$, $z = -3.40$, $p = < .001$), which meant that with more usage of the HL at home, gender accuracy in Dutch decreased.

8.2 Code-switching mode

In code-switching mode, participants were instructed to perform the task in Dutch, but name the object in their HL. This led to a total production of 1810 mixed DPs (a Dutch determiner and adjective followed by a noun in the HL). Of these, 1413 DPs contained an indefinite or absent determiner, and 397 contained a definite determiner. As in the unilingual mode, there were differences between the three groups with respect to definiteness: Spanish HSs produced relatively more definite DPs (36.44%) compared to the Papiamento (23.68%) and the Turkish HSs (7.65%). To test whether these differences were significant, a generalized mixed effects model was run with the definiteness of the determiner as the dependent variable, and the HL as the independent variable. Including HL significantly improved the model, and the effect of HL was significant for the contrast between Turkish and the other two languages ($\beta = 6.38$, $SE = 1.89$, $z = 3.38$, $p < 0.001$), meaning that the Turkish HSs used the indefinite determiner significantly more often than the other two groups when code-switching into Dutch (Figure 3).

Interestingly, in a relatively large proportion of all elicited DPs (234 cases), the Dutch adjective was placed following the noun, which is not a possible word order in Dutch. As mentioned in Section 8.1 these could not be coded as either common or neuter, given that they can probably be considered predicative constructions. They will be analysed separately below.

For all other DPs, we coded the assigned gender either based on the determiner, when it was definite, or on the (prenominal) adjective, when the determiner was

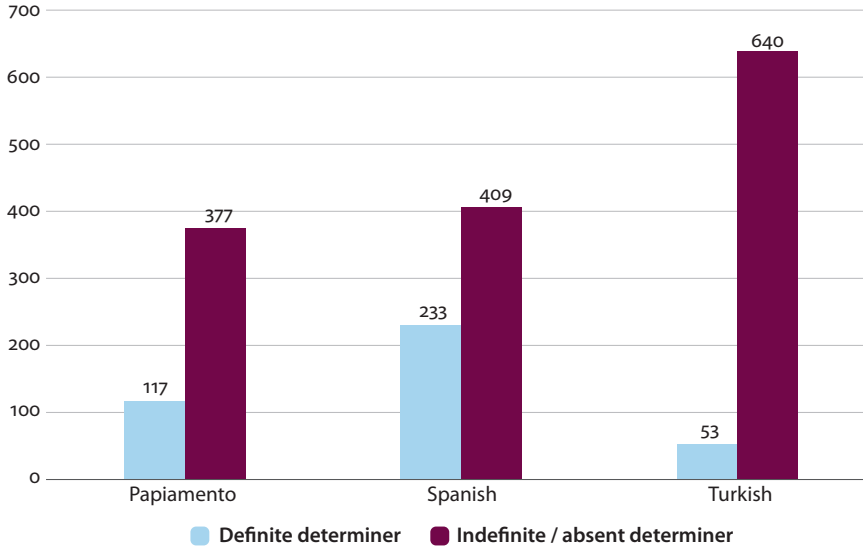


Figure 3. Production of definite and indefinite determiners in code-switched speech of Papiamento, Spanish and Turkish HSs

indefinite or absent. We excluded 34 ambiguous cases for which the gender could not be determined, for instance because an inflected adjective was produced in combination with a non-Dutch definite determiner (*el* **rode**.COM/NEUT *banco* 'the red couch'). Of all the unambiguous cases with prenominal adjectives, 86.64% was assigned common gender, suggesting an overall common default. The preference for common gender differed between the three HLs, though: 80.82% for the Spanish HSs,, 87.28% for the Turkish HSs, and 93.31% for the Papiamento HSs. The gender assigned to the noun matched the translation equivalent in Dutch 57.87% of the time, but this number also differed depending on the HL (54.60%, 57.04% and 61.44% for the Papiamento, Turkish and Spanish speakers, respectively) and depending on the gender of the translation equivalent (91.80% for common gender nouns vs. only 19.24% for neuter nouns). Regarding definiteness, the distribution of common vs. neuter gender was very similar: in definite DPs, 87.46% of all cases were assigned common gender and 12.54% neuter, whereas in indefinite DPs the distribution was 86.42% common vs. 13.58% neuter.

An analysis was performed on all (D)(Adj)N constructions for all three HL groups combined. The dependent variable was the gender assigned to the DP. Independent variables taken into consideration were the HL and the gender of the translation equivalent, as well as the interaction between them. The optimal model contained a main effect of the gender of the translation equivalent ($\beta = 1.55$, $SE = 0.33$, $z = 4.66$, $p < .001$), and an interaction between HL and the gender of the

translation equivalent, for both contrast 1 (Spanish versus the mean of Papiamento and Turkish: $\beta = 1.85$, $SE = 0.55$, $z = 3.34$, $p < .001$) and contrast 2 (Turkish versus Papiamento: $\beta = 1.39$, $SE = 0.70$, $z = 1.99$, $p = .046$). The significant main effect of the gender of the translation equivalent indicated that the group as a whole assigned common gender relatively more often to those nouns whose translation equivalent is also common. In other words, participants made use of the analogical gender strategy. However, the interaction with HL showed that this effect was more pronounced for the Spanish HSs than for the Turkish and Papiamento speakers and that it was stronger in the Turkish speakers as compared to the Papiamento speakers. Post hoc comparisons using the Tukey test revealed that the effect of the gender of the translation equivalent was significant for the Spanish HSs ($z = 5.95$, $p < .001$) and for the Turkish HSs ($z = 3.58$, $p = 0.003$), but not for the Papiamento HSs ($z = 0.411$, $p = 0.99$), indicating that only the Turkish and the Spanish HSs made use of the analogical gender strategy. These differences are illustrated in Figure 4.

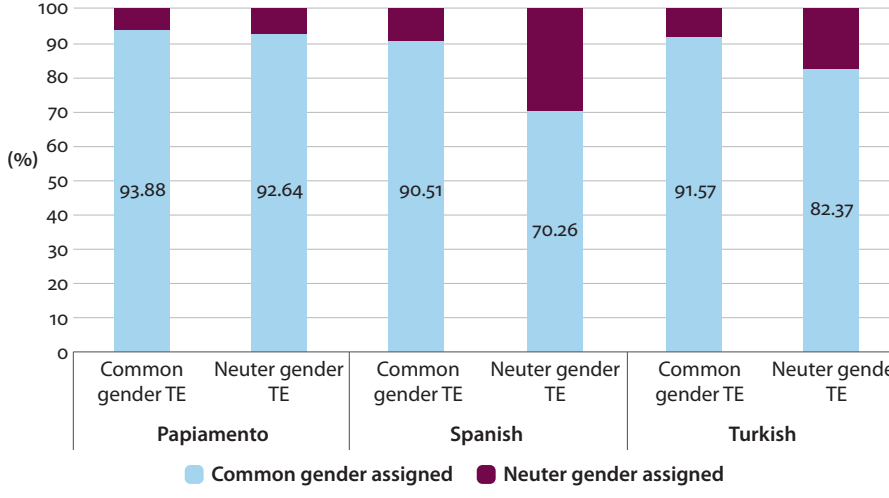


Figure 4. Gender assignment in code-switches by group and the gender of the translation equivalent (TE) in prenominal adjective constructions

The intercept of the model was significant as well ($\beta = 4.98$, $SE = 0.83$, $z = 5.98$, $p < .001$). This means that common gender was preferred across the board by all speakers and in all conditions. While this already suggests a common default strategy, we remind the reader that there were slightly more nouns with a common gender translation equivalent in our sample (805 vs. 707 neuter nouns). A more valid way to confirm a common gender default is therefore to look only at the mismatch cases, that is, those cases where the gender assigned to the DP did not match the gender of the translation equivalent. An analysis was performed on these

cases. To check whether the default effect was stronger for some groups than other, the variable group (Spanish, Papiamento and Turkish) was added to the model as a predictor, but it did not improve the model significantly. The final model included only a random effect of object, and contained a significant intercept ($\beta = 12.87$, $SE = 2.24$, $z = 5.75$, $p < .001$) indicating that, for all groups alike, common gender was assigned to neuter target nouns more often than the other way around, confirming a common gender default strategy.

Of the 234 constructions including a postnominal adjective, 126 were produced by Spanish HSs and 108 by Papiamento speakers, but none by Turkish HSs. Most of these constructions occurred in combination with an indefinite determiner, (211, of which 73 had the determiner in the HL). The adjective was almost always produced in Dutch (except for 8 cases), and interestingly, in all but one case (225 out of 226), it was uninflected) (e.g. **een** *cama* **zwart**-Ø – 'A bed black'). The postnominal construction was used slightly more with nouns that had a common gender translation equivalent (56.60%), than with a neuter gender translation equivalent (43.30%). To check whether this difference was significant, we performed an analysis with word order as the dependent variable. Gender of the translation equivalent, HL (Spanish and Papiamento), and definiteness, as well as the interactions between them, were added to the model one by one. The final model included a main effect of definiteness ($\beta = 2.35$, $SE = 0.96$, $z = 2.44$, $p = .01$), as well as an interaction between definiteness and HL ($\beta = -5.71$, $SE = 1.93$, $z = -2.95$, $p = .003$). The intercept of the model was also significant ($\beta = -11.2$, $SE = 1.72$, $z = -6.52$, $p < .001$), indicating that, overall, (D)AdjN order was more frequent than (D)Nadj order. The effect of definiteness means that postnominal adjectives were more likely to be used with indefinite determiners, and the interaction means that this effect of definiteness was stronger in the Papiamento group than in the Spanish group. These results are illustrated in Figure 5.

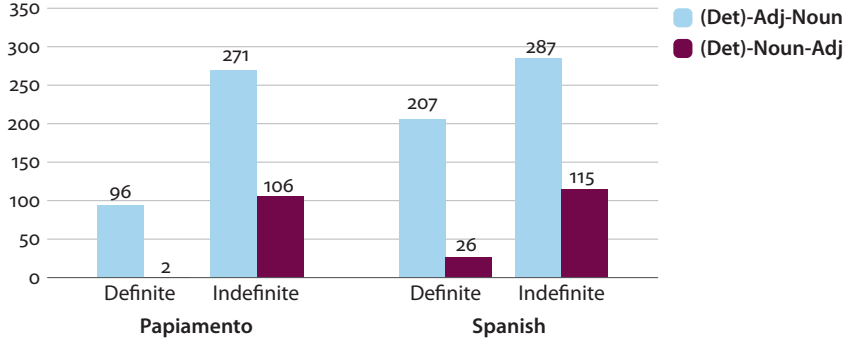


Figure 5. Prenominal and postnominal adjective constructions in code-switching mode by group and definiteness

8.2.1 *Extralinguistic variables*

To check whether any extralinguistic variables correlated with a specific type of strategy, we checked for each participant (1) how often each word order was used, and, for the prenominal adjective constructions (2) in how many cases common and default gender were assigned, and (3) how often the assigned gender matched that of the translation equivalent.

Based on this information, one or several of the following four strategies were deduced for each participant: (1) Common gender default (AN word order); (2) Neuter gender default (AN word order); (3) Analogical gender strategy (AN word order); (4) Postnominal predicative construction. A combination of two strategies occurred for instance if a participant used mostly common gender (indicating a common default strategy), but also sometimes neuter, particularly when this matched the gender of the translation equivalent (indicating the analogical gender strategy). The overview of the used gender strategies per participant is given in Appendix B.

Not surprisingly given the group results, the common default was the most frequently used strategy (43 speakers), followed by the analogical gender strategy (12 speakers), the postnominal predicative construction (9 speakers) and finally, the neuter default (5 speakers).

An exploratory analysis of the data from the present study largely confirms the patterns observed in Boers et al (2020). The common gender strategy is used more by those speakers who scored relatively low on gender accuracy in the Dutch unilingual mode and who use their HL relatively more often, both in terms of contact with relatives and friends and with respect to exposure to the HL through books, social media, TV and music. However, somewhat puzzling is the fact that this strategy also seems to be used more by speakers who have spent more time in the Netherlands.

As for the analogical gender strategy, we can only consider the data from the Turkish and Spanish HSs, as the Papiamento speakers did not make use of this strategy. Combining the Turkish and the Spanish data together clearly confirms the pattern reported in Boers et al. (2020) of this strategy being associated with a Dutch-dominant profile. For instance, out of the 12 speakers who used the analogical gender strategy as (one of) their gender assignment strategy/ies, 7 were born in the Netherlands, and had not spent any time in their home country; 4 others arrived before the age of 6. Moreover, the majority of subjects who applied the analogical gender strategy indicated using relatively more Dutch than their HL with their relatives and friends. They also indicated relatively less exposure to other input in the HL, and they also scored highest with gender accuracy in the Dutch unilingual mode.

There was some indication that the neuter default in prenominal adjective constructions also seemed to be related to being dominant in Dutch, as 4 out of 5 participants who adopted this strategy were born in the Netherlands and reported lower proficiency in the HL.

9. Discussion and conclusion

This study focused on gender assignment in Dutch by HSs of Spanish, Turkish and Papiamento in the Netherlands, with the main objective of exploring the role of CLI from the HL onto the societal language. We considered whether the linguistic properties of the HL affect gender assignment strategies in the unilingual Dutch mode as well as in code-switching. We moreover aimed to explore which extralinguistic factors best predict linguistic behaviour, both in unilingual and code-switching mode. We predicted a preference for common gender in all HS groups, both in the unilingual mode and in code-switching mode. We also expected that having a gender system in the HL would facilitate target-like gender assignment in Dutch, i.e. Spanish HSs would outperform Turkish and Papiamento HSs. We furthermore hypothesized that differences between the three HLs in terms of definiteness and noun-adjective word order would be reflected in the type of constructions produced by our participants. As for extra-linguistic variables, we expected age of arrival and amount of exposure and use of both languages to influence gender assignment in unilingual Dutch, and we hypothesized to find a relationship between language dominance and gender assignment strategies in code-switching, as was attested for the Spanish HSs in Boers et al. (2020).

In the Dutch unilingual mode, we found a clear overgeneralization of the common gender by all groups. This has been demonstrated before for simultaneous bilinguals and (early) child bilinguals acquiring Dutch gender (e.g. Hulk & Cornips, 2006; Blom et al., 2008; Unsworth et al., 2014). However, while previous work on this topic has focused on children up to age 11, our study, which included participants with an age range from 7 to 55, shows that non-target behaviour with gender assignment in Dutch can persist well into adulthood. Our results are also relevant to the field of HL studies, as they underline the importance of taking into account the other language, which is often ignored in the literature. The present study shows that not only the HL but also the societal dominant language may be vulnerable to differential acquisition and CLI as a result of reduced input and or use of the language.[10] HSs' linguistic systems are *bilingual* systems and thus, both languages should be considered, instead of focusing on only one of the languages.

10. Ideally, a monolingual control group should be included to verify if, and to what extent, these HSs differ from matched monolingually raised speakers. This aspect is important to consider in future research.

It is important to point out that there are other possible explanations for the non-target performance in Dutch demonstrated in this study, apart from, or in addition to effects related to bilingualism. Cornips (2008) mentions that the overextension of common gender in bilingual children may also be an identity marker for certain communities of ethnic minorities in the Netherlands. If this is the case, it may not be accurate to consider the linguistic behaviour of these speakers to be a consequence of the reduced input and exposure that are inherently part of the bilingual experience, but rather the adoption of a *different* variety of Dutch that is shared among young generations of HSs. Nevertheless, these bilingual speakers have undeniably been exposed to a vast amount of input of standard Dutch as well, in school, through the media, etc. It thus begs the question whether they may actually have two different varieties of Dutch at their disposal – standard Dutch and their particular ethnolect – and whether they might be able to switch between the two depending on the specific interlocutor or the context. In a similar vein, for the Papiamento speakers, it can be argued that they may have been exposed to a different variety of Dutch, given that Antillean Dutch sometimes exhibits the common gender determiner where European Dutch would use the neuter determiner, and it also frequently inflects the adjective in indefinite neuter DPs (Depuydt, 2010). With this in mind, one might wonder whether the Papiamento speakers in this study might be able to alternate between standard Dutch and Antillean Dutch and simply 'selected' Antillean Dutch during the task.

However, if the overextension of the common gender by the HSs in this study were a mere reflection of the adoption of a different variety, we might not expect to see the effects of age of arrival and usage of the HL that were attested in the analysis of the extralinguistic variables. If it is a matter of choice for these speakers to speak standard Dutch, we would not expect this choice to depend on how old they were when they arrived in the Netherlands, or how much they use Dutch in their everyday life. It thus seems reasonable to assume that the HSs' performance with gender in Dutch is at least in part a consequence of bilingualism and the reduction in use and exposure of Dutch that it entails.

A final factor that may have contributed to the observed overuse of the common determiner, is the fact that all the nouns targeted in this study were count nouns. As mentioned in Section 2, gender agreement of personal pronouns in Dutch is partly determined by the semantic properties of the noun (count vs. mass). Cornips et al. (2012) show similar effects for DP-internal agreement; both Dutch monolingual and Spanish-Dutch bilingual children correctly assign the neuter gender in Dutch more frequently with mass nouns than with count nouns (see also Roodenburg & Hulk, 2008). Given that our study only included count nouns, it is impossible to discern whether this issue applies to our data, but it is worth taking into account the effect of semantic properties of the noun in future research on gender assignment in Dutch.

In spite of the fact that all three HS groups overextended the common gender in Dutch, there was a difference between the groups with respect to the degree of overextension, that is, the Spanish speakers exhibited more target-like use of the neuter gender than the Papiamento and the Turkish speakers. This may well be related to the fact that Spanish has a two-way gender system, while the other two languages lack grammatical gender, suggesting CLI between the two languages in contact (cf. Eichler et al., 2013; Schwartz et al., 2015; Kalsta et al., 2019). In other words, the presence of a gender system in the HL may facilitate gender acquisition or at least gender awareness in the societally dominant language (cf. Egger et al., 2018; Hulk & van der Linden, 2010).

However, given the differences between the sociolinguistic profiles of the participants in each group, as described in Section 7.3., it is worth exploring the possibility that the advantage for the Spanish group might be due to something other than structural similarities between the gender systems in their two languages. For instance, Spanish-speaking participants were on average younger at the time of testing than the other two groups, and they reported a higher proficiency in the HL (at least as compared to Turkish speakers). It is unclear how these differences would benefit them in the acquisition of Dutch gender, though.

Two potentially relevant differences between the Spanish and the Papiamento HSs are the fact that the Spanish speakers reported a younger age of arrival than the Papiamento speakers, and the child participants in the Spanish group reported relatively less previous input in the HL (and thus more input in Dutch) compared to the children in the Papiamento group. However, based on these variables, one would not expect the Spanish HSs to outperform the Turkish ones seeing as the Turkish HSs reported an even earlier age of arrival than the Spanish HSs. What is more, considering other variables such as the reported length of residence in the Netherlands and the amount of use of the HL with non-immediate family (as reported for the children), the Spanish group would be expected to perform even less target-like than the Papiamento speakers, as they had spent significantly less time in the Netherlands and used Dutch significantly less frequently (at least outside of the home environment). It is therefore likely that the advantage of the Spanish speaking group can be explained by the fact that their HL exhibits a gender system.

A reviewer mentioned the possibility that the three communities in question might differ in terms of their socio-economic status (SES), which could, in turn, affect their linguistic behaviour. Although we have no reason to suspect substantial differences between the Spanish-, Papiamento- and Turkish-speaking communities in the Netherlands, we cannot verify that there are no such differences in our sample, since our background questionnaire did not specifically target information on SES. This is something that should be taken into account in future studies.

Another indication of CLI was the finding that in the unilingual mode, Papiamento and Turkish speakers used the indefinite determiner more often than Spanish speakers. This may result from an avoidance strategy that is applied whenever a speaker is uncertain about the gender of the noun. Given that for indefinite nouns, the same determiner ('*een*') is used for both common and neuter gender, the speaker may view it a safer option, even though gender marking is different for adjectives in indefinite common and neuter DPs.

Note that in the unilingual mode, there was also significant difference between the Turkish and the Papiamento: the Turkish HSs produced even more indefinites than the Papiamento HSs. This may be due to the fact that, unlike Papiamento, Turkish lacks a definite article (Kornfilt, 1997). Backus, Doğruöz, and Heine (2011) report that Turkish HSs in the Netherlands manifest an overextension of the usage context of *bir* in their Turkish using Dutch as a model. Based on our results, we can say that the Dutch-Turkish influence is reciprocal: one the one hand, Turkish HSs use *bir* more often than monolinguals in their Turkish due to the influence from Dutch; on the other hand, they use *een* frequently in their Dutch because its equivalent is available in their HL.

In code-switching mode, the most commonly adopted strategy by all three groups was the use of a common gender default. In addition, the Papiamento HSs and the Spanish HSs occasionally used a construction with a postnominal adjective, which is ungrammatical in Dutch. The Turkish HSs on the other hand, did not use this type of construction at all. This seems to be another clear manifestation of differential CLI from the HL, given that postnominal adjectives are the preferred option in Papiamento and Spanish, whereas they are ungrammatical in Turkish.[11]

These findings support the hypothesis that CLI is indeed an important factor in language contact and can also occur in the direction from the HL (the supposedly weaker language) to the societally dominant language. It is worth noting that these effects would have been hard to demonstrate convincingly by looking at only one language pair, which emphasizes the value of comparisons across language pairs.

Another difference between the groups concerned the reliance on the analogical gender strategy in code-switching. The statistical analysis revealed a three-step hierarchy between the three languages where Spanish HSs used this strategy the most, followed by the Turkish HSs, while the Papiamento HSs did not use the analogical gender strategy at all.

---

11. The Papiamento and Spanish speakers barely used this construction in the unilingual model (4 exceptions). However, quite puzzlingly, one Turkish-speaking participant used postnominal adjectives across the board in unilingual Dutch, but not in code-switching mode.

How can we explain these differences? One possibility is that the use of the analogical gender strategy is related to the level of competence in Dutch. After all, in order to be able to use the analogical gender strategy to assign gender to a HL word in Dutch, a speaker needs to possess accurate knowledge of Dutch gender. Thus, it is not surprising that the group that performed best with gender accuracy in the unilingual mode – the Spanish HSs – also made use of the analogical gender strategy the most. However, it was not the case that the Turkish group was more target-like than the Papiamento group with Dutch gender in unilingual mode – if anything, the latter group was more target-like (though not significantly).

How can we then explain the fact that the Turkish HSs adopted the analogical gender strategy more than the Papiamento HSs did? This may be related to the amount of Dutch language input during early childhood. If we compare the age of arrival for our three HL groups, we see that on average, the Papiamento speakers arrived at age 8.27, while the Turkish and Spanish speakers arrived in the Netherlands on average at ages 2.18 and 3.24 respectively. Even though the Papiamento speakers were also exposed to Dutch in their home country before arriving to the Netherlands, we can assume that this exposure was probably limited to formal registers, as Papiamento tends to be used much more than Dutch in everyday life at the Antilles (Severing & Verhoeven, 2001).

This also becomes evident when we compare the reported numbers for previous input between age 0–4: Papiamento-speaking children reported the highest percentage of use of the HL in early childhood (78.25%), compared to 75% for the Turkish-speaking children and 60.34% for the Spanish-speaking children, which seems to suggest that early input in the language that has gender is crucial for the option of the analogical gender strategy (cf. Munarriz-Ibarrola et al., 2021).

It is also in line with the results from the analysis on individual differences, which revealed a relationship between the use of the analogical gender strategy and age of arrival to the Netherlands. Other important predictors that were found in this analysis were the patterns of use of and exposure in both languages: the more Dutch was used with friends and family, and the more exposure to Dutch through books, social media, TV and music, the more the analogical gender strategy was used, suggesting that it may be a matter of dominance of Dutch more generally (cf. Boers et al., 2020), rather than just of early input. These results testify to the importance of considering extra-linguistic variables, both in terms of individual differences between speakers and at the level of the communities.

Alternatively (or in addition), the code-switching patterns in the communities may also play a role in determining the gender assignment strategies, as suggested by Króliwkowska et al. (2019), who found that frequent code-switching in the community corresponds to a higher use of the common default strategy. We did not ask our participants about their code-switching practices, but the literature shows that

code-switching is widespread in the Turkish-speaking community (Backus, 2011) as well as in the Papiamento-speaking community (Parafita Couto & Gullberg, 2019). For Spanish HSs in the Netherlands we do not have any information when it comes to code-switching habits. This issue is worth exploring in future research.

Acknowledgements

We would like to thank these students for their help at different stages of the project: Serenay Kaykaç (experiment design, data collection, transcription), Samantha Angela (data collection, transcription, coding), Eva Šipulová (data collection, transcription), Machteld van Kooten (coding), Tessa van der Meijden (coding).

References

Aalberse, S., & Weerman, F. (in preparation). *The feature [+count] in pronominal reference*. University of Amsterdam.

Aalberse, S., & Hulk, A. (2018). Introduction to the special issue–Heritage language studies and early child bilingualism research: Understanding the connection. *International Journal of Bilingualism*, 22, 491–582. https://doi.org/10.1177/1367006916654351

Aalberse, S., Backus, A., & Muysken, P. (2019). *Heritage languages: A language contact approach*. John Benjamins. https://doi.org/10.1075/sibil.58

Audring, J. (2006). Pronominal gender in spoken Dutch. *Journal of Germanic Linguistics*, 18(2), 85–116. https://doi.org/10.1017/S1470542706000043

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

Backus, A. (2011). Turkish as an immigrant language in Europe. In T. K. Bhatia & W. C. Ritchie (Eds), *The handbook of bilingualism and multilingualism* (2nd ed., pp. 770–790). Wiley-Blackwell.

Backus, A., Doğruöz, A. S., & Heine, B. (2011). Salient stages in contact-induced grammatical change: Evidence from synchronic vs. diachronic contact situations. *Language Sciences*, 33(5), 738–752. https://doi.org/10.1016/j.langsci.2011.02.004

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. C. (2021). Gender assignment strategies among simultaneous Spanish/English bilingual children from Miami, Florida. *Studies in Hispanic and Lusophone Linguistics*, 14.

Bellamy, K., Parafita Couto, M. C., & Stadthagen-Gonzalez, H. (2018). Investigating gender assignment strategies in mixed Purepecha–Spanish nominal constructions. *Languages*, 3(3), 28. https://doi.org/10.3390/languages3030028

Benmamoun, E., Montrul, S., & Polinsky, M. (2013). Heritage languages and their speakers: Opportunities and challenges for linguistics. *Theoretical linguistics*, 39(3–4), 129–181. https://doi.org/10.1515/tl-2013-0009

Berkum, J. J. van. (1996). The psycholinguistics of grammatical gender: Studies in language comprehension and production (PhD dissertation). Katholieke Universiteit Nijmegen. Nijmegen University Press.

Blom, E., Polišenská, D., & Weerman, F. (2008). Articles, adjectives and age of onset: The acquisition of Dutch grammatical gender. *Second Language Research*, 24(3), 297–331. https://doi.org/10.1177/0267658308090183

Boers, I., Sterken, B., Osch, B. van, Parafita Couto, M. C., Grijzenhout, J., & Tat, D. (2020). Gender in unilingual and mixed speech of Spanish heritage speakers in the Netherlands. *Languages*, 5(4), 68. https://doi.org/10.3390/languages5040068

Boumans, L. P. C. (1998). The syntax of code-switching: Analysing Moroccan Arabic/Dutch conversation (Unpublished PhD dissertation). Katholieke Universiteit Nijmegen, Tilburg.

Caroll, S. (1989). Second-language acquisition and the computational paradigm. *Language Learning*, 39(4), 535–594. https://doi.org/10.1111/j.1467-1770.1989.tb00902.x

Central Bureau of Statistics [CBS]. (2019, June 24). *Population; sex, age, migration background and generation*. Retrieved on 10 June from https://opendata.cbs.nl/statline/#/CBS/en/dataset/37325eng/table?ts=1614200262031

Cinque, G. (1994). On the evidence for partial N-movement in the Romance DP. In G. Cinque, J. Koster, J-Y. Pollock, L. Rizzi, & R. Zanuttini (Eds), *Paths towards Universal Grammar. Studies in honor of Richard S. Kayne* (pp. 85–110). Georgetown University Press.

Clyne, M. (1977). Nieuw Hollands or Double Dutch. *Dutch Studies*, 3, 1–20.

Clyne, M., & Pauwels, A. (2013). The Dutch language in Australia. In F. Hinskens & J. Taeldeman (Eds.), *Language and Space. An international handbook on linguistic variation Vol. 3: Dutch* (pp. 858–79). Mouton de Gruyter.

Cornips, L., & Hulk, A. (2006). External and internal factors in bilingual and bidialectal language development: Grammatical gender of the Dutch definite determiner. In C. Lefebvre, L. White, & C. Jourdan (Eds), *L2 acquisition and creole genesis: Dialogues*, (pp. 355–378). John Benjamins. https://doi.org/10.1075/lald.42.21cor

Cornips, L. (2008). Losing grammatical gender in Dutch: The result of bilingual acquisition and/or an act of identity?. *International Journal of Bilingualism*, 12(1–2), 105–124.

Cornips, L., & Hulk, A. (2008). Factors of success and failure in the acquisition of grammatical gender in Dutch. *Second Language Research*, 24(3), 267–295.

Cornips, L., Hulk, A., Reijers, C., & González, P. (2012). When semantics meets morpho-syntax. The role of count/mass feature in acquiring HET as definite determiner in monolingual and bilingual Dutch. *Estudos Linguísticos/Linguistic Studies*, 6(7), 75–89.

Depuydt, C. (2010). An emerging variety of Dutch in the Antilles: Determining Papiamento influence (Unpublished Master's thesis). Albert-Ludwigs-Universität Freiburg.

Dijkhoff, M. (2016). *Dikshonario / Woordenboek* (6th ed.). Walburg Pers.

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

Eichler, N., Jansen, V., & Müller, N. (2013). Gender acquisition in bilingual children: French–German, Italian–German, Spanish–German and Italian–French. *International Journal of Bilingualism*, 17(5), 550–572. https://doi.org/10.1177/1367006911435719

Gathercole, V. C. M., & Thomas, E. M. (2005). Minority language survival: Input factors influencing the acquisition of Welsh. In J. Cohen, K. McAlister, K. Rolstad, & J. MacSwan (Eds), *Proceedings of the 4th International Symposium on Bilingualism* (pp. 852–874). Cascadilla Press.

Geeraerts, D. (1992). Pronominale masculiniseringsparameters in Vlaanderen. In P. Cornelis, J. M. Paardekooper, H. Bennis, & J. W. de Vries (Eds), *De binnenbouw van het Nederlands: Een bundel artikelen voor Piet Paardekooper* (pp. 73–84). ICG Publications.

Göksel, A., & Kerslake, C. (2004). *Turkish: A comprehensive grammar*. Routledge. https://doi.org/10.4324/9780203340769

Gullberg, M., Indefrey, P., & Muysken, P. (2009). Research techniques for the study of code-switching. In B. E. Bullock & A. J. Toribio (Eds), *The Cambridge handbook of linguistic code-switching*, (pp. 21–39). Cambridge University Press. https://doi.org/10.1017/CBO9780511576331.003

Heusinger, K. von, & Kornfilt, J. (2005). The case of the direct object in Turkish: Semantics, syntax and morphology. *Turkic Languages*, 9, 3–44.

Houwer, A. De. (1990). *The acquisition of two languages from birth: A case study*. Cambridge University Press. https://doi.org/10.1017/CBO9780511519789

Hulk, A., & Cornips, C. (2006). Neuter gender and interface vulnerability in child L2/2L1 Dutch. In S. Unsworth, T. Parodi, A. Sorace, & M. Young-Scholten (Eds.), *Paths of development in L1 and L2 acquisition*, (pp. 107–134). John Benjamins. https://doi.org/10.1075/lald.39.06hul

Hulk, A., & van der Linden, E. (2010). How vulnerable is gender? In P. Guijarro-Fuentes & L. Domínguez (Eds), *New directions in language acquisition: Romance languages in the generative perspective* (pp. 107–132). Cambridge Scholars.

Irizarri van Suchtelen, P. (2016). Spanish as a heritage language in the Netherlands. A cognitive linguistic exploration (Doctoral dissertation). Radboud Universiteit Nijmegen. LOT.

Jacobs, B. (2012). *Origins of a creole. The history of Papiamentu and its African ties*. Mouton De Gruyter. https://doi.org/10.1515/9781614511076

Jake, J. L., Myers-Scotton, C. & Gross, S. (2002). Making a minimalist approach to codeswitching work: Adding the Matrix Language. *Bilingualism, Language and Cognition*, 5, 69–91. https://doi.org/10.1017/S1366728902000147

Kaltsa, M., Tsimpli, I. M., & Argyri, F. (2019). The development of gender assignment and agreement in English-Greek and German-Greek bilingual children. *Linguistic Approaches to Bilingualism*, 9(2), 253–288. https://doi.org/10.1075/lab.16033.kal

Kornfilt, J. (1997). *Turkish*. Routledge.

Kornfilt, J. (2017). DP versus NP: A cross-linguistic typology? In W. McClure & A. Vovin (Eds.), *Studies in Japanese and Korean historical and theoretical linguistics and beyond*, (pp. 138–158). Brill.

Kraaikamp, M. (2017). Semantic versus lexical gender. Synchronic and diachronic variation in Germanic gender agreement (PhD dissertation). Universiteit van Amsterdam. LOT.

Królikowska, M. A., Bierings, E., Beatty-Martínez, A. L., Navarro-Torres, C., Dussias, P. E., & Parafita Couto, M. C. (2019). Gender assignment strategies within the bilingual determiner phrase: four Spanish-English communities examined. Paper presented at the 3rd Conference on Bilingualism in the Hispanic and Lusophone World (BHL), Leiden, The Netherlands, 9–11 January.

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

Liceras, J. M., Fuertes, R. F., 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

Meisel, J. M. (2004). The bilingual child. In T. K. Bhatia & W. C. Ritchie (eds), *The handbook of bilingualism* (pp. 91–113). Blackwell.

Meisel, J. M. (2009). Second language acquisition in early childhood. *Zeitschrift für Sprachwissenschaft*, 28(1), 5–34. https://doi.org/10.1515/ZFSW.2009.002

Mitrofanova, N., Rodina, Y., Urek, O., & Westergaard, M. (2018). Bilinguals' sensitivity to grammatical gender cues in Russian: The role of cumulative input, proficiency, and dominance. *Frontiers in Psychology*, 9, 1894. https://doi.org/10.3389/fpsyg.2018.01894

Montrul, S. (2004). Subject and object expression in Spanish HSs: A case of morphosyntactic convergence. *Bilingualism*, 7(2), 125–142. https://doi.org/10.1017/S1366728904001464

Munarriz-Ibarrola, A., Ezeizabarrena, M. J., Castro Arrazola, V. de, & Parafita Couto, M. C. (2021). Gender assignment strategies and L1 effects in the elicited production of mixed Spanish-Basque DPs. *Linguistic Approaches to Bilingualism*. https://doi.org/10.1075/lab.20016.mun

Ortega, L. (2020). The study of heritage language development from a bilingualism and social justice perspective. *Language Learning*, 70, 15–53. https://doi.org/10.1111/lang.12347

Osch, B. A. van. (2019). Vulnerability in HSs of Spanish in the Netherlands: An interplay between language-internal and language-external factors (PhD dissertation). Universiteit van Amsterdam. LOT.

Otheguy, R., & Lapidus, N. (2003). An adaptive approach to noun gender in New York contact Spanish. In R. Cameron, L. López, & R. Núñez-Cedeño (Eds.), *A Romance perspective on language knowledge and use* (pp. 209–229). John Benjamins. https://doi.org/10.1075/cilt.238.17oth

Paolieri, D., Cubelli, R., Macizo, P., Bajo, T., Lotto, L., & Job, R. (2010). Grammatical gender processing in Italian and Spanish bilinguals. *Quarterly Journal of Experimental Psychology*, 63(8), 1631–1645. https://doi.org/10.1080/17470210903511210

Parafita Couto, M. C., & Gullberg, M. (2019). Code-switching within the noun phrase: Evidence from three corpora. *The International Journal of Bilingualism*, 23(2), 695–714. https://doi.org/10.1177/1367006917729543

Parafita Couto, M. C., Munarriz, A., Epelde, I., Deuchar, M., & Oyharçabal, B. (2015). Gender conflict resolution in Spanish-Basque mixed DPs. *Bilingualism: Language and Cognition*, 18(2), 304–323. https://doi.org/10.1017/S136672891400011X

R Development Core Team. (2017). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing.

Rodina, Y., Kupisch, T., Meir, N., Mitrofanova, N., Urek, O., & Westergaard, M. (2020). Internal and external factors in heritage language acquisition: Evidence from heritage Russian in Israel, Germany, Norway, Latvia and the United Kingdom. *Frontiers in Education*, 5, 20. https://doi.org/10.3389/feduc.2020.00020

Rodina, Y., & Westergaard, M. (2017). Grammatical gender in bilingual Norwegian-Russian acquisition: The role of input and transparency. *Bilingualism: Language and Cognition*, 20(1), 197–214. https://doi.org/10.1017/S1366728915000668

Roodenburg, J., & Hulk, A. (2008). Puzzles on grammatical gender. *Lingue e Linguaggio*, 7(1), 67–92.

Schwartz, M., Minkov, M., Dieser, E., Protassova, E., Moin, V., & Polinsky, M. (2015). Acquisition of Russian gender agreement by monolingual and bilingual children. *International Journal of Bilingualism*, 19(6), 726–752. https://doi.org/10.1177/1367006914544989

Severing, R., & Verhoeven, L. (2001). Bilingual narrative development in Papiamento and Dutch. In L. Verhoeven & S. Strömqvist (Eds), *Narrative development in a multilingual context* (pp. 255–276). John Benjamins. https://doi.org/10.1075/sibil.23.10sev

Szagun, G., Stumper, B., Sondag, N., & Franik, M. (2007). The acquisition of gender marking by young German-speaking children: Evidence for learning guided by phonological regularities. *Journal of Child Language*, 34(3), 445–471. https://doi.org/10.1017/S0305000906007951

Tat, D. (2010). APs as reduced relatives: The case of bir in (some) varieties of Turkic. In A. Simpson (Ed.), *Proceedings of the 7th Workshop on Altaic Formal Linguistics* (pp. 301–315). The MIT Press.

Treffers-Daller, J. (1993). *Mixing two languages: French-Dutch contact in a comparative perspective*. Mouton de Gruyter.

Unsworth, S., Argyri, F., Cornips, L., Hulk, A., Sorace, A., & Tsimpli, I. M. (2014). The role of age of onset and input in early child bilingualism in Greek and Dutch. *Applied Psycholinguistics*, 35(4), 765–805. https://doi.org/10.1017/S0142716412000574

Valdés Kroff, J. R. (2016). Mixed NPs in Spanish-English bilingual speech. In R. E. Guzzardo Tamargo, C. M. Mazak, & M. C. Parafita Couto (Eds.), *Spanish- English codeswitching in the Caribbean and the US* (pp. 281–300). John Benjamins. https://doi.org/10.1075/ihll.11.12val

Appendix A. Objects and colors used in the Director-Matcher task

Dutch gender	Spanish gender	Spanish canonicity	Dutch noun	Spanish noun	Turkish noun	Papiamento noun	Translation
COM	MASC	canonical	**hamer**	*martillo*	*çekiç*	*martin / martiu*	'hammer'
COM	MASC	canonical	**hoed**	*sombrero*	*şapka*	*sombré / pèchi*	'hat'
NEUT	MASC	canonical	**boek**	*libro*	*kitap*	*buki*	'book'
NEUT	MASC	canonical	**oog**	*ojo*	*göz*	*wowo*	'eye'
COM	FEM	canonical	**kaars**	*vela*	*mum*	*bela*	'candle'
COM	FEM	canonical	**fles**	*botella*	*şişe*	*bòter*	'bottle'
NEUT	FEM	canonical	**huis**	*casa*	*ev*	*kas*	'house'
NEUT	FEM	canonical	**bed**	*cama*	*yatak*	*kama*	'bed'
COM	MASC	non-canonical	**bank**	*sofá / sillón*	*koltuk*	*sofa / banki*	'couch'
COM	MASC	non-canonical	**kam**	*peine*	*tarak*	*peña*	'comb'
NEUT	MASC	non-canonical	**hart**	*corazón*	*kalp*	*kurason*	'heart'
NEUT	MASC	non-canonical	**spook**	*fantasma*	*hayalet*	*spoki / spiritu*	'ghost'
COM	FEM	non-canonical	**sleutel**	*llave*	*anahtar*	*yabi*	'key'
COM	FEM	non-canonical	**bloem**	*flor*	*çiçek*	*flor*	'flower'
NEUT	FEM	non-canonical	**kruis**	*cruz*	*çarpı*	*krus*	'cross'

Dutch	Spanish	Turkish	Papiamento	Translation
wit	*blanco / -a*	*beyaz*	*blanku*	'white'
zwart	*negro / -a*	*siyah*	*pretu*	'black'
rood	*rojo / -a*	*kırmızı*	*kòrá*	'red'
groen	*verde*	*yeşil*	*berdè*	'green'

Appendix B.

HL	Participant	Adjective – noun				Post-nominal Adjective	Main strategy
		Common gender		Neuter gender			
		Match	Mismatch	Match	Mismatch		
Papiamento	PA01	18	15	1	2		Common default
	PA02	21	16				Common default
	PA03	19	10				Common default
	PA04	17	23				Common default
	PA05	15	10	1	1		Common default
	PA06	16	13	1			Common default
	PA08	16	14				Common default
	PA09	12	13				Common default
	PA10	7	5	9	9		Neuter default
	PC01		2			25	N-Adj
	PC04	1				29	N-Adj
	PC05	15	12				Common default
	PC06					27	N-Adj
	PT01					24	N-Adj
	PT02	27	18				Common default
Spanish	SA01	2				28	N-Adj
	SA02	13	16				Common default
	SA03	3	1			25	N-Adj
	SA04	14	15				Common default
	SA05	10	7			12	Common default & N-Adj
	SA06	18	11	7			Common default & AGS
	SA07	15	14				Common default
	SA08	21	10	5	1		Common default & AGS
	SC01	10	3	3		3	Common default & AGS
	SC02		3				–
	SC03	15	1	13	1		AGS
	SC04	24	25			2	Common default
	SC05	12	13			2	Common default

HL	Participant	Adjective – noun				Post-nominal Adjective	Main strategy
		Common gender		Neuter gender			
		Match	Mismatch	Match	Mismatch		
Spanish	SC06	16	6	7	1		Common default & AGS
	SC07	4	1	11	11		Neuter default
	SC08	4	5	5	1	15	Common default & N-Adj
	ST01	8	8		1	6	Common default
	ST02	16	8	7			Common default & AGS
	ST03	9	2	11	8		Neuter default & AGS
	ST04					23	N-Adj
	ST05	15	14			1	Common default
Turkish	TA01	15	12	1			Common default
	TA02	17	12	1			Common default
	TA03	15	11	3			Common default
	TA04	17	13				Common default
	TA05	16	14				Common default
	TA06	14	9	4	2		Common default & AGS
	TA07	15	15				Common default
	TA08	16	14				Common default
	TA09	17	10	3			Common default & AGS
	TA10	16	14				Common default
	TA11	16	14				Common default
	TA12	6	4	11	9		Neuter default (& AGS)
	TA13	28	25	2	1		Common default
	TA14	16	4	10			N-Adj
	TA15	14	16	1			Common default
	TA16	16	13				Common default
	TC01	12	8				Common default
	TC02			14	15		Neuter Default
	TC04	12	9	5	3		Common default & AGS
	TC06	17	13				Common default
	TT01	15	14				Common default
	TT02	16	13				Common default

CHAPTER 10

Prediction on the basis of gender and number in Mandarin-Italian bilingual children

Jasmijn E. Bosch[1], Mathilde Chailleux[1], Jia'en Yee[2], Maria Teresa Guasti[1] and Fabrizio Arosio[1]
[1]University of Milan-Bicocca / [2]University Putra Malaysia

The present study used a visual world eye-tracking paradigm to investigate online processing of grammatical gender and number in Mandarin-Italian bilingual children, in comparison to monolingual Italian children. We examined how children anticipated upcoming nouns on the basis of grammatical gender and number information on the preceding article. While monolingual speakers are able to employ such predictive mechanisms from a very young age, to our knowledge, this is the first study that compares gender and number processing in bilingual children. The results show that, overall, participants made linguistic predictions on the basis of articles, although a post-hoc analysis focusing on a subset of our bilingual participants did not confirm the prediction effect in Mandarin-Italian bilingual children. We found a greater difference in the bilingual group than in the monolingual group, in that bilinguals tended to be slower when processing gender than number. Finally, we found that L2 proficiency had a significant effect on gender processing in the bilingual group. One interpretation of these findings is that the discrepancy between gender and number may be due to transfer, since Mandarin does not have grammatical gender while it does have a conceptual notion of number. Another factor may be L2 proficiency, and especially lexical knowledge, since gender is an arbitrary property stored in the lexicon, while number is concretely linked to the referential context.

Keywords: visual world paradigm, gender, number, child bilingualism, predictive processing

1. Introduction

Humans are able to make predictions while processing speech, making language comprehension fast and efficient. In a classic language processing study, Altmann and Kamide (1999) found that listeners who heard the sentence *The boy eats the*

https://doi.org/10.1075/sibil.63.10bos

cake inferred when hearing the word *eats* that the upcoming noun must denote an edible object. In other words, listeners anticipated how the sentence would continue based on the lexical meaning of the verb. Similarly, a native speaker of Italian who hears the sentence *Prendi la banana* 'Take the.FEM.SING banana' could anticipate certain features of the upcoming noun on the basis of morphosyntactic information on the article. Firstly, the gender feature indicates that the following noun must be feminine. Secondly, the number marking of the article indicates that the noun should refer to a single object rather than two or more. In contrast, when hearing the sentence *Prendi i mandarini* 'Take the-MASC-PLU mandarins-MASC-PLU', Italian listeners can predict, as soon as they hear the article, that the following noun will refer to more than one object with masculine gender.

While previous research has found that monolinguals make rapid use of determiner features during noun recognition from an early age (e.g. Lew-Williams & Fernald, 2007; Robertson, Shi, & Melançon, 2012), adult second language (L2) learners appear to be less efficient when processing gender (e.g. Foucart & Frenck-Mestre, 2011; Lew-Williams & Fernald, 2010). We turned our attention to bilingual children who started acquiring their L2 at an early age, by investigating linguistic prediction based on gender and number in Mandarin-Italian bilinguals, in comparison to monolingual Italian children. Mandarin is relevant for this purpose since it has neither articles nor grammatical gender, while it does express number, albeit in a different way than in Italian. This provides us with an interesting test case for investigating the effect of linguistic differences on language processing in early L2 learners.

Italian has a binary gender system, in which nouns can be classified as feminine or masculine. Gender assignment is largely arbitrary, although the gender of animate nouns tends to match the biological sex of their referents (e.g. *uomo*-MASC 'man' and *donna*-FEM 'woman'). Furthermore, there are some morphophonological regularities associated with the gender classification of Italian nouns. Most feminine singular nouns end with the phoneme /a/ (e.g. *matita*-FEM 'pencil') while most masculine singular nouns end with the phoneme /o/ (e.g. *libro*-MASC 'book'), but there are also exceptions to this rule (e.g. *moto*-FEM 'motorbike'; *poema*-MASC 'poem'). Additionally, Italian has a subclass of nouns ending in /e/, which may be either masculine or feminine (e.g. *noce*-FEM 'nut'; *sole*-MASC 'sun'), as well as loan words ending in consonants that tend to be masculine (e.g. *computer*-MASC 'computer') (Ayoun & Maranzana, 2020; Chini, 1995; Padovani & Cacciari, 2003; Panzini, 2017). The present study only includes nouns with prototypical, morphologically transparent endings, i.e. masculine nouns ending in -/o/ or feminine nouns ending in -/a/.

In Italian, number is expressed on nouns through suffixes (syncretic with gender marking), such that the vocalic endings of nouns reveal information about

both gender and number. Typically, the plural form of regular masculine nouns ends in *-i* (e.g. *libro*-MASC-SING 'book' vs *libri*-MASC-PLU 'books') and the plural form of regular feminine nouns ends in *-e* (e.g. *matita*-FEM-SING 'pencil' vs *matite*-FEM-PLU 't-shirts'). The plural form of nouns ending in *-e* typically ends in *-i,* regardless of grammatical gender (e.g. *noci*-FEM 'nuts'; *soli*-MASC 'suns'). It should be noted, however, that the system is also characterized by exceptions; some nouns do not mark plurality through suffixation (e.g. *computer-* MASC-SING 'computer' vs *computer*-MASC-PLU 'computers'; *città*-FEM-SING 'city' vs *città*-FEM-PLU 'cities'), and other nouns change their grammatical gender according to number (e.g., *dito*-MASC-SING 'finger', *dita*-FEM-PLU 'fingers') (Ferrari & Zampese, 2016; Panzini, 2017).

The gender and number of nouns can typically be determined by looking at the preceding article. Italian has definite and indefinite articles that are unambiguously marked for gender and number; singular masculine nouns are typically preceded by *il* (definite) or *un* (indefinite), singular feminine nouns by *la* (definite) or *una* (indefinite), plural masculine nouns by *i,* and plural feminine nouns by *le.* Nouns that starts with a vowel are an exception, since they require the use of the elided article *'l'*, which can either be followed by a feminine or a masculine noun (e.g. *l'amico*-MASC and *l'amica*-FEM 'the friend'). For a small group of masculine nouns, the articles *lo, uno* or *gli* are used, based on phonological rules (e.g. nouns starting with *z; lo zaino* 'the backpack'; *uno zaino* 'a backpack'; *gli zaini* 'the backpacks') (Ayoun & Maranzana, 2020; Ferrari & Zampese, 2016; Panzini, 2017).

In addition to determiners, gender and number of nouns are also marked through agreement with personal pronouns, adjectives, and past participles (Panzini, 2017). The example in (1) shows the typical agreement pattern with a masculine singular noun (e.g. *vaso* 'vase'), which includes the masculine singular article *il* and the vocalic ending [o] for the personal pronoun *mio* 'my', the adjective *rosso* 'red' and the past participle *stato rotto* 'is broken'. In the case of a feminine singular noun like *tazza* 'cup' in (2), the corresponding article *la* and the vocalic ending [a] are typically used. The agreement with masculine plural nouns, such as *vasi* 'vases' in (3), typically involves the article *i,* as well as the vocalic ending [i] for the other elements of the sentence (*miei* 'my', *rossi* 'red', *stati rotti* 'are broken'). Finally, the typical agreement pattern for feminine plural nouns such as *tazze 'cups'* is shown in (4), in which the article *le* and the vocalic ending [e] are used.

(1) *Il* *mio* *vaso* *rosso* *è*
the-MASC-SING my-MASC-SING vase-MASC-SING red-MASC-SING is
stato *rotto*
been-MASC-SING broken-MASC-SING
'My red vase has been broken'

(2) *La mia tazza rossa è stata*
the-FEM-SING my-FEM-SING cup-FEM-SING red-FEM-SING is been-FEM-SING
rotta
broken-FEM-SING
'My red cup has been broken'

(3) *I miei vasi rossi sono*
the-MASC-PLU my-MASC-PLU vase-MASC-PLU red-MASC-PLU are
stati rotti
been-MASC-PLU broken-MASC-PLU
'My red vases have been broken'

(4) Le mie tazze rosse sono state
the-FEM-PLU my-FEM-PLU cup-FEM-PLU red-FEM-PLU are been-FEM-PLU
rotte
broken-FEM-PLU
'My red cups have been broken'

Mandarin presents several contrasts with Italian. First, it lacks articles, as bare nouns may either have a generic or existential meaning, or a definite interpretation. Second, Mandarin does not have grammatical gender, but it uses nominal classifiers. These lexical items, which indicate to which class a noun belongs, appear obligatorily with numerals or quantifiers (Frankowsky & Ke, 2016). It has been argued that in some aspects classifier systems may not be qualitatively different from grammatical gender systems, as both systems entail nominal classification (Fedden & Corbett, 2017; Haspelmath, 2018). Unlike grammatical gender in Italian, however, classifiers in Mandarin do not involve morphosyntactic agreement. Nominal classification in Mandarin is based on semantic properties of the noun, such as its material and physical shape (Huang & Ahrens, 2003).

Number is unspecified in bare nouns in Mandarin, that is, they are not singular nor plural, but they have general number (Corbett, 2000; Rullmann & You, 2006). There are, however, several ways of expressing the concept of number or quantity, including using numerals such as *yī* 'one' or *liǎng* 'two', or quantifiers such as *duì* (denoting a pair) or *tào* (denoting a set). Two frequently discussed plural-encoding forms are *xīe* and *-mén* (Iljic, 1994). While *xīe* 'some' marks collectiveness, it is not obligatory, and it can also be used with mass nouns. The suffix *-mén* is obligatorily used with pronouns to denote number, but it is restricted to human and definite referents. Moreover, it has been argued that plurality may be expressed by the reduplication of classifiers, and singularity may be expressed by using a simple classifier in the absence of a numeral (Zhang, 2014).

To sum up, while the grammatical notion of gender is absent in Mandarin, number can be expressed in various ways, albeit differently than in Italian. This allows us to test the role of L1 characteristics in L2 processing in bilingual children.

This chapter will first discuss previous research on processing of grammatical gender and number, in both monolingual and bilingual adults and children. Then we will turn to our hypotheses, methods and results, before concluding with a discussion of our findings.

2. Literature review

2.1 Processing of grammatical gender

Gender processing occurs rapidly in native speakers. Evidence for this comes from studies using the visual world paradigm, in which participants' eye movements are recorded when presented with a visual scenario, while they are listening to sentences (Lew-Williams & Fernald, 2010; Dahan et al., 2000). For example, Lew-Williams and Fernald (2010) presented participants with two pictures of objects that either matched or mismatched in grammatical gender, instructing them to look for an object displayed on the screen (i.e. *Encuentra el/la …* 'Find the-MASC/the-FEM'). Monolingual Spanish speakers showed faster identification of the target referent when there was a mismatch between target and competitor, allowing the target noun to be identified based on the gender-marked article (Lew-Williams & Fernald, 2010). Using a similar paradigm, Dahan et al. (2000) found that French speakers were able to pre-activate the upcoming noun based on the gender-marked article. The results of these eye-tracking studies are supported by neurophysiological evidence showing that native speakers attend to gender cues when processing sentences, and that they are able to use this information to anticipate upcoming nouns (Wicha, Moreno, & Kutas, 2004; Barber & Carreiras, 2005).

Monolingual children also make use of grammatical gender cues during online processing. Recent eye-tracking studies have demonstrated that, similarly to adults, German eight- and nine-year-old children can predict upcoming nouns on the basis of grammatical gender marking on determiners and adjectives (Cholewa et al., 2019; Lemmerth & Hopp, 2019). Furthermore, three-year-old monolingual Spanish-speaking children have been shown to effectively exploit gender information on articles, leading to faster recognition of subsequent nouns (Lew-Williams & Fernald, 2007). Using a preferential looking paradigm, Van Heugten & Shi (2009) reported a similar effect in even younger children; French-learning two-year-olds were faster to turn their gaze to a target referent when it was preceded by a disambiguating gender-marked article. Moreover, when a noun was preceded by an incongruent article, this disagreement in grammatical gender subsequently hindered comprehension.

In contrast to native speakers, adult L2 learners tend to be less efficient when processing grammatical gender. For example, using an auditory naming task,

Guillelmon and Grosjean (2001) tested whether native English late L2 learners of French, early English-French bilinguals and French monolinguals experienced a facilitation effect when repeating nouns that were preceded by a grammatical gender-marked article, and an inhibition effect when there was a mismatch in grammatical gender. Their results showed clear facilitation and inhibition effects for both monolinguals and early bilinguals, resulting in faster reaction times when the appropriate gender was used and slower reaction times when there was a gender mismatch, while late bilinguals were completely insensitive to gender marking.

Furthermore, predictive processing may be slower in the L2, even in highly proficient speakers whose offline comprehension is on target. For example, Gruter, Lew-Williams and Fernald (2012) investigated offline comprehension, elicited production, and online processing of gender in advanced L2 speakers of Spanish (L1 English) as well as L1 Spanish speakers. Even though the L2 participants performed at-ceiling in a picture selection task, they made some errors in production, and they made significantly less use of gender cues when processing familiar nouns.

Focusing on L1 English advanced to near-native L2 speakers of German, Hopp (2012) found that predictive processing based on gender-marked articles was significantly slower in L2 speakers with variable performance in a gender assignment production task than in native speakers and L2 speakers who consistently marked grammatical gender correctly. When processing gender-marked adjectives, L2 speakers with consistent high accuracy were significantly slower than native speakers, while L2 speakers with variable accuracy did not show any predictive processing.

As for L2 learners with lower levels of proficiency, Lew-Williams and Fernald (2010) found that in contrast to monolingual adults and children, intermediate L2 learners of Spanish failed to make use of gender cues on the article during online processing. In order to test whether this may have been due to differences in frequency of exposure, follow-up experiments were conducted with novel words, which were taught to the participants before the experimental phase. The results showed that after a training, L2 participants were able to make use of the gender-marked article, albeit to a lesser extent than L1 participants. However, when participants were trained with indefinite articles and tested with definite articles, L2 speakers were no longer able to do so, indicating that they could not exploit gender cues on articles when the task required them to generalize across forms.

In addition to factors related to the language proficiency, the typological distance between two languages and the degree of phonological overlap between translation equivalents may contribute to the observed differences between monolingual and bilingual processing of grammatical gender. Dussias et al. (2013) explored the effect of language distance on Spanish L2 processing by comparing native speakers of English (a language lacking grammatical gender) to native speakers of Italian (a

language which has a grammatical gender system comparable to Spanish), while taking into account different levels of proficiency. The results showed that, like Spanish monolingual speakers, native English high-proficiency L2 learners exhibited predictive processing on the basis of gender-marked articles, while native English low-proficiency L2 learners failed to do so. Interestingly, native Italian low-proficiency L2 learners of Spanish showed anticipatory eye movements, but only with the feminine article *la,* which is identical to its Italian counterpart. This suggests that low-proficiency participants experienced positive transfer when there was overlap between their L1 and their L2. Another example of transfer in bilingual gender processing is provided by the 'gender congruency effect', which refers to the often-replicated finding that bilinguals who speak two gender-marked languages are faster when processing nouns which have the same grammatical gender in both languages, as compared to nouns with incongruent gender (Morales et al, 2016; Sá-Leite, Fraga, & Comesaña, 2019).

At the neurophysiological level, some studies have found that L2 speakers respond differently to violations in grammatical gender agreement than native speakers, depending on the morphosyntactic realization and the degree of syntactic overlap between the L1 and the L2 (Foucart & Frenck-Mestre, 2011, 2012). For example, Foucart and Frenck-Mestre (2011) found that native German advanced L2 learners of French showed a native-like P600 effect in response to grammatical gender violations when agreement rules were similar in both languages (i.e. between determiners and nouns), while there was no such effect when agreement rules differed across languages (i.e. between adjectives and nouns). However, Tokowicz and MacWhinney (2005) found that native English low-proficiency L2 learners of Spanish were sensitive to agreement violations with grammatical gender, despite the fact that this feature was absent in the L1.

Whilst there is extensive (but yet inconclusive) literature on the differences between online processing of grammatical gender in native speakers versus adult L2 learners, online processing of grammatical gender in bilingual children remains largely unexplored. A recent study investigated linguistic prediction based on gender-marked determiners and adjectives in German, by comparing eight- and nine-year-old German-Russian bilingual children to a group of age-matched monolingual controls in a visual world experiment (Lemmerth & Hopp, 2019). The bilingual children were divided in two groups; simultaneous bilinguals, who were exposed to both languages from birth, and early sequential bilinguals, who were exposed to German before the age of four. The results showed that, similarly to monolingual children, simultaneous bilinguals anticipated upcoming nouns on the basis of the grammatical gender of the article or prenominal adjective. In contrast, sequential bilinguals only did so when the grammatical gender of the target noun was the same as that of its Russian translation equivalent. This means that

native Russian early L2 learners of German experienced a gender congruency effect, i.e. their processing was delayed when there was a mismatch in grammatical gender between their two languages, due to cross-linguistic interference during the pre-activation of nouns based on articles and adjectives.

Russian and German both have a tripartite grammatical gender system (distinguishing between feminine, masculine and neuter gender), but they differ in the way in which gender is expressed. While both languages mark grammatical gender on prenominal adjectives, German also marks gender on articles, which are absent in Russian. This allowed the authors to explore the effect of linguistic differences, by testing whether participants would be more likely to show prediction based on adjectives than on articles.

A previous study on native Russian adult L2 learners of German (Hopp & Lemmerth, 2018) found that high-intermediate speakers showed predictive processing in German based on gender-marked adjectives but not on gender-marked articles, suggesting that cross-linguistic overlap in the morphosyntactic realization of grammatical gender facilitates processing in L2 learners. However, Lemmerth and Hopp (2019) observed identical patterns for both types of stimuli, which means that the delayed processing of grammatical gender in sequential bilingual children cannot be attributed to the absence of articles in Russian.

Using a similar paradigm, Bosch and Foppolo (submitted) investigated predictive processing in Italian-German bilingual children between the ages of six and nine. Participants were residing in Italy or in Germany, and the sample included both simultaneous and sequential bilinguals with varying levels of language dominance. The results showed rapid prediction based on gender-marked articles in an Italian experiment (which was conducted in both countries) and in a German experiment (which was only conducted in Germany). Although in the Italian experiment prediction was significantly delayed in the incongruent gender condition (in which there was a mismatch between German and Italian gender), participants were still able to anticipate the target referent after hearing the disambiguating article. In other words, while children were influenced by their knowledge of German when processing sentences in Italian, they exhibited prediction regardless of gender incongruency. Crucially, the efficiency of linguistic prediction was related to relative language proficiency, i.e. Italian-dominant children showed more anticipation and less interference from German in an Italian task than German-dominant children. On the other hand, no gender congruency effect was found in the German experiment, in which mostly German-dominant children participated.

To summarize, the limited empirical evidence that is available suggests that bilingual children, like monolingual children and adults, are able to process grammatical gender in a fast and efficient manner (at least if the two languages have a gender system). Although language dominance appears to be an important factor,

it still remains unclear how child L2 learners with lower proficiency may differ from simultaneous bilinguals, and to what extent bilingual children may be influenced by the linguistic properties of their L1.

2.2 Processing of number

The concept of number appears very early in life, as shown by studies reporting prelinguistic understanding of number using manual-search tasks (Feigenson & Carey, 2005; Li, Ogura, Barner, Yang, & Carey, 2009). In Li et al. (2009) infants were able to discriminate a "one" condition, in which a single object was presented, from a "more-than-one" condition, in which infants saw several objects of the same type. From the age of two, children are sensitive to the morphological marking of number (Corrêa, Augusto, & Ferrari-Neto, 2005) and they are able to distinguish singular from plural marking. Using a preferential looking paradigm, Kouider et al. (2006) found that English-speaking children were sensitive to double number marking on the verb and the article at the age of 24 months. They showed participants two pictures, one depicting a single novel object and one depicting several identical novel objects. When hearing 'There are some blinkets', 24-month-olds looked significantly longer to the target picture depicting several objects. In a second experiment, they found that only 36-month-olds (but not 24-month-olds) were sensitive to the singular/plural distinction when only the phonological marking of the noun was presented ('Look at the blinkets'). With the same type of paradigm, Robertson and colleagues (2012) investigated processing of number information carried by the determiner in French-speaking two-year-olds. To do so, they manipulated the degree of information given by the determiner. In the uninformative condition, target and competitor pictures were matched in number so that the number marking of the determiner was not disambiguating. In the informative condition, target and competitor pictures mismatched in number so that children could rely on the number features of the determiner to anticipate the noun. Their results showed that toddlers looked longer at the target in an informative condition (in which the target could be identified by relying on the number marking on the determiner) than in an uninformative condition (in which the determiner did not reveal any identifying information about the target). Although participants did not preselect the noun based on the article, they did show a facilitation effect of number agreement in processing the following noun.

Previous studies investigating agreement effects in monolingual adults consistently found a facilitation effect of number on subsequent language processing, both within the noun phrase and across phrases (i.e. between subjects and verbs). Within the nominal phrase, the number facilitation effect has been evidenced in lexical decision tasks using a grammatical priming paradigm, in which

participants are presented with two dependent words (typically Determiner-Noun or Noun-Adjective) that either match or mismatch in number, while they have to perform a lexical decision task on the second element. Lukatela, Carello, and Turvey (1987) reported an effect of number congruency between the possessive adjective and the noun in Serbo-Croatian speakers. Noun recognition was faster in congruent trials than in incongruent ones, suggesting that participants may have pre-activated the number feature of the noun on the basis of the preceding article, or that they experienced a facilitation effect of number agreement during subsequent processing. Similar results were found for French and Spanish (Cole & Segui, 1994; Faussart, Jakubowicz, & Costes, 1999). Number agreement effects have also been evidenced in neuroimaging and electrophysiological studies, reporting a sensitivity to number agreement violations (Barber & Carreiras, 2003, 2005; Carreiras, Carr, Barber, & Hernandez, 2010). ERP data have revealed that number is integrated both at the syntactic level and at the semantic level. Barber and Carreiras (2003, 2005) found that number agreement violations elicit a Left Anterior Negativity (LAN) as well as a P600, which respectively correspond to early morphosyntactic integration and syntactic reanalysis (Friederici, 2002). They also found an N400, which is generally associated with semantic integration (ibid).

Considering L2 learners, the acquisition of number agreement generally appears to be rather unproblematic. Looking at production data, White and colleagues (2004) investigated number accuracy in late L2 learners of Spanish (L1 French and English) with varying proficiency levels. All groups performed at ceiling when computing number agreement between the determiner and the noun, with a minimum accuracy of 94.4% for the French low-proficiency group. When producing the Determiner-Noun-Adjective sequence, minimum accuracy was still 83.5%, showing that number agreement is acquired even at low proficiency levels.

Moreover, different studies have shown that adult L2 learners of Spanish (Lew-Williams & Fernald, 2009) and German (Hopp, 2012) are able to use number-marking on articles to rapidly distinguish between single versus multiple referents. In other words, similarly to native speakers, late L2 learners seem to be able to anticipate nouns based on the number feature of the determiner. Studies using electrophysiological methods also provide evidence for native-like processing of number in L2 speakers, at least at the higher levels of proficiency (Gabriele, Fiorentino, Bañón, 2013; Gillon-Dowens et al., 2010; Rossi et al., 2006). For example, Gabriele et al. (2013) tested English-speaking late learners of Spanish with a grammaticality judgment task addressing number violations between the noun and the adjective (*El banco es un edificio muy *siguros* 'The bank is a building-MASC-SING very safe-MASC-PLUR') while ERP responses were recorded. Results showed that participants who were proficient in Spanish displayed a native-like pattern both in the behavioural data (97% accuracy) and in the elicitation of the P600. Using Barber &

Carreiras (2005)'s material, Gillon-Dowens et al. (2010) found a native-like biphasic LAN-P600 pattern in native English advanced L2 learners of Spanish. Consistent with this finding, Rossi et al. (2006) found a similar pattern for high-proficiency L2 learners of Italian and German.

While late learners can reach native-like processing of number, there seem to be two major modulating factors. First, effects of L1/L2 similarity have been reported. To investigate the effect of L1 characteristics on L2 processing, Gillon-Dowens et al. (2011) replicated their earlier study with native English learners of Spanish with Chinese learners of Spanish. This time, number violations elicited only the P600, showing no LAN. The authors argue that the different ERP patterns between the two groups may be due to L1 effects on the L2: contrary to English and Spanish, Chinese does not mark number through morphosyntactic agreement. Thus, the absence of the number feature in the L1 may affect the early syntactic processing of number violations.

A second factor that may modulate number processing is L2 proficiency, as studies focusing on beginner L2 learners have reported non-nativelike ERP patterns (Gabriele et al., 2013; Tokowicz & MacWhinney, 2005). For example, Gabriele et al. (2013) investigated number processing in native English late learners of Spanish with varying levels of proficiency (low, intermediate and advanced). While advanced and intermediate learners exhibited a P600 response, only a marginally significant P600 effect was found in low-proficiency learners. Another study testing native English low-proficiency learners of Spanish reported no P600 response to number violations at all (Tokowicz & MacWhinney, 2005).

From very early in life, number marking is used as a cue for predicting upcoming speech, thus facilitating language processing. Late L2 learners can reach native-like abilities, as shown in both behavioural and electrophysiological data. Yet, performance might be modulated by the degree of similarity between the L1 and the L2, as well as by L2 proficiency. Until now, the L2 literature has mostly focused on late L2 learners and their ability to acquire new linguistic features. However, it remains unknown how early bilingual children process number, and whether the L1 affects number processing in early L2 learners.

2.3 The discrepancy between gender and number

Although advanced L2 learners are able to reach native-like proficiency in the production, comprehension and online processing of both grammatical gender and number (White, Valenzuela, Kozlowska-Macgregor, & Leung, 2014; Gabriele et al., 2013; Gillon-Dowens, Vergara, Barber, & Carreiras, 2010; Gillon-Dowens, Guo, Guo, Barber, & Carreiras, 2011), grammatical gender appears to be more challenging to acquire than number.

Several studies have directly compared number processing to gender processing in L2 populations, but the results are mixed. For example, Tokowicz and MacWhinney (2005) investigated the production and online processing of grammatical gender and number agreement in native English late L2 learners of Spanish. They found that even though participants were less accurate with grammatical gender than with number in a grammaticality judgement task, their ERP patterns showed sensitivity to agreement violations with gender but not with number. The authors argued that their participants were implicitly sensitive to morphosyntactic agreement with a feature that is unique to their L2 (i.e. gender), but not with a feature that exists in both the L1 and the L2 but which is expressed differently in the two languages (i.e. number).

In contrast, other studies found an advantage for number over gender in L2 learners, both in behavioral tasks (White et al., 2004) and in online language processing (Gabriele et al., 2013; Gillon-Dowens et al., 2010; Lew-Williams & Fernald, 2009; Hopp, 2012). For example, capitalizing on the fact that articles in Spanish are marked for both grammatical gender and number, Lew-Williams and Fernald (2009) found that native English late L2 learners of Spanish made rapid use of number cues when identifying subsequent nouns, while they were not able to rely on grammatical gender cues unless the same pairs of articles and nouns were used throughout the experiment. Similarly, studying native English high-proficient L2 speakers of German, Hopp (2012) found more variable performance with gender processing than with number processing. While L2 speakers reliably anticipated nouns on the basis of the number marking on articles and adjectives, participants with sub-optimal performance in an offline gender assignment task were less likely to anticipate nouns on the basis of grammatical gender.

Moreover, several studies have compared ERP responses to gender and number agreement violations in bilingual adults. Testing native English late learners of Spanish, Gabriele et al. (2013) found no difference between gender and number for low-proficiency speakers, who showed only a marginal P600 effect for both gender and number agreement violations. The dissociation between number and gender appeared in the intermediate learners, who showed native-like ERP patterns for number but not for gender. In advanced learners, the advantage for number over gender was only quantitative, in that the P600 was significantly weaker when processing gender as compared to number agreement violations. A similar pattern was observed in the native English high-proficiency L2 learners of Spanish tested by Gillon-Dowens and colleagues (2010), who showed a delayed LAN and a weaker P600 when processing agreement violations with grammatical gender but not with number.

Thus, although there is some conflicting evidence in the literature, most studies suggest that grammatical gender develops more slowly than number in L2 learners

(White et al., 2004), which may lead to more non-nativelike processing of gender as compared to number, in particular at the lower levels of language proficiency (Gabriele et al., 2013).

This dissociation between gender and number has also been evidenced in the monolingual population, both in adults (Antón-Méndez, Nicol, & Garrett, 2002; Barber & Carreiras, 2003, 2005; Carreiras, 2010; Igoa, García-Albea, & Sánchez-Casas, 1999; Lukatela et al., 1987) and children (Dispaldro, Ruggiero, & Scali, 2015). Monolinguals tend to produce more gender errors than number errors (Antón-Méndez et al., 2002) and are slower to process gender than number (Lukatela et al., 1987). Barber and Carreiras's (2003, 2005) findings suggest that although early processes seem to be similar, differences arise in later processing. More specifically, they found that the second phase of the P600 response to agreement violations was larger for gender than for number. They argue that this difference in the ERP pattern reflects more complex reanalysis processes for gender than for number. The dissociation between the two features is also supported by neuroimaging data that showed that the processing of gender and number activate different areas of the brain (Carreiras et al., 2010).

It has been claimed that this dissociation between number and gender is related to different representations at the conceptual level (Antón-Méndez, Nicol, & Garrett, 2002; Igoa, García-Albea, & Sánchez-Casas, 1999). Grammatical gender (feminine vs masculine in Italian) is an abstract feature usually described as an intrinsic characteristic of the noun. Thus, grammatical gender is assumed to be stored at the level of the lexicon (Harris, 1991). On the other hand, in many languages number (singular vs plural) is a morphosyntactic feature that is attached to the lexical stem and that refers to the quantity of the referent. Contrary to grammatical gender, number is not a property of a noun and can be retrieved from the referential context (Ritter, 1993), making it a more meaningful feature for L2 learners.

In sum, although there are some mixed results in the literature, grammatical gender tends to be more challenging than number. This difference has been explained at the representational level, where gender is an arbitrary part of the lexicon and number is a morphosyntactic feature expressing a semantic property of the referent. The current study aims to contribute to this debate by contrasting gender and number processing in early L2 learners of Italian.

3. Methods

3.1 Hypotheses and predictions

The present study addresses how Mandarin-Italian bilingual children use grammatical gender and number marking on articles to anticipate upcoming nouns, in comparison with monolingual Italian children. Whereas monolingual children have been shown to effectively rely on gender and number cues in spoken word recognition, to date very few studies have focused on bilingual children. The combination of Mandarin and Italian provides an interesting test case: in Italian prenominal articles are marked for gender and number, while Mandarin lacks both articles and grammatical gender (but not number).

Firstly, we hypothesize that predictive processing is less efficient for child L2 learners than for monolingual children, especially if they have to rely on a grammatical category that is absent in their L1. Since Mandarin does not have articles, we predict that bilinguals will be less likely to anticipate upcoming nouns on the basis of determiner features. In other words, in a condition in which there is a mismatch in grammatical gender and number between the target picture and its competitor, we expect that participants will start directing their gaze toward the target picture after the onset of the article and before the onset of the noun, but bilingual participants may be delayed in comparison with monolingual participants.

Secondly, we hypothesize that grammatical gender is a more challenging feature to acquire for L2 learners than number, due to greater reliance on arbitrary lexical knowledge. We therefore predict that bilingual children will be slower when anticipating nouns based on a gender as compared to a number cue, and we predict that the discrepancy between gender and number processing will be greater for bilinguals than for monolinguals. The fact that Mandarin lacks grammatical gender while it does have a conceptual notion and grammatical expression of number may also contribute to such an asymmetry.

Thirdly, hypothesizing that gender processing becomes more native-like as L2 proficiency increases, we predict that for bilingual children, prediction of nouns based on gender-marked articles will be influenced by their proficiency in Italian in terms of vocabulary knowledge.

3.2 Participants

We tested a total of 63 children ranging from eight to eleven years old. After excluding two monolinguals and four bilinguals (see the Analysis section for more details), our final sample included 32 Italian monolinguals (M_{Age} = 9;7, *SD* = 1;1, Range = 8;0–10;8) and 25 L2 learners of Italian who spoke Mandarin as their L1 (M_{Age} = 9;9, *SD* = 1;2, Range = 8;6–11;9).

Bilingual children were from an immigration background and were raised in Mandarin-speaking families living within the Chinese community in Milan. In this community, families typically speak only Mandarin at home, while children are first exposed to Italian when they enter school (between the age of three and six). Based on the school's information, all children had at least three years of exposure to Italian.

All participants attended an Italian school in Milan. Bilingual children also attended a Chinese heritage language school where they received literacy training in Mandarin Chinese on Saturdays during the scholastic year and on weekdays during two months in the summer break.

Bilingual children took the Peabody Picture Vocabulary Test (PPVT-R; Stella, Pizzoli, & Tressoldi, 2000) in order to obtain a measure of their Italian proficiency. The mean standard score on this task was 85 (SD = 15),[1] indicating that our bilingual participants had substantially smaller vocabulary sizes in Italian in comparison to monolingual norms (M = 100, SD = 15).

3.3 Task and materials

Participants took part in a visual world eye-tracking task to measure their ability to predict the upcoming noun on the basis of the gender and number marking of the article.

The auditory stimuli consisted of Italian sentences divided into three parts: introduction (*Adesso trova* 'Now find…'), definite article (*il* 'the-MASC-SING', *la* 'the-FEM-SING', *le* 'the-FEM-PLU'), and target noun. The sentences were recorded by a female native speaker of Italian. The audios were manipulated with the Audacity® software (Audacity Team, 2020). We spliced the audio fragments in order to ensure that the introduction of the sentence (*Adesso trova il/la/le* 'Now find the') was always the same. The determiners *la, le* and *il* had a duration of 315 ms, 350 ms and 370 ms respectively. The noun always started exactly 750 ms after the determiner onset, so that in each condition anticipatory eye movements could be observed before the onset of the noun. Background noise was removed.

Sixty high-frequency nouns were selected, and their corresponding pictures were created. All nouns referred to inanimate concrete objects and were two or three syllables long. Pictures depicted simple objects on a white background, with either one single object or two identical objects. Targets and competitors were matched in syllable length.

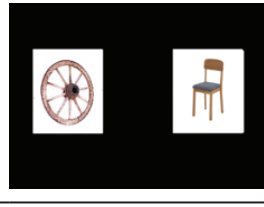
Three conditions manipulated predictability and type of cue: Early Gender, Early Number, and Late (see Table 1). In both Early conditions, there was a mismatch

1. Note that we had missing data on the vocabulary test for four children.

between the target and the competitor, allowing for prediction on the basis of the article. In the Early Gender condition, pictures showed one feminine object and one masculine object. Thus, there was a gender mismatch so that the gender marking of the article (*il* vs *la)* could lead to anticipatory eye-movements towards the target. In the Early Number condition, pictures displayed one single object on one side and two identical objects on the other side. Thus, there was a number mismatch (with matched feminine gender) so that anticipatory eye-movements to the target could be generated from the number marking of the article (*la* vs *le*). In the Late condition, the target and the competitor were matched in both gender and number in three different types of pairs : *la* vs *la, il* vs *il* and *le* vs *le*. In this condition, the target could not be identified before the onset of the noun.

There were 10 items in each Early condition, and 15 items in the Late condition. Each noun was used only once in the experiment. Gender and number of the target were balanced within the corresponding condition so that five items targeted a feminine noun in the Gender condition, and five items targeted a plural noun in the Number condition. In the Late condition, five targets were masculine singular, five were feminine singular and five were feminine plural. The target appeared on the right side in 17 trials and on the left side in 18 trials. To avoid any preference bias towards one of the pictures, we created a second list of 35 items in which targets and competitors were swapped. The two lists were balanced across participants. The presentation order of the items was randomized.

Table 1. An overview of experimental conditions

Condition	Picture	Audio
Early Gender (predictable)		Adesso trova la barca Now find the-FEM-SING boat-FEM-SING 'Now, find the boat' Competitor: *il letto* the-MASC-SING bed-MASC-SING
Early Number (predictable)		Adesso trova la foglia Now find the-FEM-SING leaf-FEM-SING 'Now, find the leaf' Competitor: *le torte* the-FEM-PLUR cakes-FEM-PLUR
Late (unpredictable)		Adesso trova la ruota Now find the-FEM-SING ruota-FEM-SING 'Now, find the wheel' Competitor: *la sedia* the-FEM-SING chair-FEM-SING

3.4 Procedure

Participants were tested individually in a quiet room with a portable computer. Data were collected with a Tobii Pro X3-120 eye-tracker that captured participants' gaze at 120 Hz. Participants sat in front of the computer screen, with their eyes 60 to 70 cm away from the display. Participants underwent a short familiarisation phase consisting of three items, after which calibration occurred. Calibration consisted of 9 red dots that participants had to fixate.

Each trial was divided into two phases. The first one was a passive phase in which gaze data was collected. Participants faced a visual setting showing a pair of pictures while listening to an audio stimulus. The pair of pictures appeared one second before the sentence started, in order to give time for visual exploration and lexical retrieval. The trial ended one second after the end of the audio stimulus, so as to capture the whole gaze pattern. In the second phase, offline accuracy was collected. At the end of the trial, a question mark appeared, indicating to the participants that they had to select the right picture by clicking on the mouse. Before moving on to the next trial, children saw a fixation cross; the experiment would only continue to the next item once they fixated on the middle of the screen. No feedback was given during the experimental phase.

3.5 Analysis

Data from four bilingual children were removed from the original sample because of poor calibration of the eye-tracker. We also excluded the data of two monolingual children who had more than 50% track-loss during the experimental phase. Thus, the analysis included 32 monolinguals and 25 bilinguals. The analysis was based on trials for which the offline response was accurate, excluding trials in which there was more than 35% track-loss. As a result, 161 trials were removed for the monolinguals (32 in the Early Gender condition, 60 in the Early Number condition and 69 in the Late condition), and 160 for the bilinguals (57 in the Early Gender condition, 42 in the Early Number condition and 61 in the Late condition), leaving respectively 959 and 711 trials for the statistical analysis.

For the purpose of our study, we created three time windows corresponding to the introduction (starting 500 ms before the onset of the article), the determiner (750 ms starting from the article onset until the noun onset) and the noun (ending 500 ms after the onset of the noun). The boundaries of each time window were shifted by 200 ms, as this is the estimated time required for the planning and execution of saccadic eye movements (Altmann, 2011). We used the *eyetrackingR* (Dink & Ferguson, 2015) and *ggplot2* (Wickham, 2016) packages to visualize the eye gaze patterns in *R* (R Core Team, 2019).

In the statistical analyses we compared looks toward the target versus competitor during the introduction and the determiner. Data were analyzed with a generalized linear mixed-effects model in which the odds of looking to the target served as the binary dependent variable (yes vs no), using the glmer function of the *lme*4 package (Bates et al., 2015) in *R*. In the first analysis, we compared bilinguals to monolinguals, to examine (1) whether they differ in the extent to which they use determiner features to predict upcoming nouns and (2) whether they differ with respect to gender versus number processing. We considered various predictors such as Time Window (Intro vs Determiner), Condition (Early Gender vs Early Number vs Late), Group (Monolingual vs Bilingual), Age, List and Item Order, as well as random intercepts for Item and Subject, and random slopes for Time window and Condition. We set the contrasts as follows: for Time Window, Intro was coded as 0 and Determiner as 1, and for Group, Monolingual was coded as +1/2 and Bilingual was coded as −1/2. For Condition, we used a ternary contrast. When comparing the two early conditions against the late condition, Late was coded as −2/3, Early Gender was coded as +1/3 and Early Number was coded as +1/3. When comparing gender versus number, Late was coded as 0, Early Gender was coded as −1/2 and Early Number was coded as +1/2. All the numerical variables were rescaled and centered around the mean. The best model was selected through a bottom-up stepwise model comparison using the Bayesian Information Criterion (BIC).

In a follow-up analysis, we focused on the subset of bilinguals who completed the vocabulary test, to test whether they were able to anticipate nouns on the basis of the gender and number of articles, and to investigate the role of L2 proficiency in their ability to do so. This time, we considered main effects and interactions of Time Window, Condition (with Late coded as the baseline) and Italian Vocabulary scores, a main effect of Item Order, and random intercepts for Subject and Item with random slopes for Time Window and Condition. Again, numeric predictors were rescaled and centered around the mean.

4. Results

The mean offline accuracy on the task was 98.9% for bilingual participants (*SD* = 10%) and 99.9% for monolingual participants (*SD* = 3.7%). The analysis of the eye-tracking data includes data from accurate trials only.

The time course of target fixations during the sentence in the three different conditions is displayed in Figure 1 for monolinguals and in Figure 2 for bilinguals.

As can be seen in Figure 1, for monolingual children the proportion of looks to the target picture started increasing during the article, before the onset of the noun, in the early gender and the early number condition, but not in the late condition.

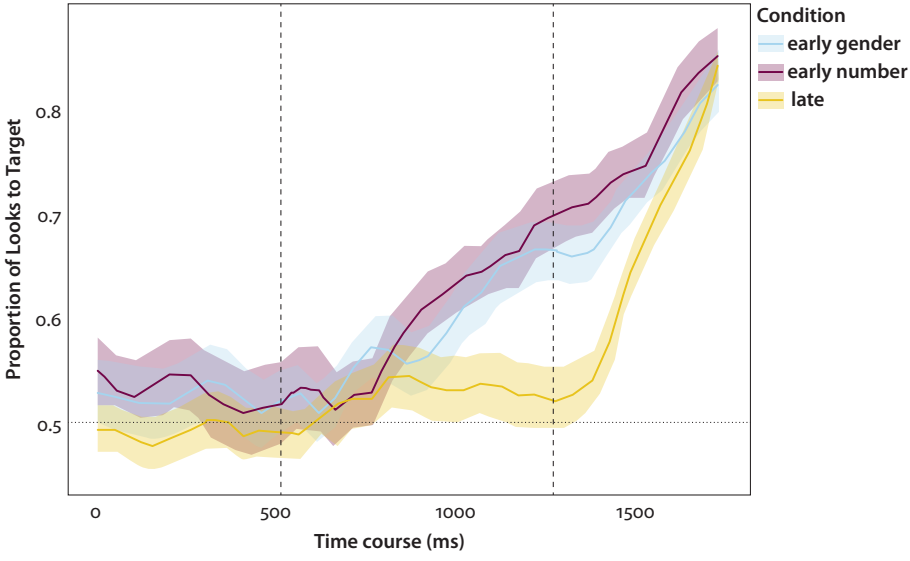


Figure 1. Time course of the proportions of looks toward the target (versus competitor) in the three conditions for monolingual participants. The first vertical line represents determiner onset and the second vertical line represents noun onset, shifted 200 ms to account for saccade planning

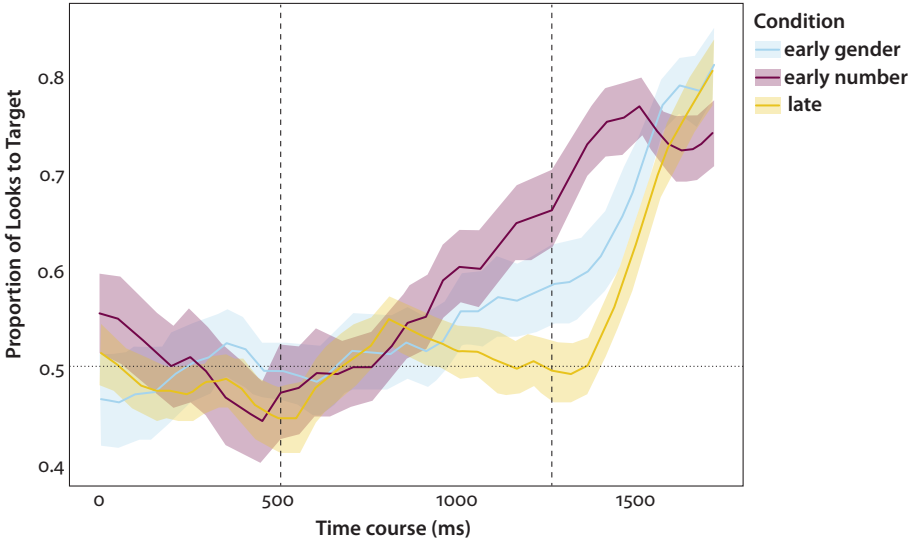


Figure 2. Time course of the proportions of looks toward the target (versus competitor) in the three conditions for bilingual participants. The first vertical line represents determiner onset and the second vertical line represents noun onset, shifted 200 ms to account for saccade planning

This suggests that monolinguals anticipated nouns on the basis of articles, with nearly identical patterns for gender and number. In contrast, as can be seen in Figure 2, bilinguals also showed anticipation of the target noun, but they seemed to be considerably slower to direct their gaze to the target picture in the gender condition than in the number condition.

The statistical analysis confirmed these patterns. The most predictive model that converged included a three-way interaction between Time window, Condition and Group as well as their main effects, a main effect of Item Order, and random intercepts for Item and Subject with random slopes for Time window and Condition. Table 2 provides a summary of the model output.

Table 2. Estimated odds ratios, 95% confidence intervals and associated p-values of main and interaction effects for the comparison between monolingual and bilingual participants

Generalized linear mixed model			
Looks to target (yes or no) ~ Time window (intro vs determiner) * Condition (early gender vs early number vs late) * Group (monolingual vs bilingual) + Item order + (1 | Item) + (1 + Time window + Condition | Subject)			
Fixed factor	**Est. odds ratio**	**95% CI**	***p***
Time window	1.20	1.1 .. .1.4	.002
Condition (early vs late)	1.17	.92 .. 1.5	.209
Condition (gender vs number)	1.08	.79 .. 1.5	.622
Group	1.09	.92 .. 1.3	.304
Item order	.933	.92 .. .94	<.001
Time window : Condition (early gender and early number vs late)	1.05	1.0 .. 1.1	.010
Time window : Condition (early gender vs early number)	.983	.94 .. 1.0	.481
Time window: Group	1.08	.86 .. 1.4	.514
Condition (early gender and early number vs late) : Group	1.15	.88 .. 1.5	.310
Condition (early gender vs early number): Group	.858	.58 .. 1.3	.446
Time window : Condition (early gender and early number vs late) : Group	1.02	.95 .. 1.1	.570
Time window: Condition (early gender vs early number): Group	1.31	1.1 .. 1.4	<.001

Overall, participants were more likely to look at the target picture during the determiner than during the introduction, as shown by the significant main effect of Time window. Item order also had a significant effect; participants became less likely to look at the target during the course of the experiment.

The significant interaction between Time window and Condition (Early vs Late) shows a prediction effect; the difference between the (predictable) Early conditions

and the (unpredictable) Late condition was significantly greater during the article than during the introduction, indicating that participants anticipated the target noun on the basis of determiner features. We found no significant difference in prediction abilities between monolinguals and bilinguals, since the interaction between Time Window, Condition (early gender and early number vs Late) and Group was not significant. When comparing early number to early gender, however, there was a significant three-way interaction between Time Window, Condition (early gender vs early number) and Group, indicating that the difference between gender and number processing was significantly greater for bilinguals than for monolinguals. In other words, overall we found a prediction effect, but bilingual participants were less efficient when processing gender as compared to number.

We then conducted a follow-up analysis, focusing on bilingual participants only, to test whether they were able to anticipate nouns based on gender- and number-marked articles, taking their Italian vocabulary knowledge into account. This analysis included only 21 bilingual participants, since we had missing data on the vocabulary test for four children. The model included main and interaction effects of Time Window, Condition and Vocabulary, a main effect of Item order, random intercepts for Subject and Item and random slopes for Condition and Time Window. The model output is provided in Table 3.

Table 3. Estimated odds ratios, 95% confidence intervals and associated p-values of main and interaction effects for bilingual participants only

Generalized linear mixed model			
Looks to target (yes or no) ~ Time window (intro vs determiner) * Condition (early gender vs early number vs late) * Vocabulary + Item order + (1 | Item) + (1 + Condition + Time window | Subject)			
Fixed factor	**Est. odds ratio**	**95% CI**	***p***
Time window	1.19	.97 .. 1.5	.094
Condition (late vs early gender)	1.09	.70 .. 1.7	.707
Condition (late vs early number)	1.15	.73 .. 1.8	.549
Vocabulary	.977	.18 .. 1.2	.808
Item order	.887	.87 .. .90	<.001
Time window : Condition (late vs early gender)	.994	.92 .. 1.1	.872
Time window : Condition (late vs early number)	.961	.89 .. 1.0	.285
Time window: Vocabulary	.877	.72 .. 1.1	.200
Condition (late vs early number) : Vocabulary	.844	.62 .. 1.2	.302
Condition (late vs early gender): Vocabulary	1.06	.76 .. 1.5	.735
Time window : Condition (late vs early gender) : Vocabulary	1.33	1.2 .. 1.4	<.001
Time window: Condition (late vs early number): Vocabulary	.931	.87 .. 1.0	.054

In the second analysis we found no significant prediction effect for bilingual participants as a group (neither for gender nor number), since there was no significant interaction between Time Window and Condition. However, the degree to which bilinguals used gender-marked articles to anticipate nouns was influenced by their Italian vocabulary knowledge, as shown by the three-way interaction between Time Window, Condition (late vs early gender) and Vocabulary. Note that there was also a marginally significant three-way interaction between Time Window, Condition (late vs early number) and Vocabulary, suggesting that there may be a relation between L2 proficiency and prediction based on number, too.

As in the previous model, there was also a significant main effect of Item order, indicating that participants became less likely to look at the target during the course of the experiment.

5. Discussion and conclusion

This study addressed linguistic prediction on the basis of determiner features in Mandarin-Italian bilingual children. We aimed to contrast the processing of grammatical gender with number, since gender has been hypothesized to be particularly challenging for L2 speakers, specifically when this feature is absent in the L1. The bilingual participants were compared to a group of monolingual Italian children.

The first aim of this study was to investigate to what extent native Mandarin-speaking children who are learning Italian as their L2 use articles to anticipate upcoming nouns in Italian, in comparison with monolingual controls. We hypothesized that bilingual children use less predictive processing than monolingual children, especially when they need to attend to a grammatical category that is absent in their L1. Therefore, we predicted that Mandarin-Italian bilingual participants would be less likely to anticipate target nouns on the basis of articles than monolingual Italian participants. This prediction was not borne out by the data: we found no significant difference between bilinguals and monolinguals in the extent to which they were able to anticipate, providing no evidence for a general difference in predictive processing.

Our results suggest that, overall, participants were able to anticipate nouns based on morphosyntactic agreement with the article, complementing previous research showing that children process speech predictively (Cholewa et al., 2019; Lew-Williams & Fernald, 2007; Mani & Huettig, 2012). However, in a post-hoc analysis focusing on the effect of language proficiency in a subset of bilingual children, we found no significant prediction effects. This result should be interpreted with caution since the post-hoc analysis only focused on a smaller sample, which may have led to insufficient statistical power. Previous research suggests that bilingual

children are able to process speech predictively (Lemmerth & Hopp, 2019; Bosch & Foppolo, submitted), and the gaze pattern of bilingual participants in the present study suggests a similar tendency, at least for number processing. Therefore, future studies that include a larger number of participants are needed to determine the extent to which child L2 learners are able to anticipate upcoming words when they have to attend to a grammatical category that is absent in their L2.

Our second aim was to compare the processing of grammatical gender and number in Mandarin-Italian bilingual children with that of their monolingual Italian peers. We hypothesized that there would be a discrepancy between gender and number processing for bilingual participants, since gender processing requires greater reliance on arbitrary lexical knowledge, and Mandarin lacks grammatical gender despite having a conceptual notion and grammatical expression of number. Our results lend support to this hypothesis. As predicted, the difference between gender and number processing was significantly greater for bilinguals than for monolinguals. Although this pattern was not statistically confirmed by our post-hoc analysis focusing on bilinguals only, bilingual children showed a tendency to be slower with grammatical gender than with number processing.

Previous studies have found a dissociation between gender and number in late L2 learners; while one study found faster processing of gender in comparison with number (Tokowicz & MacWhinney, 2005), other studies found that gender is more challenging for L2 speakers than number (Gabriele et al., 2013; White et al., 2004; Gillon-Dowens et al., 2010; Lew-Williams & Fernald, 2009; Hopp, 2012). Such an advantage for number over gender processing may be due to different representations at the conceptual level. Assuming that L2 learners with poorer language proficiency have more difficulties with lexical retrieval, they may experience more difficulties in gender processing than in number processing, since gender processing requires reliance on arbitrary lexical knowledge, whereas number is conceptually more tangible and strongly linked to the referential context. Moreover, L1 characteristics may play a role since most studies reporting a discrepancy between gender and number processing focus on L2 learners whose L1 does not have grammatical gender (typically, English).

Our third aim was to test the effect of L2 proficiency in terms of vocabulary knowledge. The results support our hypothesis that grammatical gender processing becomes more native-like as L2 proficiency increases. As predicted, the extent to which bilingual participants used the grammatical gender of the article to anticipate subsequent nouns was significantly influenced by their Italian vocabulary knowledge. This complements previous research showing that grammatical gender processing is related to language proficiency (Lew-Williams & Fernald, 2010; Dussias et al., 2013; Gabriele et al., 2013; Bosch & Foppolo, submitted).

The effect of proficiency also provides a potential explanation for the observation that the Mandarin-Italian bilinguals in our study appeared to be considerably slower when processing grammatical gender as compared to bilingual children in previous research. Other studies have found that as long as there is no interference due to cross-linguistic gender incongruency, bilingual children efficiently anticipate nouns based on grammatical gender (Lemmerth & Hopp, 2019 in Russian-German bilinguals; Bosch & Foppolo submitted in Italian-German bilinguals). However, the German-Italian children tested by Bosch and Foppolo were attending bilingual schools and living in bilingual communities with strong ties to Italy, while the Mandarin-Italian children in our study were growing up in a relatively closed Chinese community, with considerably less exposure to Italian outside of school. As a result, our participants had lower L2 proficiency, as indicated by weaker vocabulary skills in Italian. Bosch and Foppolo's participants were tested with the same standardized vocabulary test, and the results showed that their Italian language proficiency (M = 98, SD = 20) was considerably higher than that of the Mandarin-Italian bilinguals in the present study (M = 85, SD = 15). Moreover, age of acquisition may play a role. While participants in Lemmerth and Hopp (2019) were either simultaneous or very early sequential bilinguals (age of onset before age three), like the majority of participants in Bosch and Foppolo (submitted), participants in the current study only started learning Italian when they entered kindergarten after the age of three.

In addition to child-level factors such as proficiency, age of acquisition and language use, the difference between our results and previous findings about gender processing in bilingual children may be due to the different characteristics of the L1. German, Russian and Italian all have grammatical gender, thus enabling positive transfer from the L1 to the L2. Mandarin, on the other hand, does not have grammatical gender, which means that no such transfer is possible. Consequently, grammatical gender in the L2 may be more difficult for Mandarin-speaking children than for children with a gender-marked L1.

At least three limitations should be acknowledged. Firstly, our sample size was relatively small, and follow-up studies are needed to confirm the patterns observed in the present study. Secondly, we are not able to tease apart the two possible explanations for a potential discrepancy between gender and number processing for bilingual participants. In other words, bilinguals showed a larger difference between gender and number processing than monolinguals, but we do not know whether this is due to L1 effects or to child-level factors such as proficiency. More research is needed to disentangle this by systematically examining the effect of individual differences in input and language proficiency in both the L1 and the L2, or by testing different language pairs. One possibility would be to repeat the current study with bilingual children who speak two languages that have both grammatical gender

and number, as to test whether bilinguals differ from monolinguals independently of L1 effects. Thirdly, even though we only included trials that were answered correctly, we cannot be certain if participants actually knew all the nouns and their grammatical gender. Future studies may therefore want to include a production task, in order to test participants' knowledge of the nouns used in the experiment.

Acknowledgements

We thank Francesca Foppolo for her contribution to the experimental design and analysis, Silvia Silleresi for her help in creating the experimental stimuli, Giulia Mornati for her involvement in implementing the task, Abby Deng and Antonella Guarascio for helping with data collection and data entry, and Carlo Toneatto for technical support. Finally, we are very grateful to the children, their families and the school for their participation.

Funding

This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska Curie grant agreement No 765556, as a part of the MultiMind project.

References

Altmann, G. T. (2011). The mediation of eye movements by spoken language. In S. Liversedge, I. Gilchrist, & S. Everling (Eds.), *The Oxford handbook of eye movements* (pp. 979–1004). Oxford University Press.

Altmann, G. T., & Kamide, Y. (1999). Incremental interpretation at verbs: Restricting the domain of subsequent reference. *Cognition*, 73(3), 247–264. https://doi.org/10.1016/S0010-0277(99)00059-1

Antón-Méndez, I., Nicol, J. L., & Garrett, M. F. (2002). The relation between gender and number agreement processing. *Syntax*, 5(1), 1–25. https://doi.org/10.1111/1467-9612.00045

Audacity Team. (2020). Audacity(R): Free audio editor and recorder [Computer application].

Ayoun, D., & Maranzana, S. (2020). Italian grammatical gender: A corpus study from a second language acquisition perspective. *Italica*, 97(2), 306–333.

Barber, H., & Carreiras, M. (2003). Integrating gender and number information in Spanish word pairs: An ERP study. *Cortex*, 39(3), 465–482. https://doi.org/10.1016/S0010-9452(08)70259-4

Barber, H., & Carreiras, M. (2005). Grammatical gender and number agreement in Spanish: an ERP comparison. *Journal of Cognitive Neuroscience*, 17(1), 137–153. https://doi.org/10.1162/0898929052880101

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed effects models using lme4. *Journal of Statistical Software*, 67(1), 1–48. https://doi.org/10.18637/jss.v067.i01

Bosch, J. E., & Foppolo, F. (submitted). Predictive processing of grammatical gender in bilingual children: The effect of cross-linguistic incongruence and language dominance.

Carreiras, M., Carr, L., Barber, H. A., & Hernandez, A. (2010). Where syntax meets math: Right intraparietal sulcus activation in response to grammatical number agreement violations. *NeuroImage*, 49(2), 1741–1749. https://doi.org/10.1016/j.neuroimage.2009.09.058

Chini, M. (1995). Grammatiche a confronto: La categoria grammaticale del genere nella competenza di nativi italofoni e nelle interlingue di apprendenti dell'italiano come L2. In P. Desideri (Ed.), *L'universo delle lingue. Confrontare lingue e grammatiche nella scuola* (pp. 277–294). Quaderni del Giscel, La Nuova Italia.

Cholewa, J., Neitzel, I., Bürsgens, A., & Günther, T. (2019). Online-processing of grammatical gender in noun-phrase decoding: An eye-tracking study with monolingual German 3rd and 4th graders. *Frontiers in Psychology*, 10, 2586. https://doi.org/10.3389/fpsyg.2019.02586

Cole, P., & Segui, J. (1994). Grammatical incongruency and vocabulary types. *Memory & Cognition*, 22(4), 387–394. https://doi.org/10.3758/BF03200865

Corbett, G. G. (2000). *Number*. Cambridge University Press. https://doi.org/10.1017/CBO9781139164344

Corrêa, L. S., Augusto, M. R., & Ferrari-Neto, J. (2005). The early processing of number agreement in the DP: Evidence from the acquisition of Brazilian Portuguese. Talk given at the 30th Boston University Conference on Language Development, Boston.

Dahan, D., Swingley, D., Tanenhaus, M. K., & Magnuson, J. S. (2000). Linguistic gender and spoken-word recognition in French. *Journal of memory and Language*, 42(4), 465–480. https://doi.org/10.1006/jmla.1999.2688

Dink, J. W., & Ferguson, B. (2015). EyetrackingR: An R library for eye-tracking data analysis. Retrieved on 11 June from www.eyetracking-r.com

Dispaldro, M., Ruggiero, A., & Scali, F. (2015). Real-time comprehension of gender and number in four-to seven-year-old children: A study of the relationship between Italian clitic pronouns and visual picture referents. *Journal of Child Language*, 42(4), 917–931. https://doi.org/10.1017/S0305000914000427

Dussias, P. E., Kroff, J. R. V., Tamargo, R. E. G., & Gerfen, C. (2013). When gender and looking go hand in hand: Grammatical gender processing in L2 Spanish. *Studies in Second Language Acquisition*, 35(2), 353–387. https://doi.org/10.1017/S0272263112000915

Faussart, C., Jakubowicz, C., & Costes, M. (1999). Gender and number processing in spoken French and Spanish. *Rivista di Linguistica*, 11(1), 75–101.

Fedden, S., & Corbett, G. G. (2017). Gender and classifiers in concurrent systems: Refining the typology of nominal classification. *Glossa*, 2(1), 1–47. https://doi.org/10.5334/gjgl.177

Feigenson, L., & Carey, S. (2005). On the limits of infants' quantification of small object arrays. *Cognition*, 97(3), 295–313. https://doi.org/10.1016/j.cognition.2004.09.010

Ferrari, A., & Zampese, L. (2016). *Grammatica: Parole, frasi, testi dell'italiano*. Carocci.

Foucart, A., & Frenck-Mestre, C. (2011). Grammatical gender processing in L2: electrophysiological evidence of the effect of L1–L2 syntactic similarity. *Bilingualism:Language and Cognition*, 14(3), 379–399. https://doi.org/10.1017/S136672891000012X

Foucart, A., & Frenck-Mestre, C. (2012). Can late L2 learners acquire new grammatical features? Evidence from ERPs and eye-tracking. *Journal of Memory and Language*, 66(1), 226–248. https://doi.org/10.1016/j.jml.2011.07.007

Frankowsky, M., & Ke, D. (2016). Humanness and classifiers in Mandarin Chinese. *Language and Cognitive Science*, 2(1), 55–67. https://doi.org/10.17923/LCS201601003

Friederici, A. D. (2002). Towards a neural basis of auditory sentence processing. *Trends in Cognitive Sciences*, 6(2), 78–84. https://doi.org/10.1016/S1364-6613(00)01839-8

Gabriele, A., Fiorentino, R., & Bañón, J. A. (2013). Examining second language development using event-related potentials: A cross-sectional study on the processing of gender and number agreement. *Linguistic Approaches to Bilingualism*, 3(2), 213–232. https://doi.org/10.1075/lab.3.2.04gab

Gillon-Dowens, M., Vergara, M., Barber, H. A., & Carreiras, M. (2010). Morphosyntactic processing in late second-language learners. *Journal of Cognitive Neuroscience*, 22(8), 1870–1887. https://doi.org/10.1162/jocn.2009.21304

Gillon-Dowens, M., Guo, T., Guo, J., Barber, H., & Carreiras, M. (2011). Gender and number processing in Chinese learners of Spanish–evidence from event related potentials. *Neuropsychologia*, 49(7), 1651–1659. https://doi.org/10.1016/j.neuropsychologia.2011.02.034

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

Guillelmon, D., & Grosjean, F. (2001). The gender marking effect in spoken word recognition: The case of bilinguals. *Memory & Cognition*, 29(3), 503–511. https://doi.org/10.3758/BF03196401

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

Haspelmath, M. (2018). Toward a new conceptual framework for comparing gender systems and some so-called classifier systems. Talk presented at Stockholm University, Department of Linguistics on 13 April.

Hopp, H. (2012). The on-line integration of inflection in L2 processing: Predictive processing of German gender. In A. K. Biller, E. Y. Chung, & A. E. Kimball (Eds.), *Proceedings of the 36th annual Boston University Conference on Language Development* (pp. 226–245). Cascadilla Press.

Hopp, H., & Lemmerth, N. (2018). Lexical and syntactic congruency in L2 predictive gender processing. *Studies in Second Language Acquisition*, 40(1), 171–189. https://doi.org/10.1017/S0272263116000437

Huang, C. R., & Ahrens, K. (2003). Individuals, kinds and events: classifier coercion of nouns. *Language Science*, 25(4), 353–373. https://doi.org/10.1016/S0388-0001(02)00021-9

Igoa, J. M., García-Albea, J. E., & Sánchez-Casas, R. (1999). Gender-number dissociations in sentence production in Spanish. *Rivista di Linguistica*, 11(1), 163–196.

Iljic, R. (1994). *Quantification in Mandarin Chinese: Two markers of plurality*. Mouton de Gruyter.

Kouider, S., Halberda, J., Wood, J., & Carey, S. (2006). Acquisition of English number marking: The singular-plural distinction. *Language Learning and Development*, 2(1), 1–25. https://doi.org/10.1207/s15473341lld0201_1

Lemmerth, N., & Hopp, H. (2019). Gender processing in simultaneous and successive bilingual children: cross-linguistic lexical and syntactic influences. *Language Acquisition*, 26(1): 21–45. https://doi.org/10.1080/10489223.2017.1391815

Lew-Williams, C., & Fernald, A. (2007). Young children learning Spanish make rapid use of grammatical gender in spoken word recognition. *Psychological Science*, 18(3), 193–198. https://doi.org/10.1111/j.1467-9280.2007.01871.x

Lew-Williams, C., & Fernald, A. (2009). Fluency in using morphosyntactic cues to establish reference: How do native and non-native speakers differ? In J. Chandlee, M. Franchini, S. Lord, & G-M. Rheiner (Eds), *Proceedings of the 33rd annual Boston University Conference on Language Development* (pp. 290–301). Cascadilla Press.

Lew-Williams, C., & Fernald, A. (2010). Real-time processing of gender-marked articles by native and non-native Spanish speakers. *Journal of Memory and Language*, 63(4), 447–464. https://doi.org/10.1016/j.jml.2010.07.003

Li, P., Ogura, T., Barner, D., Yang, S. J., & Carey, S. (2009). Does the conceptual distinction between singular and plural sets depend on language? *Developmental Psychology*, 45(6): 1644–1653. https://doi.org/10.1037/a0015553

Lukatela, G., Carello, C., & Turvey, M. T. (1987). Lexical representation of regular and irregular inflected nouns. *Language and Cognitive Processes*, 2(1): 1–17. https://doi.org/10.1080/01690968708406349

Mani, N. & Huettig, F. (2012). Prediction during language processing is a piece of cake – but only for skilled producers. *Journal of Experimental Psychology: Human Perception and Performance*, 38(4), 843–847.

Morales, L., Paolieri, D., Dussias, P. E., Kroff, J. R. V., Gerfen, C., & Bajo, M. T. (2016). The gender congruency effect during bilingual spoken-word recognition. *Bilingualism: Language and Cognition*, 19(2), 294–310. https://doi.org/10.1017/S1366728915000176

Padovani, R., & Cacciari, C. (2003). Il ruolo della trasparenza morfologica nel riconoscimento di parole in Italiano. *Giornale Italiano di Psicologia* 30(4), 749–772.

Panzini, A. (2017). *Grammatica italiana*. Edizioni Lindau.

R Core Team. (2019). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing.

Ritter, E. (1993). Where's gender? *Linguistic Inquiry*, 24(4), 795–803.

Robertson, E. K., Shi, R., & Melançon, A. (2012). Toddlers use the number feature in determiners during online noun comprehension. *Child Development*, 83(6), 2007–2018. https://doi.org/10.1111/j.1467-8624.2012.01828.x

Rossi, S., Gugler, M. F., Friederici, A. D., & Hahne, A. (2006). The impact of proficiency on syntactic second-language processing of German and Italian: Evidence from event related potentials. *Journal of Cognitive Neuroscience*, 18(12), 2030–2048. https://doi.org/10.1162/jocn.2006.18.12.2030

Rullmann, H., & You, A. (2006). General number and the semantics and pragmatics of indefinite bare nouns in Mandarin Chinese. In K. von Heusinger & K. Turner (Eds.), *Where semantics meets pragmatics* (pp. 175–196). Brill.

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

Stella, G., Pizzoli, C., & Tressoldi, P. E. (2000). *Il Peabody Test. Test di vocabolario recettivo*. Omega.

Tokowicz, N., & MacWhinney, B. (2005). Implicit and explicit measures of sensitivity to violations in second language grammar: An event-related potential investigation. *Studies in Second Language Acquisition*, 27(2), 173–204. https://doi.org/10.1017/S0272263105050102

Van Heugten, M., & Shi, R. (2009). French-learning toddlers use gender information on determiners during word recognition. *Developmental Science*, 12(3), 419–425. https://doi.org/10.1111/j.1467-7687.2008.00788.x

White, L., Valenzuela, E., Kozlowska–Macgregor, M., & Leung, Y. K. I. (2004). Gender and number agreement in nonnative Spanish. *Applied Psycholinguistics*, 25(1), 105–133. https://doi.org/10.1017/S0142716404001067

Wicha, N. Y., Moreno, E. M., & Kutas, M. (2004). Anticipating words and their gender: An event-related brain potential study of semantic integration, gender expectancy, and gender agreement in Spanish sentence reading. *Journal of Cognitive Neuroscience*, 16(7), 1272–1288. https://doi.org/10.1162/0898929041920487

Wickham, H. (2016). *ggplot2: Elegant graphics for data analysis*. Springer. https://doi.org/10.1007/978-3-319-24277-4

Zhang, N. N. (2014). Expressing number productively in Mandarin Chinese. *Linguistics*, 52(1): 1–34. https://doi.org/10.1515/ling-2013-0055

Language index

Name index

Subject index

Gender as a morphosyntactic feature is arguably "an endlessly fascinating linguistic category" (Corbett 2014: 1). One may even say it is among "the most puzzling of the grammatical categories" (Corbett 1991: 1) that has raised probing questions from various theoretical and applied perspectives. Most languages display semantic and/or formal gender systems with various degrees of opacity and complexity, and even closely related languages present distinct differences, creating difficulties for second language learners. The first three chapters of this volume present critical reviews in three different areas – gender assignment in mixed noun phrases, subtle gentle biases and the gender acquisition in child and adult heritage speakers of Spanish – while the next six chapters present new empirical evidence in the acquisition of gender by bilingual children, adult L2/L3 learners and heritage speakers of various languages such as Italian, German, Dutch or Mandarin-Italian.

"This timely volume reflects the vitality of current research on gender, including both critical reviews and novel empirical studies. Topics include gender acquisition by bilingual children, issues of adult learners – both L2 and L3 – and heritage speakers of very different languages. The variety in subject matter is matched by the choice of research methods; these range from corpus work, semi-guided interviews and picture narration, through to eye-tracking. An interesting feature is the way in which gender serves both as the research topic and the means to address other research questions. As a result, the volume appeals to a wide range of readers."

Greville G. Corbett, *Surrey Morphology Group*

ISBN 978 90 272 1009 8

JOHN BENJAMINS PUBLISHING COMPANY