Advances in Historical Sociolinguistics
12

Spanish Socio-Historical Linguistics

Isolation and contact

EDITED BY
Whitney Chappell
Bridget Drinka

John Benjamins Publishing Company

Spanish Socio-Historical Linguistics

Advances in Historical Sociolinguistics (AHS)

ISSN 2214-1057

Over the last three decades, historical sociolinguistics has developed into a mature and challenging field of study that focuses on language users and language use in the past. The social motivation of linguistic variation and change continues at the forefront of the historical sociolinguistic enquiry, but current research does not stop there. It extends from social and regional variation in language use to its various communicative contexts, registers and genres, and includes issues in language attitudes, policies and ideologies. One of the main stimuli for the field comes from new digitized resources and large text corpora, which enable the study of a much wider social coverage than before. Historical sociolinguists use variationist and dialectological research tools and techniques, perform pragmatic and social network analyses, and adopt innovative approaches from other disciplines. The series publishes monographs and thematic volumes, in English, on different languages and topics that contribute to our understanding of the relations between the individual, language and society in the past.

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

Volume 12

Spanish Socio-Historical Linguistics. Isolation and contact
Edited by Whitney Chappell and Bridget Drinka

Spanish Socio-Historical Linguistics

Isolation and contact

Edited by

Whitney Chappell
Bridget Drinka
University of Texas, San Antonio

John Benjamins Publishing Company
Amsterdam / Philadelphia

DOI 10.1075/ahs.12

Cataloging-in-Publication Data available from Library of Congress:
LCCN 2021002696 (PRINT) / 2021002697 (E-BOOK)

ISBN 978 90 272 0864 4 (HB)
ISBN 978 90 272 5995 0 (E-BOOK)

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

Table of contents

New perspectives on Spanish socio-historical linguistics

Bridget Drinka and Whitney Chappell
The University of Texas at San Antonio

1. Introduction

This volume contributes to the growing field of research focused on the intersection of historical linguistics and sociolinguistics. Characterized under a dual designation – as socio-historical linguistics and as historical sociolinguistics – this innovative field of inquiry brings together research centered on the analysis of social motivations for change and on historical contexts for variation. Essential to this nexus of disciplines is a recognition that both historical linguistics and sociolinguistics share a commitment to empirical methods, a focus on the role of social motivation for change, and an ultimate aim to explain how and why language changes.

Since the publication of the first monograph explicitly juxtaposing the two disciplines, Suzanne Romaine's 1982 *Socio-historical linguistics: Its status and methodology*, the field has flourished, stimulated especially by the important early work of Nevalainen (Nevalainen & Raumolin-Brunberg 1996), Jahr (Jahr 1999), and by early advocates of socio-historical approaches to language contact (Thomason & Kaufman 1988), social network theory (Milroy & Milroy 1985), and geolinguistic analysis (Trudgill 1983). Since those early days, numerous innovative methodologies have emerged for carrying out corpus analysis, data mining, textual analysis, dialect mapping, and other types of research. The important collection of Hernández-Campoy and Conde-Silvestre (2012) represents a state-of-the-art compendium for recent advances in the field, pointing to an increased focus on quantitative measures, a heavier concentration on pragmatic variables, and a greater interest in historical dialectology, socio-demographics, and the diffusion of innovations. Recent volumes in the Advances in Historical Sociolinguistics series – those edited, for example, by Säily et al. (2017), Nevalainen et al. (2018), and Hickey (2019) – expand upon these issues, focusing on such topics as innovative constructions propagated by social media, the role of prescriptive forces on the vernacular, and the examination of data gleaned from the letters of immigrants and emigrants.

https://doi.org/10.1075/ahs.12.c01dri

What the present volume aims to contribute to this trajectory of inquiry is a new schematization of socio-historical tenets and methods and a new set of answers to persistent questions drawn from both disciplines. In response to pressing questions such as "how did this phenomenon arise?" or "why did it arise at this particular point in space and time?," we suggest that an answer cannot be given without reference to the socio-cultural context and, more specifically, to the interconnected factors of socio-cultural isolation and contact. Hence, we seek a refocusing of the two disciplines onto this intersectional, interstitial region with special emphasis on the Spanish-speaking world.

2. Previous research on isolation and contact

Isolation as a sociodemographic factor has been studied much less extensively than contact has, but its role is significant, especially as it interacts with numerous other factors. The concept of isolation has been of interest to researchers since the nineteenth century, when a number of early studies of Sprachinseln ('speech islands') were produced (Leck 1884; Gehre 1886; Hauffen 1895). Isolated communities in regions like the Alps have continued to attract scholarly attention ever since. Gaeta (2020: 205), for example, notes that the Walser German varieties located in northern Italy, at the southwestern edge of the German dialectal continuum, show extreme conservatism, and, yet, have developed a number of noteworthy innovations due to extensive contact with Romance languages. (See also dal Negro 2004 for additional discussion.) Geographical isolation naturally occurs on islands, several of which have been studied in some detail. Both Martha's Vineyard (Labov 1963) and Ocracoke (Wolfram, Hazen & Schilling-Estes 1999) were isolated from mainland trends until ferries and bridges made it possible for tourists and new residents to access the islands. These island communities have thus emerged from their isolation, and can be viewed as "post-insular" (Wolfram & Schilling-Estes 1998: 119).

Not only does isolation operate with regard to geographic separation, but it also functions on a social level: enclave communities can arise at the periphery of social space, just as they do at geographical peripheries. Minority groups may undergo social separation or segregation, reflecting asymmetrical social hierarchies. Labov (2010) has demonstrated, for example, that the differences between African American dialects and other varieties have been widening, signaling an increase in social separation. This social isolation has, at the same time, engendered a reinforcement of local identity and covert prestige among speakers of African American Vernacular English (Lippi-Green 1997; Labov 2010). Social separation and later contact is also evident in the "crossing" by Anglo youth in London as they acquire Caribbean linguistic features (Rampton 1995).

Isolation plays a significant role in the formation of diasporic communities, new communities, and koines. Schreier (2017: 356), following Esman (1996: 316), defines a diaspora as "a minority ethnic group of migrant origin that maintains sentimental or material links with its land of origin because of social exclusion, internal cohesion, or geopolitical factors." He identifies a "diasporic consciousness" shared by the community as it exists in space with others while not assimilating with them. The formation of new dialects likewise entails the separation of speakers from a point of origin, and an eventual renegotiation of identity. Trudgill (1986: 107) recognizes a key role for koineization in this formation of new towns and colonies; Tuten (2003: 83) likewise points to the role that isolation plays in the formation of unique local koines in these new towns, frontiers, and colonies. Tuten (2003: 54), following Trudgill (1986), describes what occurs when speakers of various dialectal backgrounds move into new territories and establish new social networks and norms:

> In a koineizing social situation […] all the speakers become innovators vis-à-vis speakers of other dialects. As a result, the number of relative innovations peaks at the same time that the strength of norm enforcement mechanisms (social networks) declines to a minimum.

It is thus the combined effect of isolation and contact which fosters the relaxation of old norms and the establishment of new ones, especially through micro-level processes like accommodation and their macro-level outcomes, such as focusing, leveling, and simplification.

Scholarly interest in language contact as a motivation for change has been more plentiful than that focusing on isolation, with work on this topic expanding substantially in recent years. Among the earliest models representing contact as significant were those of Johannes Schmidt (1872), which established the Wellentheorie 'Wave Theory,' and Hugo Schuchardt (1883, 1884), which helped initiate the study of pidgins and creoles. Other important works followed: Sandfeld's (1930) in-depth analysis of language contact in the Balkans, the work of the Italian Neolinguisti (e.g., Bartoli 1925) on the geographical distribution of innovations, and Boas's (1938) studies of areal spread across linguistic boundaries among Native American languages, among others. Yet it was not until 1953 that a systematic, comprehensive study of language contact appeared, with the publication of Uriel Weinreich's groundbreaking *Languages in contact*. Weinreich insisted that a complete explanation for the effects of contact will not be obtained without a consideration of extra-linguistic factors – a well-accepted claim by today's standards, but controversial at the time. Trudgill (1983) went on to adopt the methodology of demographers and geographers to track the spread of innovations across the map, proposing, for example, the "Gravity Model" to account for the diffusion of uvular /r/ from

Paris to other large urban centers but not to intervening territory. Expanding on these insights, Labov (2003, 2010) constructed the similar "Cascade Model," which provides insight, for example, into the diffusion of the Northern Cities Vowel Shift down the I-55 corridor from Chicago to St. Louis.

It was the publication of Thomason and Kaufman's (1988) *Language contact, creolization, and genetic linguistics* that launched the recent increase in number and quality of studies on language contact. Among the most important contributions of their book are the demonstration that social factors play an essential role in explaining change and that anything can be borrowed, given sufficient social motivation. The authors recognize two types of contact – contact with language maintenance ("borrowing") and contact with language shift ("substratum interference") (Thomason and Kaufman 1988: 37–42). They go on to establish a five-point scale of contact intensity, from casual contact with lexical borrowing alone (Category 1) to profound structural influence, including radical changes to the typological structure or morphosyntactic order of constituents (Category 5).

Building on this foundation, researchers on language contact have taken several new tacks in recent years: areal linguistics has moved in the direction of quantitative, empirical methodology, and now includes analysis of trends on a world-wide scale (Dahl 1985, 1995, 2008; Nichols 1992; Cysouw 2011); new integrative approaches have been proposed, such as the replication model of Heine and Kuteva (2005), which incorporates insights from grammaticalization with those from contact studies; and new models have been formulated, such as the "Stratified Convergence Zone" (Drinka 2017), which positions innovations on a three-dimensional map representing space and time.

Within the Spanish-speaking world, social isolation and contact play a crucial role in myriad linguistic phenomena, as all native Spanish speakers outside of the Iberian Peninsula "…are the linguistic heirs of the Spanish language diaspora that began in 1492 – with the expulsion of Spanish-speaking (Sephardic) Jews, and with the arrival of the Spanish language in the Americas" (Lipski 2010: 550). According to Lipski (2010), the Spanish spoken by over 400 million native speakers today includes Spanish as an official or co-official national language, as contact varieties across borders or due to historical colonization, and as the language of immigrants in new social contexts. In addition to the incomplete data with which all historical linguists must work, the sheer breadth of such disparate international circumstances involving isolation and contact in Hispanic contexts makes any attempt at carrying out a comprehensive socio-historical linguistic exploration a daunting task. An examination of the vast diversity of linguistic phenomena in the Spanish-speaking world has been partially undertaken by dialectologists (e.g., Zamora Vicente, 1967) and by linguists focusing on general patterns of socially-conditioned linguistic variation in Hispanic contexts (e.g., Lipski, 1994; Penny, 2000), but a comprehensive

study of the history of Spanish from a socio-historical perspective has not yet been produced.

Given the immense diversity of the Spanish-speaking world, much socio-historical linguistic work on the topic has focused on a particular variety (e.g., Media Lengua in Ecuador [Muysken 1988], Judeo-Spanish [Penny 1992; Harris 1994], Isleño Spanish [Lipski 1990]), related varieties and geographic zones (e.g., Afro-Hispanic language [Lipski 2005, 2020; Sessarego 2015, 2019], Spanish in the United States [Silva-Corvalán 1994; Lipski 2008], Spanish in contact with Amerindian languages [Granda 1995; Escobar 2012]), or particular phenomena (e.g., clitic pronouns in intense contact situations [Klee 1990], Spanish phonetics and phonology in contact [Rao 2020]). In their important contribution to socio-historical linguistics, Tuten and Tejedo Herrera (2011) contend that interdisciplinary approaches involving historical linguistics, sociology of language, interactional sociolinguistics, and variationist sociolinguistics coupled with advances in quantitative analysis have enabled a deeper understanding of some of the predominant sociolinguistic forces behind historical changes to the Spanish language, including dialect mixing and standardization. Given these interdisciplinary advances, we believe the time is right to bring together innovative scholars investigating specific varieties, regions, and phenomena in the Spanish-speaking world to create a cohesive collection of the most cutting-edge work in socio-historical linguistics across the wide Spanish diaspora.

3. Motivations for the present volume

The present volume took its launch from a special panel organized by the co-editors for the 23rd International Conference on Historical Linguistics in San Antonio, Texas, in 2017. The panel, entitled "Spanish Socio-historical linguistics: Isolation and Contact," drew together some of the world's most respected researchers on Spanish historical, socio-historical, and contact linguistics. Because of its expansive diatopic distribution, the Spanish language serves as a fascinating laboratory for scholars interested in language variation and change within and across societies, and this session sought to examine the crucial role that contact and isolation have played in the development of Spanish varieties worldwide. The event featured presentations by Marcela Flores, María Ángeles Gallego García, Chantal Melis, María Irene Moyna, Malte Rosemeyer, Sandro Sessarego, Álvaro Octavio de Toledo y Huerta, and Rena Torres Cacoullos, who also delivered a plenary lecture for the conference entitled "Contact-induced grammatical change – Far from a foregone conclusion." This panel, situated within the most important international conference on historical linguistics in the world, brought prominence to the special position of Spanish and helped establish Spanish historical linguistics as one

of the major focal points of the event. Covering a wide range of interconnected topics, the panelists explored issues such as the impact of socio-historical factors on structural patterns, for example, the expression of subject pronouns in Spanish and Portuguese *wh*-interrogatives and the accusative/dative opposition in Spanish from a historical perspective. Other papers highlighted the role of particular sociolinguistic factors on the development of the varieties of Spanish, including the linguistic consequences of slavery in the Americas, the evolution of Romance varieties in Al-Andalus, and the importance of child language acquisition in Latin American Spanish variation and change. Alongside these more specific topics, panelists investigated other issues with broader theoretical applications, such as the role of dialectology in historical syntax and the reconceptualization of convergence based on synchronic and diachronic evidence.

Not only did the panel serve as an important nexus for recent work on Spanish socio-historical linguistics, but it also played another significant role: it constituted the centerpiece of the "Tricententennial Panels" of ICHL23, entitled collectively "Las lenguas de San Antonio a 300 años: Reconstructing the linguistic roots of a multicultural city." This multifaceted exploration of the linguistic diversity of San Antonio was designated as an official Tricentennial event. Along with the three other panels (African American language, Texas German, and endangered languages), this event became a central focus of the conference. Given the crucial role of Spanish to the history and social ecology of Texas, a panel examining Spanish as a world language through a socio-historical lens made an essential contribution to the conference and to the field.

In response to the success of the panel, and recognizing the timeliness of the topic, we decided to expand upon the project: we issued a call for papers inspired by the original panel, requesting contributions focused on the crucial role the seemingly antithetical phenomena of contact and isolation as determining factors in the history of Spanish worldwide. Each submission underwent a rigorous peer-review process, and the best papers were selected for inclusion in the present volume. The co-editors wish to gratefully acknowledge the contribution of the external reviewers, whose scholarly generosity ensured the highest-quality final product: Bert Cornillie, Juliana de la Mora, Manuel Delicado Cantero, Manuel Díaz-Campos, Stephen Fafulas, Víctor Fernández-Mallat, Kimberly Geeslin, Melvin González-Rivera, Devin Grammon, Patricia Gubitosi, Silke Jansen, Matthew Kanwit, John Lipski, Jim Michnowicz, Francisco Moreno Fernández, Álvaro Octavio de Toledo y Huerta, Danae M. Pérez, Malte Rosemeyer, Israel Sanz-Sánchez, Armin Schwegler, Scott Schwenter, Sandro Sessarego, Naomi Shin, Sarah Sinnot, Eeva Sippola, Fernando Tejedo-Herrero, and André Zampaulo. This project would not have been possible without their time, effort, and expertise.

4. Chapters within the volume

Following this brief introduction, the full volume consists of eight chapters divided into two sections. The first section, "Socio-historical features in isolation and contact," highlights linguistic phenomena that intersect with socio-historical factors. The second section, "Socio-historical varieties in isolation and contact," primarily focuses on specific varieties' relationship with their socio-historical context. Together, the contributions in this volume shed light on the unique social situations in the Spanish-speaking world that have resulted in disparate linguistic outcomes over time. Each contribution is briefly summarized below.

In the first contribution to Section I, called "Complexification of the early modern Spanish address system: A role for koineization?," Donald Tuten examines the role of socio-cultural factors alongside koineization in the development of the highly complex address system of Castilian. Tuten tracks the rapid rise of *vuestra merced*, especially in 16th-century Peninsular urban centers and the Spanish colonies, where considerable demographic and social mobility existed and where status anxiety fueled a "hyper-salience" of address forms. While socio-cultural factors, then, clearly played a key role in the initial stages of the complexification of the pronominal system, Tuten also sees a significant role for koineization, especially with regard to the rapidity of the actuation of the change: the fast pace of the grammaticalization of these forms in the urban centers of Seville and Madrid and in the colonial communities occurred with little resistance. And, as he insists, it is this rapid grammaticalization of the forms that led to the complexification of the system of address. Tuten thus concludes that koineization had a role to play in the development, but it can best be viewed as a heuristic rather than a predictive device to be used alongside the analysis of socio-demographic and socio-cultural factors.

Next, in "Personal vs. personalized infinitives in Ibero-Romance: Historical origins and contact-induced change," Lamar A. Graham explores the presence of overt subjects with infinitives across several Romance languages. In Portuguese and Galician, on the one hand, inflectional morphology appears on infinitives with overt subjects, e.g. *nós dizermos* 'us to speak,' commonly referred to as personal infinitives. In Castilian and Asturian, on the other hand, overt subjects developed alongside infinitives with no morphological agreement, e.g. *nosotros decir* 'us to speak,' a structure Graham calls the "personalized infinitive." After exploring the uses and distributions of these forms, Graham argues that koineization and language contact have served as catalysts for the evolution of these different forms across Romance languages.

In the following chapter, "Language variation and change through an experimental lens: Contextual modulation in the use of the Progressive in three Spanish dialects," Martín Fuchs and María Mercedes Piñango explore the use of the Simple

Present vs. the Present Progressive in three varieties of Spanish: Mexican Altiplano, Rioplatense, and Castilian. They find that speakers' choice of marker is constrained by their assessment of what is Common Ground between themselves and their addressees and that speakers tend to use the Present Progressive to establish perspective alignment with listeners. The researchers hypothesize that the Present Progressive is preferred, overall, for expressing *event-in-progress* and as a more expressive marker demanding fewer computational resources because it requires less contextual information. The Simple Present, on the other hand, is hypothesized to be generally constrained to situations where the addressee shares perceptual access to the event. Using a real-time comprehension experimental model, the researchers find that, as predicted, subjects take longer to read passages using the Simple Present when shared perceptual access is not specified. They also find variability among the varieties, with Mexican Altiplano more advanced in the diachronic shift of the Progressive to the Imperfective.

In the final contribution to Section I, "Adult language and dialect learning as simultaneous environmental triggers for language change," Israel Sanz-Sánchez and Fernando Tejedo-Herrero examine two changes in the history of Spanish, the simplification of medieval sibilants in Early Colonial Spanish (ECS), and the reconfiguration of the 3rd person clitic system in Medieval Southern Iberian Castilian (MSIC). Their findings point to the gradational, not discrete, relationship of language and dialect, and to the fact that both are similarly affected by linguistic, cognitive, and social factors. The authors explore how the two sociolinguistic ecologies produced different results: with regard to the ECS sibilants, a series of several sibilants was reduced to one, resulting in a simplification of the system, due especially to pervasive contact both in Spain and in the colonies. With regard to the MSIC clitics, on the other hand, a more complex result came about, as case-based contrasts were maintained. The lack of group-enforcing norms resulted in variability, but loose-knit social networks and other social factors in growing urban centers like Seville led eventually to the acceptance of a western-influenced case-based system.

The second section begins with John Lipski's chapter, called "Searching for the sociolinguistic history of Afro-Panamanian Congo speech." In it, he traces the linguistic history of Congo speech along Panama's Caribbean coast, which, according to oral history, was created in order to obscure captive and maroon Africans' speech from their colonial masters. This special language lives on in a series of folkloric practices performed during the Carnival season, and this chapter analyzes both this contemporary variation alongside diachronic developments to paint a more holistic picture of Congo speech's origin and development over time. Lipski uncovers a complex picture that is not in keeping with the explanation preserved in the oral tradition, that the dialect was only a cryptolect used to hide meaning from Spanish colonists. Rather, Lipski outlines a different scenario, namely that Congo

speech once had greater communicative value and mutual intelligibility among its speakers than in contemporary Panama, but its use rapidly declined due to social, demographic, and infrastructure changes.

Next, Sandro Sessarego's chapter, "A sociohistorical perspective on the origin and evolution of two Afro-Andean vernaculars," sheds light on the socio-historical background that gave rise to Choteño Spanish, spoken in Ecuador, and Yungueño Spanish, spoken in Bolivia. Both varieties exhibit unique linguistic features, which Sessarego claims are likely the result of advanced second-language acquisition processes. Contrary to previous views that a Spanish creole formed in the colonial Andes, Sessarego outlines the numerous socio-demographic factors that make a (de)creolization phase unlikely in these regions.

In "*Vamos en Palma* 'we are going to Palma': On the persistence (and demise) of a contact feature in the Spanish of Majorca," Andrés Enrique-Arias explores the Spanish produced by Catalan-dominant bilinguals in Majorca both diachronically and synchronically to determine the likely origin and development of the preposition *en* to convey movement. Enrique-Arias contends that Majorcan Catalan speakers' less complete exposure to Spanish led to the emergence of the feature, which was then recycled within the community for over three centuries, given its relative isolation from monolingual Spanish speakers. However, the island's increased contact with normative varieties of monolingual Spanish has led to the disappearance of this contact feature among more urban, younger speakers.

The final contribution comes from Maryann Parada. In "Anthroponymic perseverance of Spanish vestigial <x>," she examines the distribution of the archaic spelling <x> to represent [x], vs. the spelling <j>, preferred in Peninsular Spanish. The use of <x> turns out to be localized especially to Mexico and to other locales with ties to Mexico, such as the southwestern United States. Parada surveys the distribution of the given name *Ximena*, as well as several surnames such as *Ximénez* and *Mexía*. She finds that the use of the archaic letter in personal names has taken on social significance in Mexico due partly to its connection with indigenous names, such as *Xochitl* and *Xitlali*, but also due to its acquisition of special meaning as an emblem of collective and individual identity, especially Mexican identity. Spanish surnames with <x>, particularly *Ximénez* and *Mexía*, are also localized primarily in Mexico. The author suggests that this higher frequency in surnames may have strengthened the feature in the Mexican public's consciousness, which could also have contributed to the rise of frequency in the name *Ximena*.

5. The intersectionality of isolation and contact

Although the aforementioned chapters discuss a range of linguistic phenomena across several continents and centuries, they are unified by their focus on two potential socio-historical motivations for linguistic change which seem to represent polar opposites: the influence of isolation and the effect of contact. Several contributors focus squarely on the role of isolation: Andrés Enrique-Arias, for example, examines the structural effect of the centuries-long separation of Majorcan Catalan from Spanish. Similarly, Sandro Sessarego points to the special sociodemographics of colonial Andean Spanish, especially this region's non-participation in the large plantation economy, as explanatory of the linguistic outcome in Choteño and Yungueño Spanish. Other authors explore the role of contact on linguistic outcomes in the Spanish-speaking world: Israel Sanz-Sánchez and Fernando Tejedo-Herrero, for example, find the simplification of the sibilant system in Early Colonial Spanish as due largely to contact among Spanish varieties, both on the Peninsula and in the colonies. Martin Fuchs and María Mercedes Piñango likewise account for the variability they document as due potentially to contact among speakers who brought a wide range of peninsular dialects to the colonies.

In each of these papers, and throughout the volume, what emerges is a recognition of the pervasive interaction of socio-historical factors in linguistic variation and change. While isolation and contact can be viewed as major motivators of change, they seldom act alone and should not be viewed simplistically as occupying opposite ends of a continuum. Rather, these factors must be seen as participating in a larger ensemble of interactive factors. The above-mentioned features of Majorcan Catalan studied by Enrique-Arias, for example, developed in relative isolation but were adopted into the Spanish of bilingual speakers through contact; more recently, their diminished use among urbanized youth reflects further contact with normative Spanish. The Afro-Spanish ethnolects described by Sessarego and John Lipski all originated in the context of isolation, but followed different trajectories due to varying socio-historical pressures, the former undergoing contact-induced changes not unlike those found in other enclaves of Spanish, such as the Judeo-Spanish of Istanbul, the latter persisting in small, isolated communities but gaining stature as the ritualized performance language of Carnival.

Given the inherent intersectionality of isolation, contact, and other motivating forces in historical sociolinguistics, related factors that inform the leitmotif of the volume are discussed as well, including the pervasive effects of koineization, the influence of social networks, and the role of identity-building. Donald Tuten, for example, constructs a nuanced chronologization of the development of the complex address system in Spanish, documenting the earlier role of social mobility in shaping the pronouns of address in cities and colonial outposts, but also the eventual

effects of rapid grammaticalization as koineization and the resultant complexification emerged. Lamar A. Graham also sees a major role for koineization alongside contact in the special development of personal and personalized infinitives in the Romance languages.

Similarly, Sanz-Sánchez and Tejedo-Herrero find a lack of group-enforcing norms to be responsible for the variability of the clitic system in Medieval Southern Iberian Castilian, but they note that a growth of urban social networks, along with other factors, led to the final outcome. Parada likewise recognizes the role of contact in the retention of the archaic <x> spelling for names such as *Ximena* and *Ximénez* in Mexico and the southwestern United States but accounts for the adoption as due more pointedly to the forces of identity-building, in part as reminiscent of local indigenous orthographies but also as emblematic of shared Mexican consciousness. The papers collected in this volume, then, contribute not only to the field of Spanish socio-historical linguistics and to the study of particular factors like isolation and contact as instrumental in fostering linguistic change, but they also demonstrate the need for multicausational explanations for change in determining the diachronic trajectory of a language. The broad geographic and sociocultural scope of the Spanish language allows for an extensive comparison of these socio-historical dimensions. We hope to demonstrate in this book that historical linguistics should not be divorced from sociolinguistics, as the intersection of the two fields helps paint a more complete picture of the Spanish language in the past, present, and future.

References

Bartoli, Matteo. 1925. *Introduzione alla neolinguistica (Princìpi – Scopi – Metodi)*. Geneva: Biblioteca dell'Archivum Romanicum II. 12.

Boas, Franz. 1938. *The mind of primitive man*. (Revised edn. 1963). New York: Collier Books.

Cysouw, Michael. 2011. Quantitative explorations of the world-wide distribution of rare characteristics, or: the exceptionality of northwestern European languages. In Horst J. Simon & Heike Wiese (eds.), *Expecting the unexpected: Exceptions in grammar*, 411–432. Berlin: De Gruyter Mouton. https://doi.org/10.1515/9783110219098.411

Dahl, Östen. 1985. *Tense and aspect systems*. Oxford/New York: Blackwell.

Dahl, Östen. 1995. Areal tendencies in tense-aspect systems. In Pier Marco Bertinetto, Valentina Bianchi, Östen Dahl & Mario Squartini (eds.), *Temporal reference, aspect and actionality. Vol. 2: Typological perspectives*, 11–27. Torino: Rosenberg & Sellier.

Dahl, Östen. 2008. An exercise in a posteriori language sampling. *Language Typology and Universals* 61. 208–220.

dal Negro 2004. *The decay of a language: The case of a German dialect in the Italian Alps*. Bern: Lang.

Drinka, Bridget. 2017. *Language contact in Europe: The periphrastic perfect through history*. Cambridge: Cambridge University Press. https://doi.org/10.1017/9781139027694

Escobar, Anna María. 2012. Spanish in contact with Amerindian languages. In José Ignacio Hualde, Antxon Olarrea & Erin O'Rourke (eds.), *The handbook of Hispanic linguistics*, 65–88. Malden, MA & Oxford: Wiley-Blackwell. https://doi.org/10.1002/9781118228098.ch4

Esman, M. J. 1996. Diasporas and international relations. In John Hutchins & Anthony D. Smith (eds.), *Ethnicity*, 316–320. Oxford: Oxford University Press.

Gaeta, Livio. 2020. Remotivating inflectional classes: An unexpected effect of grammaticalization. In Bridget Drinka (ed.), *Historical linguistics 2017*, 205–227. Amsterdam/Philadelphia: Benjamins. https://doi.org/10.1075/cilt.350.10gae

Gehre, Moritz. 1886. *Die deutschen Sprachinseln in Oesterreich*. Großenhain: Hentze.

Granda, Germán de. 1995. El influjo de las lenguas indoamericanas sobre el español. *Revista Andina* 13(1). 173–198.

Harris, Tracy K. 1994. *Death of a language: The history of Judeo-Spanish*. Newark, DE: University of Delaware Press.

Hauffen, Adolf. 1895. *Die deutsche Sprachinsel Gottschee*. Graz: K.K. Universitäts-Buchdruckerei und Verlags-Buchhandlung 'Styria.'

Heine, Bernd & Tania Kuteva. 2005. *Language contact and grammatical change*. (Cambridge Approaches to Language Contact). Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511614132

Hernández-Campoy, Juan Manuel & Juan Camilo Conde-Silvestre (eds.). 2012. *The handbook of historical sociolinguistics*. Oxford: Wiley-Blackwell. https://doi.org/9781118257227

Hickey, Raymond (ed.). 2019. *Keeping in touch: Emigrant letters across the English-speaking world*. (Advances in Historical Sociolinguistics 10.) Amsterdam: John Benjamins. https://doi.org/10.1075/ahs.10

Jahr, Ernst Håkon (ed.). 1999. *Language change: Advances in historical sociolinguistics*. Mouton de Gruyter. https://doi.org/10.1515/9783110807653

Klee, Carol A. (1990) Spanish-Quechua language contact: The clitic pronoun system in Andean Spanish. *Word* 41(1), 35–46. https://doi.org/10.1080/00437956.1990.11435812

Leck, Hans. 1884. *Deutsche Sprachinseln in Wälschtirol: landschaftliche und geschichtliche Schilderungen*. Stuttgart: Karl Aues Verlag.

Labov, William. 1963. The social motivation of a sound change. *Word* 19(3). 273–309. https://doi.org/10.1080/00437956.1963.11659799

Labov, William. 2003. Pursuing the cascade model. In David Britain & Jenny Cheshire (eds.), *Social dialectology: In honour of Peter Trudgill*. Amsterdam/Philadelphia: John Benjamins. https://doi.org/10.1075/impact.16.03lab

Labov, William. 2010. *Principles of linguistic change. Vol. 3: Cognitive and cultural factors*. Malden, MA/Oxford: Wiley-Blackwell. https://doi.org/10.1002/9781444327496

Lippi-Green, Rosina. 1997. *English with an accent: Language, ideology, and discrimination in the United States*. London: Routledge.

Lipski, John. 1990. *The language of the Isleños: Vestigial Spanish in Louisiana*. Baton Rouge: Louisiana State University Press.

Lipski, John. 1994. *Latin American Spanish*. London: Longman.

Lipski, John. 2005. *A history of Afro-Hispanic language: Five centuries, five continents*. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511627811

Lipski, John. 2008. *Varieties of Spanish in the United States*. Washington, DC: Georgetown University Press.

Lipski, John. 2010. Spanish and Portuguese in contact. In Raymond Hickey (ed.), *The handbook of language contact*, 550–580. Malden, MA/Oxford: Wiley-Blackwell. https://doi.org/10.1002/9781444318159.ch27

Lipski, John. 2020. *Palenquero and Spanish in contact: Exploring the interface*. Amsterdam & Philadelphia: John Benjamins. https://doi.org/10.1075/coll.56

Milroy, James & Leslie Milroy. 1985. Linguistic change, social network and speaker innovation. *Journal of Linguistics* 21. 339–384.

Muysken, Pieter. 1988. Media Lengua and linguistic theory. *Canadian Journal of Linguistics* 33(4). 409–422. https://doi.org/10.1017/S0008413100013207

Nevalainen, Terttu & Helena Raumolin-Brunberg (eds.). 1996. *Sociolinguistics and language history: Studies based on the Corpus of Early English Correspondence*. Amsterdam/Atlanta: Rodopi.

Nevalainen, Terttu, Minna Palander-Collin & Tanja Säily (eds.). 2018. *Patterns of change in 18th-Century English: A sociolinguistic approach*. (Advances in Historical Sociolinguistics 8.) Amsterdam: John Benjamins. https://doi.org/10.1075/ahs.8

Nichols, Johanna. 1992. *Linguistic diversity in time and space*. Chicago: University of Chicago Press. https://doi.org/10.7208/chicago/9780226580593.001.0001

Penny, Ralph. 1992. Dialect contact and social networks in Judeo-Spanish. *Romance Philology* 46. 125–139.

Penny, Ralph. 2000. *Variation and change in Spanish*. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9781139164566

Rampton, Ben. 1995. *Crossing: Language and ethnicity among adolescents*. London: Longman.

Rao, Rajiv (ed.). 2020. *Spanish phonetics and phonology in contact: Studies from Africa, the Americas, and Spain*. Amsterdam & Philadelphia: John Benjamins. https://doi.org/10.1075/ihll.28

Romaine, Suzanne. 1982. *Socio-historical linguistics: Its status and methodology*. (Studies in Linguistics 34.) Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511720130

Säily, Tanja, Arja Nurmi, Minna Palander-Collin & Anita Auer (eds.). 2017. *Exploring future paths for Historical Sociolinguistics*. (Advances in Historical Sociolinguistics 7.) Amsterdam: John Benjamins. https://doi.org/10.1075/ahs.7

Sandfeld, Kristian. 1930. *Linguistique balkanique: Problèmes et resultants*. Paris: Champion.

Schmidt, Johannes. 1872. *Die Verwandtschaftverhältnisse der indogermanischen Sprachen*. Weimar: Böhlau.

Schuchardt, Hugo. 1882–3. *Kreolische Studien I–VI*. Sitzungberichte der philosophisch-historischen Klasse der kaiserlichen Akademie der Wissenschaften in Wien.

Schuchardt, Hugo. 1884. *Dem Herrn Franz von Miklosich zum 20. November 1885: Slawo-deutsches und Slawo-italienisches*. Graz: Leuschner & Lubensky.

Schreier, Daniel. 2017. Dialect formation in isolated communities. *Annual Review of Linguistics* 3. 347–362.

Sessarego, Sandro. 2015. *Afro-Peruvian Spanish: Spanish slavery and the legacy of Spanish creoles*. Amsterdam/Philadelphia: John Benjamins. https://doi.org/10.1075/cll.51

Sessarego, Sandro. 2019. *Language contact and the making of an Afro-Hispanic vernacular: Variation and change in the Colombian Chocó*. Cambridge: Cambridge University Press. https://doi.org/10.1017/9781108661782

Silva-Corvalán, Carmen. 1994. *Language contact and change: Spanish in Los Angeles*. Oxford: Oxford University Press.

Thomason, Sarah Grey & Terrence Kaufman. 1988. *Language contact, creolization, and genetic linguistics*. Berkeley, Los Angeles, London: University of California Press.
Trudgill, Peter. 1983. *On dialect: Social and geographical perspectives*. New York & London: New York University Press.
Trudgill, Peter. 1986. *Dialects in contact*. Oxford: Blackwell.
Tuten, Donald. 2003. *Koineization in Medieval Spanish*. Berlin: de Gruyter Mouton. https://doi.org/10.1515/9783110901269
Tuten, Donald & Fernando Tejedo Herrero. 2011. The relationship between historical linguistics and sociolinguistics. In Manuel Díaz Campos (ed.), *The handbook of Hispanic Sociolinguistics*, 283–302. Malden, MA & Oxford: Wiley-Blackwell. https://doi.org/10.1002/9781444393446.ch14
Wolfram, Walt, Kirk Hazen & Natlie Schilling-Estes. 1999. *Dialect change and maintenance on the Outer Banks*. (Publication of the American Dialect Society [PADS] 80). Tuscaloosa, AL: University of Alabama Press.
Wolfram, Walt & Natalie Schilling-Estes. 1998. *American English: Dialects and variation*. Oxford: Blackwell.
Zamora Vicente, Alonso. 1967. *Dialectología española*, 2nd edition. Madrid: Gredos.

SECTION I

Socio-historical features in isolation and contact

Complexification of the early modern Spanish address system: A role for koineization?

Donald N. Tuten
Emory University

The Spanish system of address pronouns suffered a dramatic increase in complexity between the 15th and 17th centuries. This study aims to explain how and why this change (or set of changes) took place when and where it did by using the model of koineization as a heuristic device. It argues that sociocultural factors (particularly widespread social status anxiety) and the salience of address forms played a primary role in driving the changes forward. It also argues, however, that sociodemographic factors associated with koineization (demographic movement and dialect mixing) contributed significantly to the timing, rapidity, location, and specific outcomes, including the grammaticalization of *vuestra merced* to *usted* and numerous other reduced forms.

Keywords: actuation, koineization, new-dialect formation, heuristic, address forms, grammaticalization, complexification, Spanish, *vuestra merced, usted*

1. Introduction

In the 16th and early 17th centuries, the Spanish of Castile and the Indies underwent radical and relatively rapid changes, which included cases of simplification (e.g., the reduction in inventory of sibilant phonemes) as well as at least one case of complexification: that of the system of address. In this latter process, the medieval singular pronouns *tú* and *vos* were supplemented by a range of originally indirect options, particularly *vuestra merced*, its many reduced, grammaticalized forms (e.g., *voacé, vusted*), as well as others such as *él/ella* and use of null subject third-person verb forms.[1]

1. I would like to thank Peter Trudgill, Israel Sanz-Sánchez, Barbara Johnstone, three anonymous reviewers and the editors for valuable feedback on an earlier draft of this chapter. I have tried to respond effectively to their valuable suggestions. Any remaining infelicities are, of course, my own.

https://doi.org/10.1075/ahs.12.c02tut

In this study, I argue that sociocultural factors motivated the adoption of new address forms, but I also argue that early modern koineization in Castile and the Indies played a key role in the timing, rapidity, location, and specific outcomes of the complexification of the address system. This claim may be surprising, given that koineization is generally understood as dialect mixing that leads to formation of a new variety (koine) characterized by mixing, leveling to majority variants, and simplification. It is my view, however, that a certain type and degree of complexification is not inconsistent with koineization, particularly if we view koineization as a heuristic of change.

The study is organized into five sections. First, I present a brief overview of the development of the model of koineization, particularly in the work of Peter Trudgill. Since the 1980s, Trudgill has moved from a broad conception of koineization as a quasi-predictive model of language change to a more narrowly constrained, deterministic model. I argue, however, that we may also view the model of koineization as a heuristic device. As a heuristic, the model can incorporate a broader range of causal factors (particularly sociocultural factors) in the development of explanatory accounts of a broader array of potentially relevant changes within a broader array of sociodemographic contexts. In the second section, I briefly review evidence for koineization in early modern Spanish-speaking societies of Castile and the Indies, examining demographic movement and dialect mixing, as well as a prototypical case of koineizing change: simplification of the inventory of sibilant phonemes. In the third section, I examine change in address forms in koineizing and other speech communities, arguing that these forms are among those most likely to be "noticed", assigned sociocultural indexical value, and used intentionally by speaker-learners. In the next two sections, I outline the complexifying changes that took place in the early modern system(s) of address, and then sketch my account of how and why these changes took place when and where they did, using the model of koineization as a heuristic to frame my analysis.

2. What is koineization?

Koineization is a historical sociolinguistic model of language change that has been developed since the 1980s (Siegel 1985; Trudgill 1986) as one means of responding to the actuation question: "Why do changes in a structural feature take place in a particular language at a particular time, but not in other languages with the same feature, or in the same language at other times?" (Weinreich, Labov & Herzog 1968: 102). Koineization is generally understood to be a particular kind of dialect contact and mixing that leads to rapid linguistic change (i.e., over only one or two generations) and formation of a new mixed and simplified variety. It occurs

when speakers of mutually intelligible dialects/varieties come together during settlement of new communities. Prototypical koineizing communities include new towns (Kerswill 1996), frontier zones (Tuten 2003), and colonies (Trudgill 2004), but some scholars have also included communities that have been "swamped" by a large majority of settlers of mixed background (Penny 2000: 43). In koineizing communities, settlers of diverse dialectal origin form looseknit social networks (Milroy & Milroy 1985), which allow relatively unfettered patterns of interaction that lead, in turn, to an initial peak of "disorderly" variation (for features that differ between contributing varieties) at the same time that norm enforcement mechanisms weaken.

Within such communities, mutual convergence among speaker-learners is normally seen as the primary mechanism by which majority and simpler features are selected, made more frequent, and acquired. In Trudgill's (1986) initial discussion of koineization, both adults and children are assumed to engage in mutual convergence or accommodation (Giles 1973; Gallois, Ogay & Giles 2005) as they abandon socially marked and minority features and favor majority and/or simpler features. Although Giles' rational choice model of accommodation is fundamentally social and presents convergence and divergence as goal-oriented behaviors, convergence has always been seen as the default option, and it is this "set to converge" that Trudgill has emphasized from his earliest discussion of koineization.

With regard to linguistic outcomes, koineization has been understood to lead to mixing (i.e., selection of features from different contributing varieties), selection of majority features from within the overall feature pool, reallocation (selection of two or more variants from different varieties but with new phonological, grammatical or socio-stylistic functions), and simplification (Trudgill 1986; Kerswill 2013). Simplification has in practice been understood as reduction in units (bit complexity) and rules (algorithmic complexity), although this reduction is much more limited than that which is found in pidgins and creoles (cf. Tuten 2001).[2] Complexification is not generally considered a possible outcome of koineization, though reallocation seems to allow for (or be analyzable in terms of) increases in bit complexity, since this outcome allows retention of diverse dialectal forms (as

2. Mufwene et al. (2017: 4–5) refer to changes in inventories of units (e.g., words, morphemes, phonemes, allophones) as changes in "bit complexity", and to changes in inventories of rules (or relative description length) as changes in "algorithmic complexity". Although Mufwene et al. argue that there are other types of complexity that must be taken into account in linguistic research, these types do capture most identified outcomes of simplification in koineization. Trudgill (2011: 20–32) does not use these concepts in his discussion of simplification and complexification, but they appear to capture the specific types of change that he refers to (e.g., changes in regularity, transparency, redundancy).

well as inclusion of new forms generated/adopted by speaker-learners), each with new and different phonological, grammatical, semantic, or socio-stylistic functions.

This "general" model of koineization as described presents change in koineizing contexts as predictable, though not completely so, since it still recognizes a possible role for sociocultural factors in the selection of (even minority) variants (cf. Kerswill 2010; Britain 2012). However, in the past two decades, Trudgill (e.g., 2000, 2004, 2008, 2018, forthcoming) has argued consistently for a more constrained version of the model, which he refers to as the *deterministic model of new dialect formation*. According to this, certain outcomes (in phonology and grammar) are predetermined when certain conditions attain. Relevant communities are limited only to what Trudgill terms *tabula rasa* settler colonies, where settlers (who all speak mutually intelligible varieties and interact minimally with any local indigenous communities) are far removed from established uses and social meanings of linguistic forms in their communities of origin (e.g., colonial New Zealand). In such communities, the actions and beliefs of adults are held to be of little importance. Rather, children (up to age 8) are seen to play the key roles in processes of selection and focusing of new community norms, for they are generally less influenced by sociocultural indexical values (e.g., prestige, stigma) and in a mixed community they lack consistent models for both forms and their indexical values.

Trudgill sees children's acquisition and use as dependent on an automatic process of mutual alignment. Unlike Giles' notion of accommodation, which recognizes the social agency of speaker-learners and the possibility that they will diverge rather than converge, models of automatic alignment (e.g., Cappella 1997; Pickering & Garrod 2004, 2017) claim that participants in dialogue automatically align their cognitive representations and production on a number of levels: situational, semantic, lexical, syntactic and phonological-phonetic.[3] Trudgill has identified automatic alignment as the key process of feature selection because studies of koineization, particularly in relatively isolated settler communities such as colonial New Zealand, have demonstrated a consistent preference for selection of majority variants from within the pre-koine feature pool, independent of who uses them and the indexical values they may have in home communities and among the first generation of immigrants. In other words, this evidence suggests that children in *tabula rasa*

3. Alignment occurs automatically and unconsciously because it allows participants to establish an "implicit common ground," favoring mutual understanding, and because it reduces cognitive load, favoring rapid interaction. Pickering & Garrod (2017) propose that automatic "interactive alignment" may be responsible for fixing routines in the language of a community of speakers, and that alignment is more highly predictive of the behavior of (older) children than of adults (cf. Garrod and Clark 1993), who have learned or "overlaid" socially oriented strategies of non-alignment (divergence) on the basic alignment mechanism (e.g., for marking of identities, attitudes, or stances different from those of interlocutors).

communities align unavoidably to variants with highest token frequency (including simplified features generated by child speaker-learners themselves), though in Trudgill's view final selection or focusing requires at least two generations.[4] Since prestige and stigma, which might otherwise lead speaker-learners to favor minority or marked variants, seem to play little or no role in outcomes, Trudgill has argued insistently that sociocultural factors (identity, attitudes, ideologies) are irrelevant to the process and outcomes of koineization in *tabula rasa* communities.[5]

Even so, most scholars (e.g., Kerswill 2010, and from a broader perspective, Bell 2016: 397–399) have continued to argue for the need to recognize the possible impact of sociocultural factors on the outcomes of koineization, even in narrowly defined *tabula rasa* communities. In general terms, Gordon et al. (2004) argue for multiple causation, or the interaction of sociodemographic frequency-based effects with other sociocultural factors. More specifically, Schreier (2008, 2012) argues that socio-stylistic reallocation (originally discussed in Trudgill 1986) cannot be due only to frequency effects. This is illustrated in Siegel's (1997: 127) discussion of reallocation of regional variants as stylistic variants. For instance, there exist two forms of the third person possessive in Fiji Hindi: /uske/, the general, prestige variant, and /okar/, a low-style variant used to represent "country bumpkin" characters. Significantly, the first is taken from generally more prestigious Hindi varieties and the second from generally less prestigious Bhojpuri varieties. In cases such as this, older children and adolescents of the second and third generations were clearly able to notice and learn the indexical values that the two forms seemed to have in the context of their new community. This case also illustrates how reallocation can lead to an increase (albeit minimal) in "bit complexity" (i.e., the total number of variant forms that are available for certain grammatical and communicative functions).

Ultimately, Trudgill himself recognizes the possibility that sociocultural factors may have an impact even in *tabula rasa* communities: "Convergence is the default, even if attitudinal factors can on occasion lead to this being overridden"

---

4. This differs from findings in some studies of new towns. For instance, in a study of phonological change in the new town of Milton Keynes, Kerswill (1996) and Kerswill and Williams (2000) found that the second generation (or first generation of children and adolescents) were already focusing new norms, but these matched and were influenced by a broader regional process of supralocalization.

5. Reactions to this argument have been varied and intense (see, for instance, the responses to Trudgill 2008 in *Language in Society*, volume 37). Some scholars have denied outright that anything like a *tabula rasa* settler community is possible, since first-generation settlers inevitably bring sociocultural values with them. On the other hand, Kerswill (2010, 2013) recognizes that sociocultural factors may have only a weak influence on the outcomes of koineization in *tabula rasa* communities, and that the deterministic model accounts well for (most of) the phonological outcomes of New Zealand English analyzed in Trudgill (2004).

(Trudgill 2018: 279). It is important to note too that Trudgill excludes lexicon (including, presumably, address pronouns) from his claim that sociocultural factors are irrelevant, apparently because lexical items are easily learned and changed by all speaker-learners, including adults. For this reason, he has often focused his attention on change in phonetics and phonology, aspects of language which are perhaps less susceptible to "noticing" and intentional manipulation (see below).

However valuable "general" accommodation-based or "deterministic" alignment-based models may be for accounting for changes in certain communities, they have both tended to prioritize predictability as the primary goal of model building, and this has meant a lesser or greater disregard for the possible impact of sociocultural factors on change in koineizing communities. Nevertheless, recognition that sociocultural factors may have an impact, albeit exceptionally, even in *tabula rasa* communities, suggests that predictability in language change cannot be absolute and that we would benefit by viewing the model in different terms. If we understand koineization not so much as a predictive model but rather as a heuristic device, we may find it of greater use in the development of explanatory accounts of change in a broader range of real-world contexts and with regard to a broader range of phenomena.

As a heuristic device, koineization may be seen as an artificial construct or prototype, derived from extant empirical research, which serves as a means of defining fixed or frequent characteristics of the general phenomenon of dialect mixing (these characteristics include a set of sociodemographic and psychosocial factors that define contexts of interaction and learning, as well as typical outcomes). As such, the model of koineization can serve as a benchmark or point of comparison, around or within which can be situated usual and unusual features of community contexts, as well as expected and unexpected outcomes. Hence, the model of koineization may serve as a starting point for analysis, leaving open the possibility of its application to a wider range of communities and outcomes, as well as the inclusion of other causal factors or other models of change when data suggests the need to do so. Indeed, use of the model as a heuristic constitutes a method, in that it provides a set of topics for systematic consideration (cf. Johnstone 2018: 8–9). In general, then, heuristic use of the model requires examination of actual community contexts and linguistic outcomes and a determination of their closeness to or distance from prototypical contexts and outcomes of the model. To the degree that contexts and outcomes differ from the expected, we will need to consider the impact of a broader range of contributing causal factors.

In what follows, I use the model of koineization as a heuristic with the aim of developing a more cogent and plausible explanatory account of change in the early modern Spanish address system. Moving away from a primary concern with predictability allows for exploration of the possible impact of koineization

in "less than prototypical" koineizing communities, such as the "swamped" cities of early modern Seville and Madrid, and New World colonies such as Mexico, where Spanish-speaking settlers mixed with large majorities of speakers of other languages. It also allows for closer consideration of the consequences of the actions and beliefs of adult speaker-learners, the kinds of features they are likely to "notice" and use intentionally, as well as the sociocultural factors that influence their understandings and choices. Finally, it opens the possibility of linking koineization more explicitly to a particular type of complexification. As I have indicated above, reallocation as a process and outcome of koineization leaves open the logical possibility of complexification, but this is not an avenue that previous research has aimed to explore.

3. Koineization in early modern Spain and the new world

Penny (1987) argued that one of the major consequences of the events of 1492 was the foundation of new communities, with widespread and intense demographic movement and dialect mixing. These processes created contexts ideal for koineization, not only in the New World colonies but also within some areas and cities of Castile itself. In the New World, Castilians began to establish permanent settlements in the Caribbean islands and Panama from 1498, in New Spain (Mexico and Guatemala) from 1519, and in Peru from 1531. In these colonies, Castilian/Spanish was the language of administration, trade, written communications involving Europeans and their descendants, as well as the (or at least a) dominant language of everyday interaction in colonial cities where Europeans tended to concentrate (e.g., Mexico and Lima). Though we have little information about the exact origins of settlers in particular communities, the great majority of European settlers were Castilians of diverse dialectal background, from both the north and south of Castile, though persons from Andalusia and other southerly regions of Castile were predominant overall (Boyd-Bowman 1976).

Within Castile itself, the recently conquered kingdom of Granada became a target of colonial settlement, but of greater overall importance as leaders of change and centers of diffusion were Seville and Madrid. From 1503, Seville was given much administrative control of the New World colonies, with a monopoly on trade and shipping, as well as a monopoly on emigration to and return from the Indies, all of which effectively made the city into the commercial capital of the growing empire. Later, in 1561, Philip II chose the relatively small, central-northern Castilian city of Madrid as the seat of his court, converting it suddenly into the de facto political capital of the empire. As was the case in the New World colonial cities, both Seville and Madrid saw rapid increases in population with mixing of speakers of diverse

dialectal and social origin. Indeed, although these were not new cities, the increases were so great that local populations were effectively "swamped" by the newcomers of diverse origin. For instance, in 1530, Seville's population included some 45,000 inhabitants, but by 1600 it had grown to anywhere between 100,000 and 135,000 (Poussou 1997: 264). However, demographic mobility was much greater than these figures indicate, since between 1498 and 1600 (and even after) large numbers of emigrants (and small numbers of returnees) departed from or moved through the city (often after stays of months or years). In addition, the city suffered numerous plagues which increased the death rate. Exact estimates are impossible to make, but at least 250,000 emigrants, and possibly many more, left for the New World during the 16th century (González Martínez 2002: 89, Casey 1999: 25). From 1561 Madrid too saw massive and rapid increases in population, growing from an estimated population of 15,000 inhabitants in 1561 to anywhere between 65,000 to 100,000 inhabitants by 1600 (Poussou 1997: 264), the great majority of whom came from central-northern areas of Castile (Penny 2002a).[6]

In Seville and Madrid, these demographic changes led to conditions favoring koineization: the creation of looseknit social networks, and a breakdown of sociolinguistic norm enforcement with a concurrent peak in "disorderly" sociolinguistic variation. Indeed, it seems highly likely that the final collapse of the medieval system of sibilant phonemes, generally dated to the 16th century (Penny 2002b: 99), is attributable to early modern processes of koineization (at least in Castile). In Madrid and the north, where dialectal mixing was less diverse, the six medieval sibilant phonemes (dentoalveolar /s̪/ /z̪/, apicoalveolar /s̻/ /z̻/, prepalatal /ʃ/ /ʒ/) were reduced through processes of devoicing to just three (/s̪/, /s̻/, /ʃ/). In Seville (and much of Andalusia, including part of the old kingdom of Granada) and in the New World colonies, the six sibilant phonemes were reduced to just two (/s̻/, /ʃ/), through devoicing and *seseo*, or merger of dentoalveolar and apicoalveolar sibilants.

In an important study, Sanz-Sánchez (2019) analyzes the development of these simplifications in the New World as beginning with the mixing of diverse inputs from the first generation(s) of settlers who came from different areas of Castile, so that the New World simplifications were not imported as completed changes from Seville, as has often been assumed (cf. Penny 2002b: 101) but rather developed concurrently with those in the Peninsula. These settlers included only a tiny minority

6. Unlike Lass (1990), I use "swamping" to refer to a massive influx of settlers of diverse dialectal origin that effectively overwhelms the original local population and leads to koineizing changes in the local variety. Swamping may lead to weakening or even erasure of the Founder Effect (Mufwene 1996) in communities affected by it, and it may favor and accelerate changes that were only marginal in the original community, as was apparently the case with *seseo* in early 16th-century Seville.

of users of the full six-phoneme medieval system, and a large majority of users of varied, partly simplified versions of that system. These included sibilant devoicing in one, two, or all pairs of phonemes (especially typical of extreme northern regions near Burgos but already widespread for at least one point of articulation), and possible merger of dentoalveolars and apicoalveolars, which had begun to spread in Andalusia. These input mixes would have been similar in Seville and in New World cities. In Seville and Madrid, where native speakers of Castilian and related varieties clearly predominated, the prototypical model of koineization appears to account well for the final collapse of the medieval sibilant system (though, again, in Madrid, where settlers and inputs were primarily of northern origin and therefore less diverse, simplification was less extreme). In contrast, and despite the similarity of outcomes in Seville/Andalusia and the New World colonies, Sanz-Sánchez (2016, 2019: 30–33) argues that the "classical" model of koineization is too narrow to account for the colonial outcomes, where native speakers of diverse varieties of peninsular Spanish were a tiny minority in relation to the much larger communities of indigenous inhabitants, enslaved Africans, mestizos and mulattos, whose learning and interactions must be factored into any account of the development of new varieties of Spanish there.

Sanz-Sánchez is undoubtedly right to critique the expectation that application of a highly-constrained model of koineization to all of the New World colonies can account adequately for actual conditions and outcomes, though he does recognize some communities (e.g., colonial New México) where conditions and outcomes appear to have corresponded more closely to those which are typical of koineization (cf. Sanz-Sánchez 2016). However, if we view the model as a heuristic, we may begin by asking to what degree New World conditions and outcomes correspond to those typical of koineization (rather than aiming to declare it relevant or irrelevant). Certainly, within the New World colonies, mixes of adult speakers of various peninsular varieties provided initial models for children who learned Spanish as a "native" language as well as for adults who learned Spanish as a second language, though children would have been influenced by both peninsular and local non-native models. In this sense, we can say that koineization contributed to the development of simplifying changes in the colonies, but that any thorough account of changes there must also attend, at the very least, to processes and effects of language contact and shift, as Sanz-Sánchez argues.

4. Forms of address in koineization

The simplification of address systems has been identified as a possible outcome of koineization. For instance, Mesthrie (1993: 40) identified the general loss of respect features, including deferential address pronouns, as an outcome in South African Bhojpuri.[7] But how might it be possible for an address system to increase in complexity in koineizing communities (such as those of the early modern Spanish empire)? The answer to this question is not immediately evident, since in the typical "disorderly" pre-koine community, correlations and the indexical links that depend on them cannot be easily identified or learned for highly variable features. Their social meanings are neutralized (Mesthrie 1994) and made highly idiosyncratic, and therefore speaker-leaners tend, overall, to align to the most high-frequency forms, generally understood as either pre-koine majority features or simpler features that they themselves generate.

However, as the case of socio-stylistic reallocation in Fiji Hindi makes clear, there are exceptions to this dominant tendency. These exceptions are very likely limited to forms and features that come to be "noticed" (Johnstone 2013) or, as discussed by other scholars, that become "sociolinguistically salient" (Trudgill 1986; Kerswill & Williams 2002; Auer 2014). Johnstone emphasizes that many if not most linguistic features are *not* noticed, and adds: "In order for a feature that is used in a particular context to evoke or create that context, people have to notice the correlation and attribute meaning to it" (2013: 69). In principle any feature can become "noticed" or "salient", although it may be true that lexical items are generally more available for "noticing" than fine-grained phonetic features.

How, then, do speaker-learners – in general and in koineizing contexts in particular – come to notice specific forms/features? Johnstone (2013) argues that connections between linguistic form and meaning can be highlighted as people interact, through the use of explicit metatalk (as when people complain about being addressed with one address form rather than another) but also through more implicit means, such as text metricality:

> Because…items [including forms of address –DNT] occur in the same context we are led to compare or contrast them. People collaboratively create text metricality in everyday talk, reusing bits of others' sounds and grammar so that the places where their contributions differ – the places where they are contributing something new – pop out…it can thus create indexical links between form and meaning.
>
> (Johnstone 2013: 110)

7. It should be noted, however, that Mesthrie (1993) ties loss of the respect feature to the loss of the need to use that feature within new, socially mixed and leveled communities.

Noticing of features is more likely to happen as speakers enter into contact with speakers from different communities. Importantly, once noticed and enregistered (i.e., assigned indexical values), forms can be used, exaggerated in frequency (or degree), or avoided. In other words, they can be taken up as sociolinguistic resources and used intentionally to evoke identities, personas, and situations.[8]

Any feature may come to be noticed, but pronouns of address are especially likely to be highlighted in this way, as Johnstone makes clear in her discussion of Pittsburghese *yinz* and US Southern *y'all/yall:*

> …personal pronouns occur frequently in talking and writing, and they both reflect and reinforce distinctions among categories of people that can easily spark controversy …personal pronouns have a history of becoming controversial because of the social meaning with which they become enregistered.
>
> (Johnstone 2013: 232)

Agha (2007) too points to reasons for what we might call the "sociocultural hyper-salience" of address forms:

> …there are few linguistic categories for which explicit metapragmatic reports are more ubiquitous and readily available than for person referring terms. Person deictics are relatively salient to native metapragmatic scrutiny because they exhibit a dense overlap of functional properties: they refer to the very persons picked out as the focus of stereotypic social indexicality. They formulate reference to the very individual to whom deference is paid. (Agha 2007: 280)

These observations have four key implications for change in address forms/systems. First, terms of address are more likely to be intentionally manipulated by speakers than are many other features. Second, this is particularly so in communities that require asymmetrical displays of deference and superiority, where divergence in address (in otherwise similar contexts) is actually necessary for social integration (see below). Third, this noticing and manipulation is more likely for adult (and adolescent) speakers, who have a much broader range of social experiences and contacts than do children. Fourth, noticing and manipulation by adult speakers

8. Given space limitations, I make explicit reference here only to a basic notion of indexicality. A more detailed presentation would include discussion of the fundamentally diachronic notion of indexical order (Silverstein 2003), the synchronic notion of indexical field (Eckert 2008), as well as a dynamic version of the hierarchical model of the indexical field developed by Sinnott (2013) in her analysis of the modern peninsular address forms *tú* and *usted*. Sinnott's model ties indexicality to pragmatics and usefully ranks meanings from "default" (most tied to form, which I refer to as "prototypical") to "nonce" (most heavily reliant on contextual information). All of these models of indexical meaning inform and/or are largely consonant with my analysis of change in address in early modern Spanish (see following sections).

may favor adoption of new forms of address and increases in frequency of those forms, which may in turn lead to grammaticalization of lexical items/sequences by younger generations and eventual complexification of the address system.

5. Complexification of the early modern Spanish address system

The development of early modern Spanish pronouns of address has been the object of much research. Scholars have used a wide array of evidence (e.g., letters, literary texts, legal proceedings, metapragmatic comments) to explore morphological development of address forms and formulae (e.g., Pla Cárceles 1923; Líbano 1991; De Jonge 2005; De Jonge and Nieuwenhuijsen 2009) as well as patterns of use in specific texts and contexts (e.g., Ly 1981; Eberenz 2000; Anipa 2001: 187–228, Bentivoglio 2002; Hammermüller 2010; Leon 2011; Hidalgo 2018). Though much work remains to be done in these areas, we now have a reasonably good understanding of their morphological and pragmatic development (i.e., what changes occurred in form and use). Much less attention has been directed at answering the broader question of why and how the changes in address forms and use occurred when and where they did. Indeed, León (2011: 173) has characterized the actuation of these changes as a sociolinguistic enigma. In this section, I review the complexifying changes in singular address forms and pronoun systems.

Modern Spanish is characterized by a number of singular address pronoun systems.[9] In México, the Caribbean, and Spain, *tú* 'you.SG' has conventionally been used to index solidarity with or superiority to addressee, while *usted* 'you.SG' has conventionally indexed deference or distance to addressee.[10] The pronoun *vos* 'you.SG' also appears in many varieties of South and Central America. In Argentina *vos* generally substitutes *tú* and contrasts with *usted*, while in Cali, Colombia a three-way contrast between *tú*, *vos* and *usted* survives (Newall 2016). In some

9. I do not discuss plural forms here, since their inclusion would require more extended discussion without substantially altering the main arguments.

10. In using the conventional parameters of power-status (equal vs. unequal relations) and distance/solidarity (Brown and Gilman 1960; Brown and Levinson 1987), I do not intend to imply that the potential indexicalities of address forms are limited to these values. As Agha (2007: 279) argues, the indexicalities of deictic forms are limitlessly varied. Nevertheless, as the same scholar argues, "person-referring terms" are also highly organized and non-random in social life, and the particular history of the Spanish address forms attests to the central importance of the parameters of power and distance in their development. Neither do I want to claim that more parsimonious analyses of default or prototypical meanings are not possible (cf. Sinnott 2013, who argues that social indexical values of modern Peninsular *tú* and *usted* depend on and derive from a consistent, marked association of *usted* with 'distance').

varieties of Colombia (including Bogotá up to the late 20th century), *su merced* 'your grace' has continued in use alongside other forms (cf. Uber 2011; Placencia 2010; García-Godoy 2019). All of these pronoun systems are distillations of much more complex early modern systems.

Early modern Spanish systems developed from the address pronoun system of Latin. Classical Latin made a simple distinction between TŪ 'you.SG' and VŌS 'you.PL'. In Late Imperial Latin, VŌS retained its general use for plural reference, but came to be used in the singular to index deference/distance (using a common pluralization strategy to mark deference; cf. Brown and Levinson 1987), while TŪ was left to index superiority/solidarity. This system was maintained in most varieties of medieval Romance, including medieval Castilian:

– *tú* 'you.SG' for superiority/solidarity
– *vos* 'you.SG' for deference/distance

Modern French *tu*/*vous* has maintained this system to this day, whereas it was much altered during the early modern period in Spanish. Change in Spanish singular forms began in the 15th century with the introduction of numerous new forms of deferential, indirect address among the court elite. These new forms included *vuestra*, the (feminine) singular possessive of *vos*, with a (feminine) noun originally specifying a praiseworthy attribute of the addressee. These forms included the following, all of which were first used in address to the monarch and subsequently to other members of the court elite (Ly 1981):

vuestra majestad 'your majesty'
vuestra alteza 'your highness'
vuestra excelencia 'your excellency'
vuestra señoría 'your lordship/ladyship'
vuestra merced 'your grace'

During the course of the 16th century, the system of singular forms of address became progressively more complex, coming to include the following forms by the end of that century, which are listed with brief indications of their most conventional (though often changing) indexical values:

– *tú* 'you.SG' or null subject second person singular verb forms (*cantas* 'you.sing'): superiority/solidarity;
– *vos* 'you.SG' or null subject second person plural verb forms (*cantades*/*cantáis* 'you.sing'): changing from deference/distance to superiority/solidarity during 16th century;
– *él*/*ella* 'he/she' > 'you.SG' with third-person verb forms;

- *el señor / la señora* 'the gentleman / the lady' or similar terms of impersonal address with third-person verb forms;
- null subject ("bare") third-person verb forms (*canta* 'he/she.sings > 'you.sing'): ambiguous if not preceded at least once by a specific address term or pronoun;
  - Note that *él/ella*, *el señor/la señora* and use of null subject third-person verbs developed during the 16th century to index mitigated deference/distance to addressee and to avoid the greater self-lowering implied by use of *vuestra merced* (Ly 1981: 47; Hodcroft 1993–1994);
- *vuestra merced* 'your grace > you.SG': early in century, deference/distance to/between court officials, then to/between ever-increasing numbers of persons lower in the social hierarchy;
- *su merced* 'his/her grace' > 'your grace': late 16th century remodeling of *vuestra merced* to avoid relation to increasingly derogatory *vos*.

In and near elite court circles, use of the forms *vuestra señoría* and *vuestra excelencia* for reference to titled nobles also increased greatly during this period, but the form that saw the most dramatic and widespread rise in frequency of use was *vuestra merced*. Although *vuestra merced* and similar forms were introduced in the 15th century, and were quickly adopted in letters circulating in the court (Ly 1981), frequent use of them appears to have remained limited to epistolary texts and to (some) spoken interaction until the 16th century.

We can get a rough sense of the rapid rise and extension in use of *vuestra merced* over the course of the 16th century by comparing the frequency of use in similar genres of texts. For instance, Table 1 shows a selection of form counts from León's (2011) study of address terms in late medieval and early modern literary texts. Unlike the letters circulating in the court, medieval literary texts show little use of *vuestra merced*, a fact which reflects specific genre conventions but also the limitation of use of indirect address to certain interactions among members of the court elite. By around 1600, however, *vuestra merced* had become the most commonly used form in many literary texts, as we can see in Cervantes' usage in *Don Quijote*, and this fact seems to reflect both changes in genre conventions and the rapid rise and extension of use of *vuestra merced* (and its reduced forms) across Spanish-speaking societies. This can also be seen in Hidalgo's (2018) study of the Spanish of New Spain during the mid- and late 16th-century, in which she analyzes a broad range of non-literary text types (e.g., letters, applications to travel, denunciations, testimonies, inquisitorial trials). As seen in Table 2, she finds similar evidence of a drastic increase in use of *vuestra merced*, with overall majority use of *vuestra merced* (or more accurately, its often-abbreviated forms *vm* or *vmd*).[11]

11. These studies do not analyze use of *él/ella*, reduced forms or null subject third-person verbs as forms of address.

Table 1. Explicit use of singular address forms in seven 15th-century literary texts compared with use in *Don Quijote de la Mancha*, Parts I (published in 1605) and II (1615) (based on León 2011: 81–139, 181, 185).[12]

Address form	Form counts in seven 15th-century texts	Form counts in Cervantes' *Don Quijote*
tú	1231 (78%)	260 (17%)
vos (sg.)	330 (21%)	447 (29%)
vuestra merced/vuesa merced	9/0 (0.06%)	605/198 = 803 (52%)
su merced	0	19 (1%)
vuestra señoría	0	18 (1%)
TOTAL	1570	1547

Table 2. Explicit use of singular address forms in New Spain, 1550–1600 (Hidalgo 2018: 246)

Address form	*tú*	*vos*	*vuestra merced*	*su merced*
Form counts	19 (1.95%)	374 (38.43%)	567 (58.27%)	13 (1.33%)

This relatively rapid rise in use of *vuestra merced* occurred not only in writing but in speech, and its increasingly high frequency of use favored its rapid grammaticalization. Indeed, its development from a lexical sequence to a grammatical form is characterized by all the parameters of grammaticalization of personal pronouns (Heine & Song 2011). *Vuestra merced* suffered phonetic erosion (see Figure 1), first to *vuessa merced* (by the 1530s, see below). Subsequently, a great variety of reduced forms developed in two broad lines (with some apparent blending across these): those without /t/, leading ultimately to *voacé* (first attested in 1598) and *ucé*, and those with /t/, leading ultimately to *usted* (first attested in the 17th century, though there is disagreement about exact date of first attestation; cf. De Jonge & Nieuwenhuijsen 2009: 1651; Hammermüller 2010). *Vuestra merced* also showed clear evidence of desemanticization, or semantic bleaching of the possessive and the attribute (i.e., *merced*), and of decategorialization, with loss of syntactic categories of adjective and noun, along with the feminine gender of the noun. Though phonetic reduction begins with the form *vuessa merced* (in which the relation to the possessive *vuestra* is less transparent), there is some evidence of recategorialization already in late 15th-century elite usage. For instance letters sent to Miguel Pérez de Almazán, secretary of the Catholic Monarchs, regularly closed with *Servidor de v.m. que sus manos besa* 'Servant of your grace who kisses your hands', where *manos* is preceded by use of the third-person possessive *sus* rather than second-person *vuestras* (Ly 1981: 52).

12. The 15th-century texts include four poetic texts and three prose texts. These last include *El corbacho* (1438), *Cárcel de amor* (1492), and *Grisel y Mirabella* (before 1495).

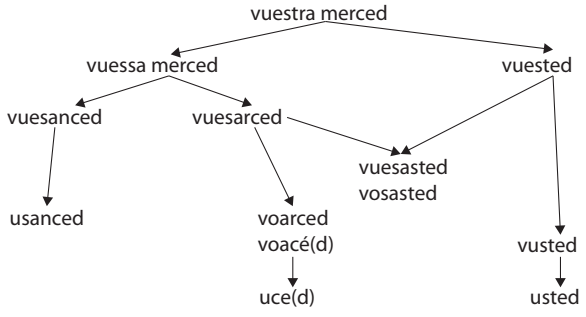


Figure 1. Examples of reduced forms of *vuestra merced* attested in the 16th & 17th centuries (adapted from De Jonge & Nieuwenhuijsen 2009: 1651 and Hammermüller 2010)

Ultimately, the grammaticalization of *vuestra merced* introduced new pronouns of address, leading to further complexification. The numerous reduced forms, each with unique and changing indexical values (see below) would make the use of address forms a challenging concern of everyday social life in the early modern period (only from the 18th century would the reduced form *usted* be selected in the developing standard). Systemic complexification was also increased by the re-analysis of *él/ella* as pronouns of address and the introduction of remodeled *su merced* as yet another possibility, along with the use of null subject third-person verbs. In addition, early modern systems of address included use of such forms as *vuestra señoría* and *vuestra excelencia* for address to members of the titled nobility and grandees (a select group of the highest titled nobility created in 1519). Their use, though socially limited, became commonplace among those interacting with nobles, so much so that they also suffered grammaticalization with reduction to such pronouns as *usía* and *ucencia*.

It is important to note how these developments compare with those in other Romance and European languages in the early modern period. Other Iberian languages (Portuguese and Catalan) show developments similar to those of Spanish (though the reasons for this require further study). Italian develops use of *vostra signoría* 'your lordship' from the 13th century, but rather than grammaticalize this phrase, Italians ultimately select *Ella* or *Lei*, pre-existing 3rd-person feminine pronouns, to pronominalize the feminine noun phrase. German-speakers followed a similar pattern, selecting third-person plural pronoun *Sie* to pronominalize *eure gnaden* 'your.graces=you.SG', which combined pluralization and indirect third-person reference to doubly mark deferential address. However, French showed no change, and even today maintains the medieval Romance system, while English, which had borrowed the indexical values for its singular pronouns *thou/thee* and *ye/you* from Medieval French, saw its address pronoun forms reduced to

just one: *you*. These disparate developments underscore the need for comparative study of the actuation of the changes in each language community, though here I consider only the case of Spanish.

6. Actuation of changes in the address pronoun system(s)

We have seen, then, that increase in frequency of use of *vuestra merced* and its ultimate grammaticalization contributed to complexification of the address pronoun system, but that this happened during a time in which we might rather have expected either maintenance or simplification of that system, as happened in French and early modern English, respectively. These unexpected changes can be accounted for, I believe, by taking into account the interaction of a number of factors:

1. The development of the state and spread of new notions and practices of "negative civility" from the late Middle Ages;
2. Prior 15th-century incorporation of *vuestra merced* into the "court sociolect", which in turn assured its inclusion in 16th-century pre-koine feature pools;
3. Massive demographic and social mobility, with dialect/sociolect mixing, leading to koineizing conditions favoring change in Peninsular cities (especially Seville and Madrid) and (in combination with language contact phenomena) in the Spanish colonies;
4. The maintenance of a medieval social ideology of static hierarchy, promoted and reinforced by the imperial administration and Counter-Reformation Catholic Church;
5. The development of extreme status anxiety in Peninsular and colonial society, along with behaviors that aimed to compensate for this insecurity;
6. The "hyper-salience" of address forms, which favored an increase in their intentional use by speakers in negotiations of social status and relations.

The formulation of this combination of factors – notably the integration of the last three – depends crucially on an understanding of koineization as heuristic rather than as a primarily predictive model. In what follows I present an integrated, cohesive account of the interaction of these diverse factors and the changes in address to which they contributed in early modern Spanish-speaking societies.[13]

---

13. This discussion follows a presentation of similar arguments in Tuten (2008), with some modifications and additions. Here, of course, I am primarily interested in considering the implications of these developments for our understanding of the model of koineization.

6.1 Setting the stage: Change in 15th-century court society and address

Castilian use of *vuestra merced* and similar expressions begins among the 15th-century court elite, but expressions of address which include a possessive and noun attribute have a long history in European discourse. In the late Roman Empire, expressions such as *maiestas tua/vestra* 'your majesty' arose for reference to the Emperor (Svennung 1958: 72–85). In the Middle Ages, these and other formulae were revived or continued by the chanceries of the Holy Roman Empire and Papacy (Niculescu 1974: 13), as revealed in the use of Italian forms *vostra ma(g)està* (for the Emperor) and *vostra santità* 'your holiness' (for the Pope). In German, *euer Majestät* had become normal in address to the Emperor from the 14th century, and *eure Gnaden* 'your graces' (already with singular meaning, as revealed by its Latin equivalent *gratia vestra*) had become current for address to lesser rulers by the 15th century (Svennung 1958: 104–106). Italy appears to have led (at least in Romance-speaking zones) in the diffusion of the use of these abstract formulae as indirect address terms, with the first manifestations of *vostra signoria* 'your lordship' in the 14th century (Coffen 2002: 66). Niculescu (1974: 90) also suggests that the poetic tradition of courtly love favored the development of indirect address, since in this tradition the abstract qualities of the beloved were praised, and we know that expressions such as *la vuestra merced* 'your grace/favor' y *la vuestra cortesía* 'your courtesy', with reference to the indicated attribute, are found in Castilian documents of the 13th and 14th centuries (Líbano Zumalacárregui 1991; Eberenz 2000: 106). These phrases would eventually come to be used for deferential, indirect address. Ly (1981: 50) cites a 15th-century letter to king John II that includes both the traditional use and the new use as a form of address:

- *veso vuestras manos e me encomiendo a vuestra merced* 'I kiss your hands and commend myself to your favor/grace';
- *la que vuestra merced me escribe* 'the one that your.grace=you.SG writes me'.

Indeed, although Koch (2008: 68–69) identifies some 14th-century examples, use of indirect formulae appear to rise dramatically during the reign of John II (1406–1454). In letters to this monarch, a range of formulae are employed to address him, including *vuestra alteza, vuestra señoría*, and *vuestra merced*, although all appeared alongside the traditionally deferential *vos* (Ly 1981: 51, 62, 73). These mixed forms were quickly extended for address to other members of the court. Already in letters dating from the late 1420s, Fernán Gómez de Cibdarreal, physician to Juan II, addresses the king as *vuestra señoría* (also occasionally as *vuestra merced*), archbishops and titled nobles as *vuestra merced*, and esteemed (but non-noble) poet Juan de Mena as *vos* and *vuestra merced* (Pons 2016). By the end of the 15th century,

however, there had occurred a ranking and ordering of (at least some) of these forms. This is suggested in the letters that military leader Gonzalo de Ayora wrote to the Catholic Monarchs and their secretary Miguel Pérez de Almazán (Ochoa 1924: 61–74), in which Isabella and Ferdinand receive *vuestra alteza*, titled nobles and archbishops receive *señoría*, and court officials, represented by the royal secretary himself, receive *vuestra merced*.

The adoption of such indirect forms of address was widespread among western European court elites. It appears to be a linguistic manifestation of the developing Renaissance concept of civility (Elias 1994; Thomas 2018: 14, 84). In this "civilizing process", the late medieval and early modern development of states required ever growing control of the conduct of individuals, beginning in expanding royal and noble courts. This control was manifested in fundamentally hierarchical terms: the principal idea was to please – or at least avoid offending – one's social superiors, beginning with the monarch, with only a secondary concern for one's social equals and almost none for social inferiors (Burke 2000: 40 refers to this as "negative civility"). A change in the address terms directed to superiors can easily be related to the developing norms of civility. For instance, in early stages of this process, a superior could remain nude, defecate, or urinate before his inferiors, but his/her inferiors were increasingly obliged to avoid such behavior in the presence of superiors (Elias 1994: 117). Such control of bodily conduct has a direct analogue in the initial imposition of use of expressions of indirect address on social inferiors (note that within Castile the monarch could continue to address all inferiors with *tú* or *vos*).

Although this cultural change likely favored the ever more widespread use of deferential, indirect address in the 15th, 16th and 17th centuries, the development of the state and of civility was a pan-European phenomenon and therefore cannot explain the timing and rapidity of the adoption of indirect address in the Castilian court (nor its widespread adoption in the 16th century). It would seem that rapid adoption of such forms in the Castilian court was favored principally by the political insecurity of the Trastamaran kings and the extreme social mobility and insecurity within the 15th century court (as well as increasing awareness of Italian customs via contact with the Aragonese-Catalonian court). The Trastamaran dynasty was descended from Henry II (1369–79), who was a bastard and regicide, and his heirs John II and Henry IV (1454–1475) were generally regarded as weak kings (even though state institutions continued to develop under them; Valdeón Baruque 2001: 228–231). Their need to secure political support led them to concede some 97 titles of nobility (of viscount, count, marquis, and duke) in only six decades (Boase 1978: 159). Thus, within a very short period, they created a hereditary titled nobility in Castile, where almost none had existed before.

At the same time, the Trastamaras employed and relied on a growing number of *letrados*, university-educated men, as royal functionaries in charge of the administration of the court and state (Boase 1978: 79, Valdeón Buarque 2001: 22, Ruiz 2017: 86). The simultaneous creation of a new class of titled nobles and another of court officials had two important effects. First, it reinforced the idea of a patron-client relationship centered on the king, as well as the idea that overt acts of loyalty and deference should be rewarded with overt *mercedes* 'favors'; both of these attitudes were evident in the exhibitions of extreme deference made before John II. Second, the creation of these two newly constituted groups – one of new titled nobles, the other of a new lower, bureaucratic nobility – led to a considerable level of social mobility within and beyond the court, and with it, increasing social insecurity and competition for status (Casey 1999: 154).

The new nobility lacked legitimacy in the same way that the kings lacked political legitimacy, since they had acquired a new status whose justification depended, paradoxically, upon a conservative medieval ideology of social hierarchy and stasis, which was often articulated in terms of the idealized "three estates": ranked, God-given categories that supposedly determined and fixed the place of each person in society (Maravall 1972: 12–17). In this uncomfortable position, the new titled nobles were obliged to compensate for their evident lack of inherited status by means of overt, compensatory performances of noble status. They could, for instance, participate in a revived tradition of courtly love, writing poems and wearing the color black to show their "noble sensitivity" (Boase 1978; Ruiz 2017: 80). Of course, performances of nobility were not solo acts: they required the cooperation and participation of others, and these "co-performances" would have involved the obligatory use of indirect address terms by inferiors to their (newly) noble superiors. The status anxiety of the new titled nobles was only increased by the need to distinguish themselves from the growing class of court officials, particularly since a number of them had their origins in that group.

Court officials faced similar challenges as they sought to secure and defend their own new social positions. Their concerns about social status were based primarily on the fact that some of them entered as commoners of "middling" social rank while others were already members of the low-level (untitled) nobility. These included *hidalgos*, a status based on (claimed) lineage and "purity of blood", and *caballeros* 'knights', whose status was acquired by means of a ceremony of investiture. Ultimately, the commoners among the court bureaucracy also became *hidalgos* or *caballeros*, with an enhanced need to perform their new status and have it recognized and performed by others. Their anxieties (shared by some new titled nobles) were exaggerated by the fact that many of them were *conversos*, persons of Jewish descent whose status was especially resented and suspect within Castile. As this group gradually negotiated its status with those above and below them, a de facto

codification of address appears to have taken place, with, as we have seen, *alteza* reserved for the monarch, *señoría* for titled nobles, and *merced* for court officials (which, appropriately enough, emphasized what these individuals could do or offer over who they were). The fact that the new bureaucratic nobility overlapped with the older low-level nobility of *hidalgos* and *caballeros* was significant for later developments, since within Castile members of these ranks made up almost 10% of the total population – a far higher percentage than in other European countries (Ruiz 2017: 89). By the time of the great population movements of the 16th century, awareness of an increasing need to use *vuestra merced* (at the very least in letters written to court officials; Ly 1981: 52) appears to have spread beyond the court to elites across Castile.

The adoption of indirect address forms in and beyond the court, particularly *vuestra merced*, was therefore the result of a unique combination of tensions in Castilian court society: a discrepancy between a conservative social theory of static hierarchy and a fluid social reality. These tensions led to a struggle to secure and defend newly gained status by both of the newly constituted elites, as well as to severe status anxiety accompanied by overcompensating discursive and other behavior. As we will see, this fundamental dynamic is repeated on a much larger scale in early modern Castilian and colonial society.

6.2 Early modern developments

From the 1530s there is increasing evidence pointing to widespread, explicit awareness of the social value of using *vuestra merced* in writing and speech in order to advance socially. Juan de Valdés, in the *Diálogo de la lengua* (1535), confirms that current court speech had already come to prefer the slightly reduced form *vuessa merced* (Pla Cárceles 1923: 246). In 1539, Antonio de Guevara, a writer and ecclesiastic closely connected to the court of Charles V, recommended greater use of *vuestra merced* by those who desired to achieve social success:

> [there was] a valiant and generous *caballero*, albeit not well brought up, that I always heard saying to each one *vos*, *vos*, *él*, *él* and who never said *merced*. I said to him: "By my life, sir, I often think to myself that because of this neither God nor the King ever grant you a *merced*, because you (*vos*) never call anyone *merced*.
> (quoted in Ly 1981: 84; my translation)

The Inca Garcilaso de la Vega also reported that in the 1530s and 1540s Gonzalo Pizarro (brother to Francisco) became known for his generous use of *vuestra merced* to his Spanish companions in Peru, in a evident attempt to secure their loyalty (del Castillo Mathieu 1982: 615). From the middle of the century, Italians observed the high frequency of use of *vuestra merced* – roughly equivalent to their *vostra*

signoria – among the stereotypically proud and honor-obsessed Spaniards, and even blamed the Spaniards for their own increasing use of *vostra signoria* (Niculescu 1974: 108).

By century's end, there is also evidence that the different forms of address were acquiring a diverse range of indexical values (beyond those indicated above), of which users were very aware. *Vos*, of course, had become unacceptable for use to superiors and even among equals of upper and middling social ranks, and its use was increasingly interpreted as an affront to the addressee's honor (Ly 1981: 42). *Vuestra merced* and its reduced forms had become associated with (and apparently initially limited to) urban usage, while avoidance of them through use of *él/ella* had become restricted to address to/among nuns and monks (at least in letters) and otherwise associated with rural usage (Lapesa 1970: 152, 158–160). Indeed, use of *él/ella* had become an urban stereotype of rural usage, as revealed by its comic use in the plays of Lope de Vega (Ly 1981: 50). Especially in cities such as Seville and Madrid, forms of *vuestra merced* had become the norm for symmetrical address among members of the lower-level elite and middling ranks of society: *caballeros* and *hidalgos*, but also merchants (many of whom aimed to become or became *caballeros*), notaries, lawyers, scribes, doctors and even artisans (Ly 1981: 96). In general, reduced forms of *vuestra merced* were associated with more informal, colloquial usage, but beyond this there was much variation in their indexical values. For instance, Cervantes' use of the much-reduced form *voacé* in a poem ("Al túmulo de Felipe II en Sevilla", 1598) indexes lower-class Seville speech and simultaneously evokes the decline of Seville and of all Spain. In 1597, César Oudin writes in his *Grammaire espagnole* that *vosasted* is the reduced form most commonly used in speech, but this observation is almost surely based on his prior experience with local usage in the Madrid court (Pla Cárceles 1923: 254). Eventually, the prevalence of reduced forms made spoken use of the full form *vuestra merced* into another stereotype of a lack of urbanity.

This rapid diffusion of *vuestra merced* through urban society, to increasingly lower ranks, can largely be accounted for if we again consider the tension between the dominant social ideologies of the period and the koineizing contexts in which urban speakers were interacting, particularly in Seville and Madrid. The conservative social ideals of the 15th century were largely retained in the 16th century. If anything, there was a strengthening of the idea that a fixed, ancestral social hierarchy was necessary for good government. Following the defeat of the *comuneros* in 1521, the growing imperial state implemented ever more policies to (attempt to) restrict social mobility, including statutes of *limpieza de sangre*, or blood purity, and sumptuary laws (Ruiz 2017: 78, 239). It did so in alliance with the Counter-Reformation Catholic Church, which taught belief in the fundamentally static and hierarchical nature of both the spiritual and temporal worlds (Casey 1999: 138–164).

Ironically, 16th-century Castile also witnessed an unprecedented moment of social mobility. This mobility is usually attributed to the affluence of wealth from the Indies, but the great demographic movements to and through cities in Castile and the New World were equally important. Demographic mixing, particularly in Seville, Madrid, and the colonies, offered increased social anonymity, led to generalization of looseknit social networks, and created a situation in which social positions could and did change incessantly, with constant claims being made for higher social status (e.g., Crawford 2014 studies numerous attempts to claim *hidalgo* status in Seville). Importantly, social mobility was generally perceived as threatening to society, in accord with the dominant social ideologies of the time (and as revealed in picaresque novels). This reality occasioned the intensification and extension to broader society of the status anxiety – and stereotypical Spanish obsession with "honor" or face – that had characterized the 15th-century Castilian nobility.

Demographic mixing also led to mixing of diverse dialects and sociolects (including that associated with the court). In these koineizing contexts, speakers from different locales, even if employing similar forms, almost certainly had different ideas about norms for selection and use of address forms, as is suggested by the many attested cases of conflict regarding appropriate address in real life and in drama (e.g., Bentivoglio 2002; Hodcroft 1993–1994; Ly 1981). Still, the use of salient address terms could be "noticed" and talked about, and new uses easily learned or invented. The lack of community norms and norm enforcement, and the ease of intentional, goal-oriented use of address forms, would allow rapid change to occur, particularly among those moving to the cities.

Many of those who moved to the cities were members of the middling ranks of society. In the fluid social contexts of the cities, recent arrivals (and even established residents) who had the necessary economic means could begin a new life with a new social status (Ruiz 2017: 85–89). A merchant could become a *caballero* or a *converso* commoner might obtain *hidalgo* status. Such a change required, among other things, a performance of the new status through choice of clothing, food, housing, activities, social relations, and features of language. With regard to address, social climbers (below the titled nobility) would have needed to insist that others serve as co-participants in their performance, requiring that both inferiors and equals address them with (forms of) *vuestra merced*. Such demands, of course, would have been difficult or impossible to satisfy if the social climber had remained in a (generally smaller) community of origin, where his or her social origins were known and more stable, closeknit social networks served to preserve more conservative norms and resist inflation in use of *vuestra merced* (as seen in the rural resistance to explicit use of *vuestra merced*).

Another factor favoring the increase in intentional use of *vuestra merced* in these urban contexts was the fact that a majority of new arrivals did not possess great

economic and social resources. To satisfy their needs and achieve desired goals in a society and culture dominated by relations of patronage, in what was conceived of as a "pyramid of deference and obligation" (Casey 1999: 148–165), new arrivals were obliged to seek *mercedes* or favors from officials and established residents, often by means of intermediaries. They would have found it advantageous to exaggerate their explicit use of deferential forms, particularly *vuestra merced*, offering it as a form of symbolic capital in exchange for a desired favor (cf. the above quote from Sevillian Antonio de Guevara). And of course, in a society dominated by extreme status anxiety and concern for possible loss of honor or face (Casey 1999: 170), failure to use an adequately deferential form could have truly detrimental consequences. Such pressures made use of (forms of) *vuestra merced* not only increasingly acceptable but necessary in contexts where previously *vos* or *él/ella* would have sufficed. Of course, those who succeeded in rising in the social hierarchy would also expect to receive *vuestra merced* from those equal or inferior to them.

While a generalized and extreme status anxiety favored ever-increasing use of *vuestra merced*, ongoing koineization (with looseknit social networks in heterogeneous urban contexts characterized by dialect mixing and weakened norm enforcement) allowed the increase to take place without significant opposition. Importantly, explicit use of *vuestra merced*, as opposed to pronominalization with other existing pronouns (as in Italian or German), appears to have been favored by this lack of conservative resistance and the increasing social danger of using options that were coming to be seen as insufficiently urban and deferential (e.g., *él/ella* or null subject verbs). Koineizing contexts favored not only increases in frequency but also the grammaticalization of not one but numerous reduced forms of *vuestra merced*, which also appear to have acquired conventional indexical values of place (e.g., Seville vs. Madrid), of social rank, and of style (aside from endless contextual tropes upon such conventional values). Ultimately, the radical increase in use of such grammaticalized forms led to even greater complexification of the overall system of address pronouns, although over time certain forms diffused more or less widely than others, and consequently there developed a diversity of complex address systems.[14]

7. Conclusion: A role for koineization?

In the preceding account I have employed the model of koineization as a heuristic. This has involved integrating and foregrounding the impact of sociocultural factors in an explanatory account of a complexifying change not normally relatable to koineization, and to extension of the model of koineization to

14. At this point, a separate set of questions arises: when, where, how and why does complexity in address begin to decline? These are questions that I will explore in future work.

"less-than-prototypical" contexts. In reviewing the arguments presented, it will be helpful to relate them to the question that constitutes the second part of my title: what exact role(s) did koineization play in the complexification of the early modern Spanish system(s) of address?

Evidently, one cannot claim that a prototypical process of koineization "led to" or "caused" this complexification in a predictable way, that is, by attributing its occurrence solely to the constrained set of sociodemographic and psychosocial factors that are normally understood to occasion the outcomes of mixing, selection of majority variants, reallocation (though see below), and simplification. First (and this is particularly so if we adopt Trudgill's deterministic model for *tabula rasa* communities as our prototype), it is quite evident that early modern Seville, Madrid and New World colonial communities do not match prototypical definitions of a koineizing community. Nevertheless, they were characterized by key features of such communities: demographic movement, mixing and integration of speakers of related dialects/varieties, an initially "chaotic" feature pool and weak norm enforcement, and smooth transmission of the great majority of language features from older to younger generations.

Nor can koineization be said to have been the primary factor motivating the adoption of new address forms: this is clearly attributable to the driving influence of contemporary ideologies of social stasis and hierarchy in a context of great geographic and social mobility. The tension between ideals and reality led to a steep and widespread rise in status anxiety. Speakers, particularly status-conscious adults, sought to compensate for their social insecurity through compensatory, intentional performances of status, including increasing use of inflated forms of address by self and expectations of their use by others.

However, koineization does seem to have played a key role in the timing, rapidity, location, and specific outcomes of this process of change. Koineization in early modern Castile, particularly the cities of Seville and Madrid, as well as the New World colonies, favored typical outcomes (simplification in sibilant phoneme inventory) and atypical outcomes (complexification in address), both of which occurred over the course of just two or three generations (mid-16th to early 17th centuries) in the relevant communities.[15]

---

15. More comparative research is required with regard to the possible impact of early modern koineization on change in other European languages. Portuguese developed a complex address system like that of Spanish, and this may have been influenced by a similar conjunction of sociodemographic and sociocultural factors, but may also be attributed in part to the heavy cultural influence of Castile on Portugal. In early modern England, London saw large demographic inflows (swamping) and dialect mixing, but this occurred in a cultural context quite different from that of Castile.

The specific outcomes of changes in address include rapid grammaticalization of indirect address forms such as *vuestra merced*, with the creation of numerous new pronouns over the course of just two or three generations. Although sociocultural factors were key to the initial adoption and extension of these forms, koineizing conditions – notably the lack of stable sociolinguistic norms and norm enforcement mechanisms – allowed these inflationary tendencies in adult behavior to progress unchecked, with subsequent learning and grammaticalization of the most frequent forms by subsequent cohorts of children and adults.

The role of koineization is particularly evident if we compare the lack of constraints placed on use of explicit forms of *vuestra merced* with resistance to use of full indirect forms in other societies and languages. In comparatively more stable Italian and German communities, similar sociocultural factors also favored inflation and complexification in address, however, in those languages pronominalization was generally limited to adaptation of existing third-person pronouns to second-person use (Italian *vostra signoria* > *Ella/Lei*, German *eure Gnaden* > *Sie*). This can be attributed to the use of the pronoun as a mitigating device that avoided the extreme self-lowering implied by use of the full indirect forms. Similar resistance to use of the full indirect forms was also found (for a time) in the more stable and conservative rural communities of Castile. It was only in urban Castilian (Seville, Madrid) and colonial communities that koineizing conditions allowed entirely new forms of address to be adopted, extended, and rapidly grammaticalized without significant resistance.

It is grammaticalization of new forms that led to the complexification of the system of address. Although complexification is rarely seen as a possible outcome of koineization, the general recognition of reallocation as a possible outcome suggests that complexification is at least a possibility even within prototypical models of koineization. More to the point, socio-stylistic reallocation leads to the selection of two or more pre-existing variants from different varieties (or newly generated variants), with each assigned new and different socio-stylistic functions. This certainly seems to describe the complexification of the early modern Spanish address system: selection of both pre-existing and newly adapted address forms, each with different indexical values. Nevertheless, socio-stylistic reallocation appears to be dependent on the influence of sociocultural factors, and is therefore not predictable. Moreover, complexification in koineization would appear to be limited to increases in bit complexity: I see little possibility for opacity-producing increases in algorithmic complexity.

Ultimately, this study has had two interrelated aims. One goal has been to demonstrate how use of the model of koineization can aid in developing a cogent and plausible explanatory account for the rapid complexification of the early modern Spanish address system. To do this, however, it has also been necessary to

argue for the advantages of using koineization as a heuristic device. For research on language change in new communities that have undergone demographic movement and dialect mixing, the advantages include the potential application of the model to "less-than-prototypical" communities, integration of sociodemographic with sociocultural factors in accounts of outcomes considered atypical of koineization, as well as consideration of the potential impact of adult beliefs and behavior, along with "noticing" and intentional use of salient features. Rather than reducing possibilities for understanding how and why changes occur when and where they do, a movement away from strong expectations of predictability may actually enhance the explanatory potential of the model.

References

Agha, Asif. 2007. *Language and social relations*. Cambridge & New York: Cambridge University Press.

Anipa, Kormi. 2001. *A critical examination of linguistic variation in Golden Age Spanish*. New York: Peter Lang.

Auer, Peter. 2014. Anmerkungen zum Salienzbegriff in der Soziolinguistik. *Linguistik Online* 66(4), https://bop.unibe.ch/linguistik-online.

Bell, Allen. 2016. Succeeding waves: Seeking sociolinguistic theory for the twenty-first century. In N. Coupland (ed.), *Sociolinguistics: Theoretical debates*, 391–416. Cambridge/New York: Cambridge University Press. https://doi.org/10.1017/CBO9781107449787.019

Bentivoglio, Paola. 2002. Spanish forms of address in the sixteenth century. In Irma Taavitsainen & Andreas Jucker (eds.), *Diachronic perspectives on address systems*, 177–191. Amsterdam/Philadelphia: John Benjamins. https://doi.org/10.1075/pbns.107.09ben

Boase, Roger. 1978. *The troubadour revival: A study of social change and traditionalism in late medieval Spain*. London: Routledge & Kegan Paul.

Boyd-Bowman, Peter. 1976. Patterns of Spanish emigration to the Indies until 1600. *The Hispanic American Historical Review* 56. 580–604.

Britain, David. 2012. Koineization and cake-baking: Reflections on methods in dialect contact research. In Andrea Ender, Adrian Leemann & Bernhard Wälchli (eds.), *Methods in contemporary linguistics*, 219–238. Berlin/Boston: De Gruyter Mouton.

Brown, Roger, and Albert Gilman. 1960. The pronouns of power and solidarity. In Thomas Sebeok (ed.), *Style in language*, 253–276. Cambridge: MIT Press.

Brown, Penelope, and Stephen C. Levinson. 1987. *Politeness: Some universals in language use*. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511813085

Burke, Peter. 2000. A civil tongue: Language and politeness in early modern Europe. In P. Burke, B. Harrison & P. Slack (eds.), *Civil histories: Essays presented to Sir Keith Thomas*, 31–48. Oxford and New York: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780198207108.003.0002

Cappella, Joseph. 1997. The development of theory about automated patterns of face-to-face human interaction. In G. Philipsen & T. Albrecht (eds.), *Developing communication theories, SUNY series in human communication processes*, 57–83. Albany: State University of New York Press.

Casey, James. 1999. *Early modern Spain: A social history*. London and New York: Routledge. https://doi.org/10.4324/9780203255148

Castillo Mathieu, Nicolás del. 1982. Testimonios del uso de vuestra merced, vos y tú en América (1500–1650). *Thesaurus* 37. 602–644.

Coffen, Béatrice. 2002. *Histoire culturelle des pronoms d'adresse: vers une typologie des systémes allocutoires dans les langues romanes*. París: Honoré Champion.

Crawford, Michael J. 2014. *The fight for status and privilege in late medieval and early modern Castile, 1465–1598*. University Park, PA: Pennsylvania State University Press.

De Jonge, Bob. 2005. El desarrollo de las variantes de vuestra merced a usted. In Bob de Jonge (ed.), *Actas del II Congreso de la Región Noroeste de Europa de la Asociación de Lingüística y Filología de América Latina* (ALFAL). Estudios de Lingüística Española 22.

De Jonge, Bob and Dorien Nieuwenhuijsen. 2009. Formación del paradigma pronominal de las formas de tratamiento. In Concepción Company Company (ed.), *Sintaxis histórica de la lengua española. Segunda parte: La frase nominal*. Vol. 2, 1593–1671. Mexico: Universidad Nacional Autónoma de México and Fondo de Cultura Económica.

Eberenz, Rolf. 2000. *El español en el otoño de la edad media: sobre el artículo y los pronombres*. Madrid: Gredos.

Eckert, Penelope. 2008. Variation and the indexical field. *Journal of Sociolinguistics* 12(4). 453–476. https://doi.org/10.1111/j.1467-9841.2008.00374.x

Elias, Norbert. 1994. *The civilizing process*. Oxford: Blackwell.

Gallois, Cindy; Ogay, Tania; & Giles, Howard. 2005. Communication Accommodation Theory. In William B. Gudykunst (ed.), *Theorizing about intercultural communication*, 121–48. Los Angeles: Sage.

García-Godoy, María Teresa. 2019. Una nueva hipótesis diacrónica sobre su merced como tratamiento informal. *Bulletin of Hispanic Studies* 96(2). 107–205. https://doi.org/10.3828/bhs.2019.7

Garrod, Simon, & Clark, Aileen. 1993. The development of dialogue co-ordination skills in schoolchildren. *Language and Cognitive Processes* 8. 101–26.

Giles, Howard. 1973. Accent mobility: A model and some data. *Anthropological Linguistics* 15. 87–105.

González Martínez, Rosa María. 2002. *La población española: Siglos XVI, XVII, y XVIII*. Madrid: Actas.

Gordon, Elizabeth, Lyle Campbell, Jennifer Hay, Margaret Maclagan, Andrea Sudbury, Peter Trudgill. 2004. *New Zealand English: Its origins and evolution*. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511486678

Hammermüller, Gunther. 2010. Evolución de las formas de tratamiento del español medieval hasta el siglo XVI. In M. Hummel, B. Kluge & M. E. Vázquez Laslop (eds.), *Formas y fórmulas de tratamiento en el mundo hispánico*, 507–529. Mexico: El Colegio de México.

Heine, Bernd, and Kyung-An Song. 2011. On the grammaticalization of personal pronouns. *Journal of Linguistics* 47. 587–630.

Hidalgo, Margarita. 2018. *Diversification of Mexican Spanish: A tridimensional study in new world sociolinguistics*. Boston and Berlin: Mouton de Gruyter.

Hodcroft, F. W. 1993–1994. ¿A mí un él? Observations on vos and él/ella as forms of address in Peninsular Spanish. *Journal of Hispanic Research* 2. 1–16.

Johnstone, Barbara. 2013. *Speaking Pittsburghese: The story of a dialect*. Oxford and New York: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199945689.001.0001

Johnstone, Barbara. 2018. *Discourse analysis*. 3rd ed. Oxford: Wiley-Blackwell.

Kerswill, Paul. 1996. Children, adolescents, and language change. *Language Variation and Change* 8. 177–202.
Kerswill, Paul. 2010. Contact and new varieties. In Raymond Hickey (ed.), *The handbook of language contact*, 230–251. Oxford: Blackwell. https://doi.org/10.1002/9781444318159.ch11
Kerswill, Paul. 2013. Koineization. In J. K. Chambers & Natalie Schilling (eds.), *The handbook of language variation and change*, 519–536. Oxford: Wiley Blackwell. https://doi.org/10.1002/9781118335598.ch24
Kerswill, Paul, and Ann Williams. 2000. Creating a new town koine: Children and language change in Milton Keynes. *Language in Society* 29. 65–115.
Kerswill, Paul, and Ann Williams. 2002. Salience as a factor in language change: Evidence from dialect leveling in urban England. In M. C. Jones & E. Esch (eds.), *Language change: The interplay of internal, external and extra-linguistic factors*, 81–110. Berlin/New York: Mouton de Gruyter. https://doi.org/10.1515/9783110892598.81
Koch, Peter. 2008. Tradiciones discursivas y cambio lingüístico: el ejemplo del tratamiento vuestra merced en español. In J. Kabatek (ed.), *Sintaxis histórica del español y cambio lingüístico*, 53–87. Frankfurt and Madrid: Vervuert/Iberoamericana.
Lapesa, Rafael. 1970. Personas gramaticales y tratamientos en español. *Revista de la Universidad de Madrid* 19. 141–167.
Lass, Roger. 1990. Where do extraterritorial Englishes come from? Dialect input and recodification in transported Englishes. In Sylvia Adamson, Viven Law, Nigel Vincent & Susan Wright (eds.), *Papers from the 5th International Conference on English Historical Linguistics. Cambridge, 6–9 April 1987*, 245–280. Amsterdam/Philadelphia: John Benjamins. https://doi.org/10.1075/cilt.65.16las
León, Ana Emilia. 2011. *Extinción de vos en el español peninsular*. New York: Peter Lang. https://doi.org/10.3726/978-1-4539-0786-3
Líbano Zumalacárregui, Ángeles. 1991. Morfología diacrónica del español: las fórmulas de tratamiento. *Revista de Filología Española* 71. 107–121.
Ly, Nadine. 1981. *La poétique de l'interlocution dans le théatre de Lope de Vega*. Bordeaux: Université de Bordeaux.
Maravall, José Antonio. 1972. *Estado moderno y mentalidad social*. Volume II. Madrid: Revista de Occidente.
Mesthrie, Rajend. 1993. Koineization in the Bhojpuri–Hindi diaspora – with special reference to South Africa. *International Journal of the Sociology of Language* 99. 25–44.
Mesthrie, Rajend. 1994. Koinés. In R. E. Asher (ed.), *The encyclopedia of language and linguistics*, 1864–1867. Oxford: Pergamon Press.
Milroy, James and Lesley Milroy. 1985. Linguistic change, social network, and speaker innovation. *Journal of Sociolinguistics* 21(2). 339–384.
Mufwene, Salikoko. 1996. The founder principle in creole genesis. *Diachronica* 13. 83–134.
Mufwene, Salikoko, Christophe Coupé, and François Pellegrino. 2017. Complexity in language: A multifaceted phenomenon. In *Complexity in language: Developmental and evolutionary perspectives*, Salikoko Mufwene, Christophe Coupé & François Pellegrino (eds.), 1–29. Cambridge and New York: Cambridge University Press. https://doi.org/10.1017/9781107294264.001
Newall, Gregory. 2016. Second person singular forms in Cali Colombian Spanish: Enhancing the envelope of variation. In María Irene Moyna & Susana Rivera-Mills (eds.), *Forms ofaAddress in the Spanish of the Americas*, 149–169. Amsterdam/Philadephia: John Benjamins. https://doi.org/10.1075/ihll.10.08new

Niculescu, Alessandro. 1974. *Strutture allocutive pronominali reverenziali in italiano*. Florence: Olschki.

Ochoa, Eugenio de (ed.). 1924. *Epistolario español*. Biblioteca de Autores Españoles 13. Madrid: Sucesores de Hernando.

Penny, Ralph. 1987. *Patterns of language-change in Spain*. London: University of London, Westfield College.

Penny, Ralph. 2000. *Variation and Change in Spanish*. Cambridge: Cambridge University Press.

Penny, Ralph. 2002a. Contacto de variedades y resolución de la variación: aspiración y pérdida de /h/ en el Madrid del s. XVI. In Juan Sánchez & María Teresa Echenique Elizondo (eds.), *Actas del V Congreso Internacional de Historia de la Lengua Española (Valencia, 31 de enero–4 de febrero de 2000)*, Volume 1, 397–406. Madrid: Gredos.

Penny, Ralph. 2002b. *A history of the Spanish language*. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511992827

Pickering, Martin, & Garrod, Simon. 2004. Toward a mechanistic psychology of dialogue. *Behavioral and Brain Sciences* 27. 169–226

Pickering, Martin, and Simon Garrod. 2017. Priming and language change. In Marianne Hundt, Sandra Mollin & Simone E. Pfenninger (eds.), *The changing English language: Psycholinguistic perspectives*, 173–190. Cambridge and New York: Cambridge University Press. https://doi.org/10.1017/9781316091746.008

Pla Cárceles, José. 1923. La evolución de tratamiento vuestra merced. *Revista de Filología Española* 10. 245–280.

Placencia, María Elena. 2010. El estudio de formas de tratamiento en Colombia y Ecuador. In M. Hummel, B. Kluge & M. E. Vázquez Laslop (eds.), *Formas y fórmulas de tratamiento en el mundo hispánico*, 341–373. Mexico: El Colegio de México.

Pons, Lola. 2016. Centón epistolario de Fernán Gómez de Cibdarreal. *Lemir: Revista de Literatura Española Medieval y del Renacimiento* 20. 345–428.

Poussou, Jean-Pierre. 1997. Migrations et mobilité de la population en Europe a l'époque moderne. In Jean-Pierre Bardet & Jacques Dupaquier (eds.). *Histoire des populations de l'Europe*, 262–286. Paris: Fayard.

Ruiz, Teófilo. 2017. *Spanish society, 1348–1700*. Abingdon, Oxon/New York: Routledge. https://doi.org/10.4324/9781315180960

Sanz-Sánchez. Israel. 2016. A diachronic perspective on Latin American Spanish Verbal Morphology: Reassessing the role of koineization. In Eva Núñez-Sánchez (ed.), *Diachronic applications in Hispanic linguistics*, 239–281. Newcastle-upon-Tyne: Cambridge Scholars.

Sanz-Sánchez, Israel. 2019. Documenting feature pools in language expansion situations: sibilants in early colonial American Spanish. *Transactions of the Philological Society* 117(2). 199–233. https://doi.org/10.1111/1467-968X.12154

Schreier, Daniel. 2008. *St. Helenian English: Origins, evolution and variation*. Amsterdam and Philadelphia: John Benjamins. https://doi.org/10.1075/veaw.g37

Schreier, Daniel. 2012. The impact of migratory movements on linguistic systems: Transplanted speech communities and varieties from a historical sociolinguistic perspective. In J. M. Hernández-Campoy & J. C. Conde-Silvestre (eds.), *The handbook of historical sociolinguistics*, 534–551. Oxford: Wiley-Blackwell. https://doi.org/10.1002/9781118257227.ch29

Siegel, Jeff. 1985. Koines and koineization. *Language in Society* 14. 357–378.

Siegel, Jeff. 1997. Mixing, leveling, and pidgin/creole development. In A. K. Spears & D. Winford (eds.), *The structure and status of pidgins and creoles*, 111–149. Amsterdam: John Benjamins. https://doi.org/10.1075/cll.19.09sie

Silverstein, Michael. 2003. Indexical order and the dialectics of sociolinguistic life. *Language and Communication* 23. 193–229.

Sinnott, Sarah. 2013. Generalized conversational implicatures and indexical fields: The case of address forms. In Ana M. Carvalho & Sara Beaudrie, *Selected Proceedings of the 6th Workshop on Spanish Sociolinguistics*, 53–62. Somerville, MA: Cascadilla Proceedings Project.

Svennung, Josef. 1958. *Anredeformen. Vergleichende Forschungen zur indirekten Anrede in der dritten Person und zum Nominativ für den Vokativ*. Uppsala: Lund.

Thomas, Keith. 2018. *In pursuit of civility: Manners and civilization in early modern England*. Waltham, MA: Brandeis University Press. https://doi.org/10.2307/j.ctvzpv6kq

Trudgill, Peter. 1986. *Dialects in contact*. Oxford: Blackwell.

Trudgill, Peter, Elizabeth Gordon, Gillian Lewis, and Margaret Maclagan. 2000. Determinism in new-dialect formation and the genesis of New Zealand English. *Journal of Linguistics* 36. 299–318.

Trudgill, Peter. 2004. *New dialect formation: The inevitability of Colonial Englishes*. Oxford and New York: Oxford University Press.

Trudgill, Peter. 2008. Colonial dialect contact in the history of European languages: On the irrelevance of identity to new-dialect formation. *Language in Society* 37(2). 241–254. https://doi.org/10.1017/S0047404508080287

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

Trudgill, Peter. 2018. Tabula rasa new-dialect formation: On the occasional irrelevance of language regard. In B. Evans, E. Benson & J. Stanford (eds.), *Language regard: Methods, variation and change*, 266–281. Cambridge: Cambridge University Press. https://doi.org/10.1017/9781316678381.016

Trudgill, Peter. Forthcoming. Dialect convergence and the formation of new dialects. In B. Joseph, R. Janda & B. Vance (eds.), *The handbook of historical linguistics, volume II*. Oxford: Wiley-Blackwell. https://doi.org/10.1002/9781118732168.ch6

Tuten, Donald N. 2001. Modeling koineization. In Laurel J. Brinton (ed.), *Historical linguistics 1999: Selected papers from the 14th International Conference on Historical Linguistics*, 325–336. Amsterdam: John Benjamins. https://doi.org/10.1075/cilt.215.22tut

Tuten, Donald N. 2003. *Koineization in medieval Spanish*. Berlin & New York: Mouton de Gruyter. https://doi.org/10.1515/9783110901269

Tuten, Donald N. 2008. Factores socioculturales en el desarrollo de vuestra merced / usted. In Concepción Company & José G. Moreno (eds.), *Actas del VII Congreso Internacional de la Historia de la Lengua Española*, 2189–2200. Madrid: Arco/Libros.

Uber, Diane. 2011. Forms of address: The effect of the context. In Manuel Díaz-Campos (ed.), *The handbook of Hispanic sociolinguistics*, 244–262. Oxford: Wiley-Blackwell. https://doi.org/10.1002/9781444393446.ch12

Valdeón Baruque, Julio. 2001. *Los Trastámaras: el triunfo de una dinastía bastarda*. Barcelona: Temas de Hoy.

Weinreich, Uriel, William Labov, Marvin Herzog. 1968. Empirical foundations for a theory of language change. In Winfred P. Lehman & Yakov Malkiel (eds.), *Directions for historical linguistics*, 95–189. Austin: University of Texas Press.

Personal vs. personalized infinitives in Ibero-Romance

Historical origins and contact-induced change

Lamar A. Graham
UNC Chapel Hill

The infinitives of several Romance languages can appear with an overt subject. Languages such as Portuguese and Galician feature inflectional morphology on infinitives with overt subjects – the commonly named *personal* infinitives, such as *(nós) dizermos* 'us to speak'. Other languages, such as Castilian[1] and Asturian, feature overt subjects alongside infinitives with no corresponding agreement morphology on the verb – a structure I call the *personalized* infinitive, such as *nosotros decir ~ decir nosotros* 'us to speak'. Though superficially similar in use, personal and personalized infinitives differ among Ibero-Romance languages in their history and their uses in the modern dialects. In this paper I distinguish both structures, illustrating the morphosyntactic differences between the two. I also argue for the influence of koineization and language contact as impetus for the historical development of these forms in the various languages.

Keywords: Spanish/Castilian, Portuguese, Galician, Asturian, inflected infinitives, infinitives with subjects, historical change, contact, koineization

1. Introduction

One of the more striking differences between the verbal paradigms of Castilian and other Romance languages, most notably Portuguese and Galician, is the *personal* or inflected infinitive structure (Sabatini 1984; Parafita Couto 2004; Sitaridou 2009; Fidalgo Enrique 2012; Vázquez Diéguez 2012, *inter alia*), which is present and productive in Galician and Portuguese but is said to have never been attested in Castilian.

1. I primarily use "Castilian" throughout this chapter to refer to the language commonly called "Spanish."

https://doi.org/10.1075/ahs.12.co3gra

(1) *nada para **comeres*** (Pt, Gal, *Cast, *Ast)
nothing for eat.INF.2SG
'nothing for you to eat'

I consider this a striking difference among the languages due to the otherwise parallel evolution of verbal paradigms from Latin to Romance. Castilian, Portuguese, Galician, and Asturian feature all the same simple verbal paradigms *except* for the inflected infinitive.[2] It is not surprising that Galician and Portuguese feature personal infinitives due to their history, as they were originally one common language. What is unusual is that Asturian originally did feature inflected infinitives but no longer does. Instead, in Castilian and Asturian one finds uninflected infinitives with overt subjects, a phenomenon I will call the *personalized* infinitive because of the verb taking an overt subject despite its inherent inability to show agreement.

(2) a. *después de salir José y Ana* (Cast)
then from leave.INF
b. *dempués de salir Xosé y Ana* (Ast)
c. *depois de sair Josê e Âna* (*Pt)
d. *despois de saír Xosé e Ana* (%Gal)
'after José and Ana leave'

Note the differences between (1) and (2) above. A personal infinitive (1) inflects on the verb to show agreement with a subject, whether null or overt. Because of the lack of morphological agreement in (2), however, an overt subject must surface with a personalized infinitive.

The overarching question regarding infinitives with subjects is this: how is it that languages in such proximity, with otherwise similar morphology and syntax, differ so significantly with regard to infinitives with subjects? Portuguese, for its part, is simpler to explain: the nation of Portugal became a sovereign state in 1179, effectively stifling any outside influence on the language. Though Galician and Portuguese were one and the same in antiquity, the establishment of the nation of Portugal left Galicia under the control of the crown of Castile and, consequently, allowed Galician to diverge to some extent from Portuguese due to external influence from Castilian. Finally, the community of Asturias endured even more Castilian influence than did Galician. The predominance of Castilian as a national language during the *Reconquista* and also during the Francoist regime left traces on the

2. Modern Asturian does not have a future subjunctive paradigm *per se*. However, the future subjunctive in Portuguese, Galician, and Old Spanish inherited its morphology from the Latin imperfect active subjunctive, and this morphology in Asturian surfaces as variations on the pluperfect indicative and the imperfect subjunctive. (See Section 6)

grammars of the minority languages of Spain.[3] The continuous contact of Galician and Asturian with Castilian facilitated dialect leveling between the languages over time, a phenomenon called *koineization* (Tuten 2001, 2003, 2021 [this volume]).

This chapter is organized as follows. Section 2 begins with a description of infinitives in Ibero-Romance and how they manifest themselves morphosyntactically, defining differences between *personal* and *personalized* infinitives across the languages. Section 3 is a discussion of the syntax of infinitives with subjects, both personal and personalized, with an explanation of how they can surface grammatically. Section 4 introduces the theories of koineization and prestige effects between Castilian and the Western Iberian languages. Section 5 describes the theories of their genesis of the personal infinitive in Portuguese and Galician, with special attention paid to the disappearance and resurgence of the Galician personal infinitive due to koineization and language revitalization. I conclude in Section 6 with a discussion of personalized infinitives in Castilian and Asturian, exploring the Old Asturian personal infinitive disappearance as a result of contact with Castilian and the observable superstrate influence in the region.

2. Defining personal and personalized infinitives in Ibero-Romance

The personal infinitive, also called the inflected infinitive, is an idiosyncrasy in Romance in that infinitives are generally uninflected, communicating neither tense nor agreement. While it is true that all infinitives are tenseless forms (Stowell 1982), personal infinitives very clearly have subjects and are marked for agreement with those subjects. Other languages besides Galician and Portuguese feature infinitives that take overt subjects, such as Castilian and Asturian. However, there is no agreement morphology on Castilian and Asturian infinitives, which is why they are designated as personalized infinitives in this chapter.

In Portuguese, for example, the personal infinitive is disposed to various functions depending on the matrix verb of the sentence. Wireback (1994: 544) describes these uses in the following examples:

(3) <u>After impersonal expressions</u>
É imprescindível ***partirmos*** *já.* (Pt)
be.PRS.3SG indispensable leave.INF.1PL already
'It is absolutely necessary for us to leave now.' (Thomas 1969: 189)

3. For example, Roseman (1995) and Kabatek (1997a) mention the integration of Castilian elements into Galician lexicon and grammar due to centuries of contact. Konvalinka (1985) discusses the same for Asturian, in great detail.

(4) In declarative predicates
*Garantiu os libros **estarem** no porão.* (Pt)
guarantee.PRF.3SG the-books be.INF.3PL in-the hold
'He guaranteed the books to be in the hold (of the ship).' (Koike 1983: 93)

(5) After perception verbs
*Vi os rapazes **fazerem** a comida.* (Pt)
see.PRF.1SG the-boys make.INF.3PL the-food
'I saw the young people make food.' (Koike 1983: 93)

(6) After emotion verbs
*Eu lamento os deputados **terem** **trabalhado** pouco.* (Pt)
I lament.PRS.1SG the-deputies AUX.INF.3PL work.PRTC little
'I regret the deputies have hardly worked.' (Raposo 1987: 87)

(7) After prepositions
*Ele abriu a cancela para os cavalos **entrarem** no curral.*
he open.PRF.3SG the-gate for the-horses enter.INF.3SG in-the corral
(Pt)
'He opened the gate for the horses to enter the pen.' (Thomas 1969: 189)

The general restriction on placement of personal infinitives is that they cannot be the matrix verb in a sentence. Nor can they be under the control of a matrix modal verb (e.g., **queremos comermos* 'we want to eat'). That is, the subject of the matrix verb must differ from the subject of the personal infinitive. Finally, personal infinitives cannot occur in subordinate clauses with overt complementizers. Thus, according to Wireback, Portuguese personal infinitives are restricted to subject positions (extraposed or not), ungoverned argument positions, and within prepositional phrases.

Galician, according to Longa (1994) and Vázquez Diéguez (2012), is superficially similar to Portuguese in that infinitives can and do feature morphological agreement with subjects. This parallelism is only superficial due to the contexts within which Portuguese structures are equally acceptable in Galician. The following examples from Longa detail instances of Galician personal infinitive structures, namely in epistemic and factive contexts, that differ from Portuguese in acceptability[4]:

(8) a. **Xoan pensa **xantaren** os pais moito.* (Gal)
b. *O João pensa **comerem** os pais muito.* (Pt)>
John think.PRS.3SG eat.INF.3SG the-fathers much
'John thinks his parents eat a lot.' (Longa 1994: 27, Example (4b))

4. Translations of Longa's examples from Galician to Portuguese are mine.

(9) a. **Lamentei* ***traballaren*** *os meus amigos.* (Gal)
b. *Lamentei* ***trabalharem*** *os meus amigos.* (Pt)
lament.PRF.1SG work.INF.3PL the-my-friends
'I regretted my friends working.' (Longa 1994: 27, Example (5b))

Given the asymmetries between Galician and Portuguese of acceptable contexts of personal infinitive use, is the justification for their differences of a historical or sociolinguistic nature, or a combination of both?

There are certain contexts in which the Castilian personalized infinitive coincides with the Portuguese personal infinitive, but in the majority of cases the Castilian equivalents of Portuguese structures involve subordinate clauses with finite verbs. To illustrate one notable exception in Castilian, I reproduce the Portuguese Example (5) below alongside its Castilian equivalent.

(10) a. *Vi os rapazes* ***fazerem*** *a comida.* (Pt)
b. *Vi a los muchachos* ***hacer*** *la comida.* (Cast)
'I saw the young people make food.' (Koike 1983: 93)

Perception verbs, such as *ver* above but also *oír* 'to hear', *escuchar* 'to listen', and *sentir* 'to feel', readily admit overt subjects with predicative infinitives in Castilian – personalized infinitives in the parlance of this chapter. In the case of Example (10), both Portuguese and Castilian show 'the young people' as the direct object of 'saw' and simultaneously the agent of 'make'. Outside of these contexts, however, finite verbs in subordinate clauses are the only grammatical equivalents to Examples (3, 4, 6, 7) in "standard" Castilian, as I provide here:

(11) After impersonal expressions
Es imprescindible que partamos ya.
'It is absolutely necessary that we leave now.

(12) In declarative predicates
Garantizó que los libros están en la bodega.
'He guaranteed that the books are in the hold (of the ship).'

(13) After emotion verbs
(Yo) Lamento que los diputados hayan trabajado poco.
'I regret that the deputies have worked little.'

(14) After prepositions
Él abrió la cancela para que los caballos entren en el corral.
'He opened the gate for the horses to enter the pen.'

Observations by Viejo Fernández (2015) show that Asturian patterns morphologically like Castilian in that agreement morphology does not surface on infinitives in the presence of an overt subject. However, he also notes that the more widespread

use of infinitives with subjects shows Asturian usage to be more similar to Galician and Portuguese than to Castilian.

(15) *Alégro =me de **venir** tu* (Ast)
make-happy.PRS.1SG =REFL.1SG of come.INF you
'I'm happy that you come' (Viejo Fernández 2015: 11, Example (11a))

(16) *Alegro-me por você **estar** passando bem…* (Pt)
by you.FORM be.INF.3SG pass.PROG well
'I'm glad that you are doing well'
(*De Reynaldo Purim para Artur Purim*, in Purim 2012)

(17) *Me =alegro de que **vengas** (*de venir tú/tú venir)* (Cast)
REFL.1SG =make-happy.PRS.1SG of that come.PRS.SBJV.2SG
'I'm happy that you come' (Viejo Fernández 2015: 11, in-line example)

Despite the divergence of Castilian in (17), as we will see in Section 6, there is more parallelism between Castilian and Asturian personalized infinitives than Viejo Fernández communicates.

In the following section, I present details as to how personal and personalized infinitives can be syntactically possible using notions of government, Case assignment, and abstract agreement.

3. Syntax of infinitivals with subjects

3.1 Government and control

Under certain circumstances the phenomenon of infinitives with explicit subjects can be grammatical. The most important operation in permitting overt subjects with infinitival verbs is that of *government*. This relationship is what allows for the assigning of certain features such as Case. It is held that matrix verbs, in general, govern all their arguments. The formal definition of government employed in this study is as follows (adapted from Haegeman 1994)[5]:

Government. A *governs* B if and only if
1. A is a head,
2. A m-commands B, and
3. there is no barrier between A and B.

Infinitives are commonly observed in Castilian as arguments of finite matrix verbs, either as subjects or complements, frequently marked as having accusative Case.

5. From Haegeman (1994), A *m-commands* B if and only if A c-commands B and the node that directly dominates A also dominates B. Maximal projections (XP) act as barriers to government.

Let us observe Example (18), which presents one grammatical and one ungrammatical sentence:

(18) a. *Lamento perder los documentos.* (Cast)
lament.PRS.1SG lose.INF the-documents
'I regret losing the documents.'
b. * *Lamento (ellos)* ***perder*** *(ellos) los documentos.* (Cast)
'I regret them losing the documents.'

In (18a) one can see that the constituent *perder los documentos* is the complement (marked as Accusative) of the matrix verb *lamento*. More specifically, *perder* itself is the accusative argument of *lamentar* under government (Raposo 1987). Accordingly, it is true that *lamentar* governs *perder*, prohibiting it from taking an overt subject; this is evidence that infinitives in complement position cannot be personalized. How does this structure differ from those in which a personalized infinitive is acceptable?

Accounts by de Mello (1995), Pöll (2007), and Sitaridou (2007) claim that such structures should not exist due to infinitives' prescriptively governing the silent category PRO as their subjects. According to Ortega Santos (2002: 21), an infinitive governed by a matrix verb cannot take its own subject, whereas it is possible for a non-governed infinitive to take a subject. Evidence of this can be seen in the following Examples (19)–(20) from Castilian, in which the infinitive in (19) is not governed and takes its own subject, while the infinitive in (20) is governed/controlled by the matrix verb and cannot take a subject:

(19) *Todo el mundo se levantó al* ***leer*** *el juez el veredicto.*
… to-the read.INF the-judge the-verdict
'Everyone stood up for the judge's reading of the verdict.'
(Ortega Santos 2002: 23, Example (4))

(20) * *Pedro les= obligó a* ***ir*** *los taxistas.*
Pedro DAT.3PL=oblige.PRF.3SG to go.INF the-taxi-drivers
'Pedro made the cabbies go.' (Ortega Santos 2002: 25, Example (16))

In Example (19), *levantarse* does not govern *leer* within the prepositional phrase. *Levantarse* can only take one predicate argument, which is shown in the lexical-conceptual structure below in (21a). By contrast, *obligar* in (20) takes two mandatory predicate arguments as shown in (21b).

(21) a. *levantar(se)* X cause [Y rise][6]
b. *obligar* X cause [Y must do [Z]]

6. In the case of reflexive verbs such as *levantarse*, X = Y.

It can be said, then, that *levantarse* has no capacity to govern another element due to its only argument necessarily being the subject. The third argument of *obligar*, Z, in Example (20) is the verb *ir*. The ungrammaticality of (20) lies in the argument Y, which is *los taxistas*. This should be the dative argument of *obligar* but is not; assigning it to the subject position of *ir* causes the failure.

Though Castilian verbs absorb the agreement of the infinitives they govern, there is still a possible way for the infinitive to have its own subject. This involves the notions of abstract agreement (Ortega Santos 2002; Pöll 2007) and coindexation with the silent infinitive subject PRO. I discuss these theories in the following subsection.

3.2 Licensing and abstract agreement, and coreference with PRO

We have observed in the previous section how the predicate structure makes it possible for an infinitive to take an overt subject. To first explain the interpretation of NPs as subjects of infinitives, we turn to the notion of *licensing*. To say that a constituent is licensed is to claim that it has an appropriate interpretation at Logical Form (LF) and Phonetic Form (PF). For example, the Extended Projection Principle (EPP) states that any finite VP must function as a predicate and must take a subject (Alexiadou & Anagnostopoulou 1998). Although agreement is normally a licensing characteristic of finite verbs (Alers 2009), this property extends to the Portuguese and Galician personal infinitives. Let us examine two examples from Portuguese, one inflected and the other uninflected.

(22) a. *Será* *difícil* *eles* ***aprovarem*** *a proposta*.
be.FUT.3SG difficult they approve.INF.3PL the-proposal
'It may/will be difficult for them to approve the proposal.'

b. *Será* *difícil* ***aprovar*** *a proposta*.
be.FUT.3SG difficult approve.INF the-proposal
'It may/will be difficult to approve the proposal.'

(Raposo 1987: 86, Example (2a)–(c))

Example (22a) shows morphological agreement with an overt subject, while (22b) does not. According to Raposo (1987), the distribution of inflected and uninflected infinitive clauses in Portuguese is complementary with respect to the subject that it can take. Inflected infinitives in Portuguese obligatorily agree with a subject, whether overt or null. Contrastively, the uninflected infinitive takes PRO as its subject by default. In comparison to Castilian, both languages exhibit the same pattern regarding uninflected impersonal infinitives:

(23) a. *Será difícil PRO* ***aprobar*** *la propuesta.* (Cast)
b. *Será difícil PRO* ***aprovar*** *a proposta.* (Pt)
'It may/will be difficult to approve the proposal.'

It is impossible for Castilian infinitives to license overt subjects under regular Agreement. Various accounts (Ortega Santos 2002; Pöll 2007) maintain that Castilian infinitives license overt subjects under *abstract agreement*. As was mentioned earlier, the infinitive governed by the matrix verb of the phrase is understood to have the same subject, as in (18a) reproduced here.

(24) *(Yo) Lamento* ***perder*** *los documentos.*
'I regret losing the documents.'

So that it is understood that the subject (*yo*) of the phrase is the agent associated with the infinitive (*perder*), it is said that the matrix verb "absorbs the abstract agreement of the infinitive, thus depriving it of its ability to license an overt subject" (Ortega Santos 2002: 21–22). But if the infinitive is ungoverned by the matrix verb, there is no Agreement absorption. Consider Example (25):

(25) *Al* ***entrar*** *yo por la puerta, todos se= callaron.* (Cast)
to-the enter.INF I through the-door all REFL.3PL=hush.PRF.3PL
'Upon my entrance through the door, everyone became silent.'
(Ortega Santos 2002: 20, Example (1a))

The lexical-semantic scheme of the verb *callar* shows that it can only govern two arguments: a subject (nominative) and an object (accusative). In this case, due to the reflexive nature of the verb (*callarse*), the subject *todos* is assigned both Cases, thus satisfying government. The result is that any other constituent within the phrase is an adjunct and not a complement, and matrix verbs cannot govern adjuncts by convention. As such, the verb *entrar* is ungoverned and can consequently take an overt subject. How, then, is Case assigned to this subject? It is assigned in Surface Structure (SS) and it is understood that the subject *yo* is the agent of the action of *entrar*.

Ortega Santos (2002) and Pöll (2007) maintain that, syntactically, the subject of an infinitival clause is not the overt, pronounced subject. Both their accounts mention the presence of PRO, the actual subject of all infinitives under EPP. Pöll states that the NP in argument position of the matrix verb is not the true subject of the embedded infinitive. Instead, this NP is coindexed with PRO which is already in subject position of the infinitive. In (26) below, we see the relationship between the direct object of the matrix verb and PRO:

(26) *Vi* *a Juan$_i$ PRO$_i$* ***trabajar*** (Cast)
see.PRF.1SG Juan work.INF
'I saw Juan working'

What is observed here is the licensing of *Juan* as the direct object of *ver*. Since this NP already has an assigned Case, it cannot receive another because *ver* is not reflexive. Therefore, the coreferentiality between *Juan* and subsequent PRO permits the same entity to function as the direct object of the matrix verb *ver* and the embedded infinitive *trabajar*.

4. Theories of contact, convergence, and divergence: Koineization in Galician and Asturian

Though Asturian and Galician have historically held more in common with each other with regard to morphosyntax than they have with Castilian (Spanish), the former two have been in constant contact with the latter for centuries. As Tuten (2003) explains, population growth and dwindling land resources in the northeastern Iberian Peninsula led inhabitants of Asturias to migrate to the south in the 8th and 9th centuries, encroaching upon Castile and its capital city of Burgos. During later stages of the *Reconquista*, the Kingdoms of Galicia (later becoming part of León) and Asturias were consolidated into the rapidly expanding Kingdom of Castile. As Castile grew in power and influence, the Castilian dialect did as well, eventually coming to be the official language of the newly annexed lands. Thus, natives of Galicia and Asturias who, until that time, had spoken their regional languages ultimately were required to learn Castilian for legal, educational, and business purposes.

The expansion of Castile into the northwestern regions of Iberia, along with the migration of Asturian and Galician citizens into the original Castilian territory, placed various dialects in contact with one another, which according to Tuten "increased the variation in social and linguistic norms … at the same time that it weakened social networks: the very conditions that promote koineization" (2003: 100; also cf. Tuten 2001). *Koineization* is defined as the genesis of a new language or dialect through contact between two or more mutually intelligible dialects in constant contact. As speakers accommodate each other's linguistic differences and speak using features common to both dialects, this new dialect spreads through the community "within one or two generations" (Tuten 2021 [this volume]). It then becomes regularized through acquisition, leveling, and simplification by subsequent generations (Tuten 2003: 90–91). Tuten's work focuses mainly on phonological and morphological leveling between Castilian and other Ibero-Romance languages, particularly diphthongization patterns (phonology) and pronominal paradigms

such as the system of possessives. The idea of koineization, though, is also applicable to the other languages with which Castilian has come into contact on the Iberian Peninsula.

For example, Kabatek (1997a) discusses the decision to standardize Galician as an independent language, similar to yet distinct from Portuguese and with certain Castilian elements discarded. several other phenomena present in Galician that arose in the period during which the language was still undergoing koineization. Galician phonology is known for featuring a fricativized allophone of /g/: speakers instead realize this phoneme as a voiceless pharyngeal or velar fricative [ħ ~ x] in a phenomenon called *gheada*. Thus, the word *grande* 'large' with an underlying representation of /gɾande/ would instead be realized as [ħɾaNdɪ ~ xɾaNdɪ]. Use of *gheada* is considered to be a marker of Galician "authenticity" free from Castilian influences, though it is commonly reserved for informal situations.

Another sociolinguistically conditioned form in Galician is the colloquial dative of solidarity, in which the second-person singular dative clitic *che* surfaces in a phrase with no grammatical referent.

(27) *Pois o que é a min non* ***che=*** ***me=***
well it that be.PRS.3SG to me.DAT NEG COLLOQ 1SG.DAT
deu *nada*
give.PRF.3SG nothing
'Well, as far as I am concerned he/she never gave me anything.'
(in Kabatek 1997b: 190, Example (8))

The interlocutor in (27) (as strong pronouns, nominative *tu* and dative *ti*) is not part of the argument of the verb in this example. Instead of fulfilling a function of coreference, it instead serves as a sort of pragmatic "bridge" between speaker and hearer. This pronoun is frequently encountered in spoken Galician, in great contrast to neighboring Castilian and Portuguese which do not feature any solidarity pronoun of this sort; as such, according to Kabatek, "already Castilianized authors of the nineteenth century consider[ed] the 'solidarity pronoun' as a typically Galician element and to use it as a marker of distance between Galician and Spanish […]" (1997b: 191).

Aside from the morphological manifestations of infinitives taking subjects (the topic of the current chapter), Galician features a high level of variability in clitic positioning in modal and restructuring contexts. Consider the following examples:

(28) a. *teño que dicilo* (Cast *tengo que decirlo*, Pt *tenho que dizê-lo*)
b. *teño que o decir* (Cast **tengo que lo decir*, Pt *tenho que o dizer*)
c. *teño o que decir* (Cast **téngolo que decir*, Pt *tenho-o que dizer*)
'I must say it' (Kabatek 1997b: 189, Examples (4)–(6))

Example (28a), with the clitic *o* remaining attached to the right of the infinitive *dicir*, is the arrangement most akin to Castilian and is also grammatical in Portuguese. We see *o* procliticized to *dicir* in (28b), an arrangement not found in Castilian but present in Portuguese. Finally, Example (28c) shows the clitic attached to the right of the main verb, which is standard in (European) Portuguese but no longer so in Castilian. Kabatek observes that all three forms are observed in spoken Galician, with (a) and (c) the most widely encountered. Form (b) is much more frequent in rural zones than urban areas, and it tends to be a feature of formal written texts that prioritize the preservation of Galician forms that are in decline in modern use much like the inflected infinitive.

Departing from Tuten's focus on koineization in medieval Spanish, I propose that koineization affected the salience of personal infinitives in medieval Asturian and personalized infinitives in Castilian. Sitaridou (2007, 2009) and Beardsley (1921) note that many Romance languages have employed infinitives with overt subjects; it is peculiar, though, that of all the languages that feature infinitives with subjects, only those from Western Iberia have or have had evidence of inflected forms. My theory of koineization affecting the prevalence of personal infinitives is largely incident upon Asturian, which originally featured inflected infinitives but no longer does so, and Galician, which underwent a resurgence in inflected infinitive structures after its restitution as a legal language.

In the following two sections, I will discuss the divergences of Galician and Asturian from Portuguese regarding infinitives with subjects due to influence from Castilian.

5. History and distribution of personal infinitives: Portuguese and Galician

There have been several differing accounts of the origins of the inflected infinitive in Galician-Portuguese. One such argument is the *creative* argument, which proposes that the inflected infinitive arose from the standard (uninflected) infinitive and gained the full complement of Galician-Portuguese verbal agreement morphology. The second theory is the *analogy* argument, which claims that the personal infinitive arose due to similarities between infinitives and the future subjunctive paradigm. The third theory claims that the personal infinitive is a descendant of the Latin imperfect active subjunctive, both in morphology and use. Following Wireback (1994), Pires (2002), and Harris (2013), I will detail these arguments in brief here.

5.1 Creative argument

According to Harris (2013), the creative argument involves the insertion of overt subjects followed by the affixation of agreement morphology to otherwise uninflected infinitives. This phenomenon would likely have followed a process of parameter resetting, first to allow subjects and then to change the value of the Agreement feature of the infinitive, much like the following stages as outlined below in Portuguese[7] (from Harris 2013: 304, Examples (5)–(7)).

(29) *É* *necessário* *PRO* ***terminar***$_{[\text{-T(ense)}][\text{-Agr(eement)}]}$ *a tarefa.*
be.PRS.3SG necessary finish.INF the-work
'It is necessary to finish the work.'

(30) *É* *necessário* *nós* ***terminar***$_{[\text{-T}][\text{-Agr}]}$ *a tarefa.*
be.PRS.3SG necessary we finish.INF the-work
'It is necessary for us to finish the work.'

(31) *É* *necessário* *(nós)* ***terminarmos***$_{[\text{-T}][\text{+Agr}]}$ *a tarefa.*
be.PRS.3SG necessary (we) finish.INF.1PL the-work
'It is necessary for us to finish the work.'

The process shown here begins as (29) with an uninflected infinitive with an impersonal subject. The second stage of the process, Example (30), involves an overt subject associated with the infinitive, the statement now communicating who will be finishing instead of making a general comment. Note the morphosyntactic features on the infinitive are identical between (29) and (30), with neither tense nor agreement; this is a personalized infinitive. The greatest change is observable in (31) which contains the overt subject *nós* plus agreement morphology on the infinitive. In this case, the infinitive now contains an agreement feature while still not being marked for tense, identifying it as a personal infinitive.

Both Wireback (1994) and Harris (2013) discount the creative approach based on relatively low frequencies of inflected infinitives with overt subjects in Old Portuguese. Harris claims that a reasonable expectation of a personal infinitive derived from an uninflected infinitive would include a higher frequency of overt subjects in Old Portuguese. Furthermore, uninflected infinitives with overt subjects such as (30) are unacceptable in modern European Portuguese.[8]

---

7. These examples would be identical in Galician but for the spelling of *necessário* (Gal *necesario*).

8. However, there is variation between inflected and uninflected infinitives with overt subjects in colloquial Brazilian Portuguese (see Gorski 2000 for an empirical study). Also, (30) is acceptable in modern Galician; I return to this later.

5.2 Analogy argument

A second argument as to the history of the inflected infinitive in Portuguese and Galician is via *analogy* with other forms, most conspicuously the modern Portuguese and Galician future subjunctive. Examples of the relevant paradigms are shown in (32–33):

(32) Personal infinitive, '___ to speak'
Gal *falar, falares, falar, falarmos, falardes, falaren*
Pt *falar, falares, falar, falarmos, falardes, falarem*

(33) Future subjunctive, 'that ___ will speak'
Gal *falar, falares, falar, falarmos, falardes, falaren*
Pt *falar, falares, falar, falarmos, falardes, falarem*

Regular verbs in Galician and Portuguese are phonologically identical between the personal infinitive and future subjunctive paradigms. According to the analogy argument, personal morphology as already present on the future subjunctive was added to the noninflected infinitive, thus creating the syncretism.

While an argument for parallels between the future subjunctive and the personal infinitive might be sustainable for verbs with regular stems, such as Galician and Portuguese *falar* 'to speak' which show morphological syncretism in the personal infinitive and future subjunctive paradigms, more argumentation than what analogy can provide is necessary to extend the analysis to cover irregular verbs.

(34) Personal infinitive, '____ to do, ____ to make'
Gal *facer, faceres, facer, facermos, facerdes, faceren*
Pt *fazer, fazeres, fazer, fazermos, fazerdes, fazerem*

(35) Future subjunctive, 'that ___ will do/make'
Gal *fixer, fixeres, fixer, fixermos, fixerdes, fixeren*
Pt *fizer, fizeres, fizer, fizermos, fizerdes, fizerem*

Note that the verb chosen for the examples – Galician *facer* and Portuguese *fazer* – illustrates the principal shortcoming of the analogical history theory. Harris' greatest argument against the analogy theory, with which I agree, is that there is no accounting for the lack of irregular stems in personal infinitives which appear in the future subjunctive. In fact, most verbs in Galician and Portuguese with rhizotonic preterite forms, such as *ser* and *estar* 'to be', *poder* 'can', Galician *poñer* and Portuguese *pôr* 'to put, place', *ir* 'to go', and *querer* 'to want' will present similar stem allomorphy to that of *facer/fazer* between the personal infinitive and the future subjunctive. Moreover, Harris, following Wireback (1994), notes that the Portuguese personal infinitive and future subjunctive only share one syntactic context: following the preposition *depois* 'after', an insufficient basis for an analogical argument.

5.3 Latin imperfect subjunctive argument

The most sustainable and widely accepted argument for the origin of the personal infinitive in Portuguese and Galician, based on Wireback (1994) and Harris (2013), is that the personal infinitive reflexes evolved from the Latin imperfect active subjunctive. The morphophonological evolution from Latin to Portuguese and Galician appears in the following table:

Table 1. Imperfect subjunctive to personal (inflected) infinitive: 'to do, to make'

P/N	Latin reflex	Portuguese reflex	Galician reflex
1SG	facerem	fazer	facer
2SG	facerēs	fazeres	faceres
3SG	faceret	fazer	facer
1PL	facerēmus	fazermos	facermos
2PL	facerētis	fazerdes	facerdes
3PL	facerent	fazerem	faceren

The Latin imperfect subjunctive paradigm was "available" due to the evolutionary conversion of the Latin pluperfect subjunctive to the Portuguese imperfect subjunctive (Martins 2001), which was a change that had already been occurring for centuries in Latin.[9] Harris notes that the imperfect-subjunctive theory, itself not perfect, provides the strongest morphological argument for the appearance of the inflected infinitive in Western Iberian due to the regularity of the verbal stems found in his corpus. His corpus study revealed instances of the inflected infinitive occurring in syntactic contexts analogous to Latin constructions containing the imperfect subjunctive, particularly in complement clauses and prepositional phrases:

(36) a. *Pervenit, priusquam Pompeius sentire*
arrive.PRF.3SG before Pompei.NOM sense.INF
posset. (Lat)
can.IMPERF.SBJV.3SG
'He arrived before Pompei could become aware of his approach.'

b. *el rey d. diniz chegou… ante de* ***darem***, *e*
… before of give.INF.3PL and
pronunciarem *[as] dictas*. (OPt)
pronounce.INF.3PL … the maxims
'King Diniz arrived… before giving and pronouncing the maxims.'
(Harris 2013: 309, Example (18))

9. Harris (2013) states that the Latin pluperfect subjunctive was the primary past subjunctive form dating back to the 3rd century AD. This changeover appears to have affected the past subjunctive morphology for other Romance languages as well.

Example (36a), which Harris cites from Scida (2004), is an instance of a Latin imperfect subjunctive in an adverbial clause, while (36b) is a similarly constructed occurrence from Old Portuguese, but with the clearly inflected infinitives *darem/pronunciarem* instead of imperfect subjunctive forms *dissessem/pronunciassem*.

This parallelism is significant in that, according to Scida (2004: 106–109), the Latin imperfect subjunctive and Portuguese personal infinitive could be found in several similar contexts, such as complement, temporal, purpose, resultative, concessive and gerundive clauses. Scida notes that both the Latin imperfect subjunctive and Galician-Portuguese personal infinitive occur following certain prepositions. Harris's corpus study notes the same, but he also found personal infinitives in subject clauses, a use which does not overlap with the Latin imperfect subjunctive but more closely parallels noninflected infinitives in Romance. Subject clauses aside, the correspondence between the forms in the two languages lends to the credibility of the Latin imperfect subjunctive as the source for the Portuguese and Galician personal infinitives.

5.4 Assimilation and dissimilation in Galician: Effects of contact with Castilian

Though the personal infinitive exists in Galician, as mentioned throughout this text, its use is much less frequent than in Portuguese (Brea 1985; Merlan 2007). The Galician personal infinitive is ordinarily found more frequently in high literature or formal style and not as frequently in colloquial speech. Despite the common ancestry of Galician and Portuguese, this is one of the greatest divergences in use between the two languages; the Portuguese personal infinitive surfaces in all registers of the language, not just formal or literary contexts. The regression of the personal infinitive in Galician is generally attributed to influence from Castilian, which for a considerably long period has enjoyed higher status in Galicia than Galician itself (Roseman 1995). This was never of concern in Portuguese, as Portugal remains a sovereign nation (and has been since the 13th century), whereas Galicia has been under Spanish control for much of its history.

Nevertheless, with renewed interest and pride in Galician, use of forms not found in Castilian has increased. Kabatek (1996) notes that the following verse from a Galician translation of the Bible employs the inflected infinitive more frequently than does a Portuguese equivalent:

(37) *e pecharon os seus ollos, para non* ***veren*** *con ollos, nin*
and close.PRF.3PL the-their-eyes for not see.INF.3PL with eyes nor
oíren *con oídos, nin* ***entenderen*** *co seu corazón*
hear.INF.3PL with ears nor understand.INF.3PL with their heart

nin ***se=converteren***
nor REFL.3PL=convert.INF.3PL
'and their eyes they have closed; lest at any time they should see with their eyes and hear with their ears, and should understand with their heart, and should be converted' (Matthew 13:15; Kabatek 1996: 384)

Kabatek's explanation for this is that the supposed overuse of the personal infinitive where it is not found in Portuguese is to "create a text with a high or solemn diaphasic mark" (1996: 384, translation mine). Use – and overuse – of characteristically Galician forms instead of the Castilian equivalents may be a manner in which Galician writers restore prestige to their regional dialect as pride in the language is revived (Kabatek 1996, 1997a, 1997b).

Besides the different contexts in which personal infinitives can be found in Galician while not in Portuguese, it is notable that Galician does allow the personalized infinitive (see Example (30)), which was only sporadically attested in Old Portuguese (Wireback 1994) and now is disallowed in Modern European Portuguese (Raposo 1987; Harris 2013). On the other hand, the personalized infinitive remains the only option in Castilian. I take the position that, instead of being a vestige of a sporadic form from antiquity, the Galician personalized infinitive reflects a lasting Castilian influence on the grammar and an effect of the koineization phenomenon.

6. The history and syntax of personalized infinitives: Castilian and Asturian

As discussed in Section 5, various theories exist concerning the origin of the personal (inflected) infinitive in Portuguese and Galician. One such theory detailed by Wireback (1994) and Harris (2013) is the *creative* theory, which states that the personal infinitive arose from uninflected infinitives that later took overt NPs as their subjects. Though Harris concluded that this was unlikely to be the reason for the development in Galician and Portuguese, there is still the possibility that the creative theory led to the personalized infinitives found in Castilian and modern Asturian. I consider this possibility to be unlike Galician and Portuguese because, as Harris stated, noninflected infinitives appeared very infrequently with overt subjects in Galician and Portuguese while being salient in Old Castilian, according to Sitaridou (2009).

What I propose, however, is an alternate theory of the creation of the Asturian personalized infinitive. The evidence available to us suggests that the creative theory is insufficient to explain the personalized infinitives in Asturian. I arrive at this conclusion based on the already-present personal infinitive in medieval Asturian which

is now lost, unlike that of Galician and Portuguese. Instead of an explanation based on morphosyntactic analysis or theory, I offer an analysis based on the linguistic contact between Asturian and Castilian. I will extend this analysis to Galician, itself in a diglossic situation with Castilian, which in spoken language does not readily admit the personal infinitive nearly as frequently as does Portuguese (Kabatek 1996, 1997a; Merlan 2007).

6.1 Castilian

The Castilian personalized infinitive has often been ignored by grammarians and descriptivists due to the lack of uniformity of its distribution. Their use is considered marginal at best and aberrant at worst (Penas Ibañez 2014), and the structure is not taught or discussed in schools. However, personalized infinitives are extremely common in the Caribbean, notably in Puerto Rican Spanish with preverbal subjects which are not usually found in other regions (Brown & Rivas 2019). Lipski (1991) expounds upon this further, giving examples of personalized infinitives in the Spanish of Mexico, Colombia, Paraguay, Chile, and even Spain, as well as non-Hispanophone countries such as the Philippines and Trinidad and Tobago.

The Castilian personalized infinitive is a perfectly acceptable construction in predicate arguments, in subject clauses, and within prepositional adjuncts (cf. de Mello 1995; Pöll 2007; Sitaridou 2009). The following examples illustrate constructions found quite frequently in Castilian:

(38) Within predicate arguments

a. *Vi a Juan* ***trabajar*** *demasiado.*
see.PRF.1SG Juan work.INF too-much
'I saw Juan work too much.'

b. *Lo=dejó* ***leer*** *la carta.*
ACC.3SG=leave.PRF.3SG read.INF the-letter
'S/he let him read the letter.' (Pöll 2007: 94, Example (1))

(39) Within prepositional adjuncts

a. *Juan se =fue antes de* ***cantar*** *María.*
Juan REFL.3SG =go.PRF.3SG before of sing.INF Maria
'Juan left before Maria sang.'

b. *Sin* ***saber*** ***=lo*** *yo / yo* ***saberlo***
without know.INF =ACC.3SG I / I know.INF = ACC.3SG.
Pedro se =compró un coche.
Pedro REFL.3SG =buy.PRF.3SG a-car
'Without me knowing, Pedro bought himself a car.'
(Pöll 2007: 94, Example (2))

(40) Within infinitival subject clauses

a. ***Presentar =se*** *Julia a las elecciones fue un error.*
present.INF=REFL.3SG Julia to the-elections be.PRF.3SG an-error
'It was a mistake that Julia ran in the election.'

b. ***El cantar*** *María la Traviata sería novedoso.*
the sing.INF Maria *La Traviata* be.COND.3SG novel
'Maria singing the Traviata would be unexpected.'

(Pöll 2007: 95, Example (3))

Note that, within adjuncts and subject clauses, the overt subject is customarily located after the infinitive. Many variants of Spanish disallow preverbal subjects in these contexts. It is worth mentioning, however, that in Caribbean Spanish preverbal subjects are routinely found in these contexts (Berger 2015; Ordóñez 2018). According to Brown and Rivas (2019), this is a logical extension of the tendency of Caribbean Spanish speakers to express subjects before verbs in otherwise impermissible situations such as interrogatives.

(41) *Adónde él ha ido?* (Puerto Rican Spanish)
to-where he AUX.3SG go.PRTC
'Where has he gone?' (Brown and Rivas 2011: 25, Example (5))

Furthermore, de Mello (1995) presents evidence of preverbal subjects with infinitives in higher registers (*habla culta*) of Spanish spoken in La Paz, Bogotá, Buenos Aires, Madrid, and other major cities outside the Caribbean.

Sitaridou considers the possibility of the infinitive with a nominative subject having been inherited by Castilian from Latin, in the form of a phenomenon called "accusativus cum infinitivo" (AcI). In this construction, the accusative complement of the finite matrix verb also serves as the subject of another infinitive. In the following example from Latin, it is shown that, even in a precursor language to Castilian, the infinitival verb does not agree morphologically with its subject.

(42) *Gallos* ***esse*** *altissimos omnes credunt* (Lat)
Gauls.ACC be.INF very-tall men.NOM believe.PRS.3SG
"Everybody thinks that the Gauls are very tall"
(Pillinger 1980: 56, Example (1a), in Sitaridou 2009: 39, Example (6))

Note that in this example *gallos* is the direct object of the verb *credere* and that *altissimos* within this complement modifies the noun; it also is declined in accusative case to agree with said noun. Most interesting about the example is the presence of *esse* within the constituent; it is understood that *gallos* is the subject of this verb. This AcI arrangement, claims Sitaridou, was one of the origins of the personal infinitive in complement position in Old Castilian and was a learned structure; it has since been lost, though, in Modern Castilian.

As in Modern Castilian, the personalized infinitive in Old Castilian appears in adjuncts and in subject clauses. It is evident, however, that some of the contexts within which personalized infinitives were found in Old Castilian are no longer permissible in the modern variant. We consider the following three examples:

(43) *Tan grant verguenca al menor dellos commo al mayor del mundo por ende ssennor por yo* ***ffazer*** *seruicio a cassio…* (OCast; %Cast)
… for I do.INF service to Cassius
"Such embarrassment for both great and small, for my having helped Cassius"
(*General estoria VI*, Alfonso X; in Sitaridou 2009: 40, Example (7))

(44) *ca bien sabedes vos que*
non es gujsado de ***besar*** *yo la mano a vos.*
… NEG be.PRS.3SG advised of kiss.INF I the hand to you.FORM (OCast; %Cast)
"Because you know well that it is not advisable of me to kiss your hand"
(*Cronica de veinte reyes*; in Sitaridou 2009: 40, Example (8))

(45) *Los quales creerían yo no* ***haber leído*** *las reglas*
them which think.COND.3PL I NEG AUX.INF read.PRTC the rules
'… those who would believe me to have not read the rules' (OCast; *Cast)
(*Los prouerbios de D. Yñigo Lopez de Mendoca con su glosa*; in Sitaridou 2009: 38, Example (5))

Examples (43)–(44) are variably acceptable in Modern Castilian where personalized infinitives persist, if we recall the theories of control and abstract agreement from Section 3 of this chapter. However, Sitaridou mentions that a personalized infinitive in complement position is acceptable in Old Castilian yet ungrammatical in modern Castilian (cf. Examples (11)–(14)), as evidenced by the acceptability judgments in (45). Moreover, this is a crucial distinction between the personalized infinitive in Castilian and the inflected infinitive in Portuguese and Galician.

Returning to the problem of whether the personalized infinitive in modern Castilian was entirely a direct descendant of AcI from Latin, Sitaridou argues against it for three reasons: (a) the subject did not carry accusative Case, (b) it was not selected by different main verbs, and (c) the construction was hardly productive. While not presenting any complications for personalized infinitives in subject position, these reasons prevent the AcI construction from being the source of the personalized infinitive in adjunct position. Instead, Sitaridou (2004, 2009) posits an acquisitional scenario in which subjects following prepositional adjuncts containing infinitives are reanalyzed as the subjects of those infinitives themselves, as in the following example:

(46) a. *[Antes de morir] ÉL*[10] era feliz
'Before dying, he was happy'
b. *[Antes de morir ÉL] era feliz*
'Before his dying, he/she/it was happy'
(Sitaridou 2009: 51, Example (22b)c)

In (46a), the pronoun *él* is the subject of the main verb *ser* (*era*). After reanalysis, *él* becomes the subject of the infinitive within the adverbial clause, and *era* takes a null subject coreferent with *él*. Sitaridou's position, also expressed in Pires (2002) for Brazilian Portuguese, is that the reanalyzed structure in (46b) was passed down through the generations and became the default structure in the grammar.

6.2 Asturian

In contrast to Portuguese and Galician, Asturian does not feature agreement morphology on infinitives with overt subjects. It would be reasonable to describe this language, then, as possessing a personalized infinitive much like Castilian. Viejo Fernández (2015) outlines some of the uses of the personalized infinitive in Asturian, examples of which follow. He notes correctly that, though these structures make Asturian appear more similar to the Western Ibero-Romance languages of Portuguese and Galician, the lack of agreement morphology on the infinitive points to a similarity between Asturian and Castilian.

(47) ***Lladrar*** *los perros alertó=nos* *de la presencia d'estraños*
bark.INF the-dogs alert.PRF.3SG=ACC.1PL of the-presence of-strangers
'The barking of the dogs alerted us to the presence of strangers' (Ast)
(Viejo Fernández 2015: 9, Example (6))

(48) *Depués de* ***venir*** *Xuan, les coses meyoraron*
after of come.INF Xuan the-things improve.PRF.3PL
'After Xuan's coming, things improved' (Ast)
(Viejo Fernández 2015: 14, Example (20a))

(49) *Ye culpable de* ***perde=se*** *los neños*
be.PRS.3SG guilty of lose.INF=REFL.3SG the-children
(y ***manca=se****)* (Ast)
(and hurt.INF=REFL.3SG)
'He is guilty of the children getting lost (and hurting themselves)'
(Viejo Fernández 2015: 13, Example (16b))

10. Sitaridou capitalizes the pronoun here to show contrastive focus, the context of which I have omitted here.

Example (47) is structurally similar to Example (42), with a nominalized infinitival phrase in subject position taking an overt subject. The personalized infinitive in (48) appears in an adverbial phrase in adjunct position. In (49) we have a change in subjects between the matrix *ser* and the complement *perdese*.

Based on accounts of the personalized infinitive in Castilian, it would appear on the surface that the equivalent construction in Asturian developed in the same manner. As stated earlier, various authors (Merlan 2007; Viejo Fernández 2015) correctly remark that modern Asturian lacks the personal infinitive construction found in Portuguese and Galician. However, regarding the history of the Asturian personalized infinitive, García Arias (2003) remarks that during the medieval era, agreement morphology on infinitives *was* productive, in contrast to the modern language. Lorenzo González (2016) provides the following example from medieval Asturian, among others which were taken from 13th-century documents:

(50) *otorgamos et prometemos de non* ***venirmos*** *contra ello en nengun tiempo*
… of NEG come.INF.1PL against it in no time
'we grant and we promise not to come against it at any time'
(*Archivo Catedrático y Diocesano de Salamanca* 1276;
in Lorenzo González 2016: 45)

Though there is no overt subject in (50), the reflex in this example and the syntactic environment in which it occurs demonstrates the presence of the personal infinitive in medieval Asturian. The agreement morphology, furthermore, is identical to that of Portuguese and Galician: {Ø, -es, Ø, -mos, -des, -en}. The question of the loss of the personal infinitive in Asturian remains, however. What could have caused the complete loss of this grammatical category in the language? I propose an explanation of how contact between Castilian and Asturian produced lasting effects on the verbal morphosyntax of the latter.

6.3 Contact with Castilian and koineization in Asturian

Recalling an argument by Uriagereka (1995a, 1995b) that languages which are morphologically "strong" routinely feature inflected infinitives,[11] I first establish that the Western Iberian group – traditionally comprising (European) Portuguese, Galician, Asturian, Leonese, and other neighboring dialects – is usually regarded as "strong." Castilian, featuring no inflected infinitive at any stage of its existence, is considered

11. Uriagereka (1995ab) lists inflected infinitives as one property of "strong-morphology" languages. Other such properties are enclisis to finite verbs, interpolation, and bare affirmation of imperatives. The presence or absence of one or more of these phenomena is an indicator of the relative morphological strength of a language (Graham 2021).

morphologically "weaker" than the Western Iberian group. During the *Reconquista*, the Kingdom of Asturias was annexed into the neighboring Kingdom of León and would remain under their control until the merger with the Kingdom of Castile in the 14th century. Castile, being the most powerful kingdom in Spain during that era, imposed Castilian as the official language of government, education, and business throughout its reign, promulgated by efforts by King Alfonso X. In effect, the linguistic situation in Asturias became one of diglossia, with the superimposed Castilian holding prestige and Asturian relegated to a lower status.

With the expansion of Castilian into Asturian territory as a leveled dialect that gained prestige alongside official status, it is probable that the koine dialect displaced certain properties of the now-minority language, including the morphological strength which, when weakened, held several other ramifications for the syntax of the language in addition to the presence of personal infinitives (Graham 2021). Indeed, Lorenzo González (2016) describes the distribution of personal infinitives in Old Asturian as limited to predicate adjuncts, prepositional phrases, and nominal or adjectival complements/arguments. As simplification is one of the regularly observed processes or phenomena associated with koineization, speakers at the time may have decided that overt subjects were redundant with infinitive agreement in Old Asturian. Remarkably, these contexts mirror those of Castilian personalized infinitives.

Konvalinka (1985: 50) further remarks that where there is linguistic assimilation in Asturian, there are two distinct directions: (1) toward Castilian norms or, less frequently, (2) toward the Central Asturian variant. The sociohistorical effect of this assimilation tendency was that, like Castilian, Asturian infinitives with expressed subjects were no longer personal and inflected but personalized and uninflected. Therefore, in assimilating to the socially prestigious form, the Asturian dialect drifted from the rest of the Western Iberian group with the loss of the inflected infinitive. This is not to say that Asturian personalized infinitives pattern exactly like Castilian, as there exist certain structures in Asturian that would be unacceptable in Castilian (see Viejo Fernández 2015 for examples). The divergence in acceptability between Asturian and Castilian is evidence of koineization regarding infinitives with subjects. Though the morphology now follows the Castilian model in that the infinitive shows no agreement with the subject, the infinitive complement is able to take a different subject than the main verb unlike standard Castilian but more closely like Portuguese and Galician.

From a sociolinguistic standpoint, modern Asturian has not undergone a resurgence in prestige to a level that has led to a restitution of verbal paradigms such as the personal infinitive in higher registers, as has been the case in Galician. In fact, Castilian remains the prestige language, and Asturian the minority language, in the diglossic situation of Asturias. This is likely an effect of lasting status of Asturian

as a "recognized language or dialect" in Spain (Konvalinka 1985: 32, 25), meaning specifically that the language is not *de jure* co-official with Castilian as is Galician. Thus, although Asturian is gaining in official recognition, as evidenced by the establishment of *L'Academia de la Llingua Asturiana* and the subsequent increase in use of the language in media, business, official capacities, and educational environments (d'Andrés 2002: 83–90), the continual pressure from Castilian militates against the restitution of a form long lost.

7. Conclusion

In this chapter I have outlined the history of infinitives co-occurring with argumental overt subjects, defining inflected infinitives as *personal* infinitives as found in Portuguese and Galician and defining uninflected infinitives as *personalized* infinitives as found in Castilian and modern Asturian. I have described the decline and resurgence of personal infinitives in Galician, noting how it diverged from Portuguese and how this paradigm has been resurrected as a marker of linguistic pride in literary and high-register contexts. I have, additionally, presented an alternate account for the disappearance of the personal infinitive from Asturian, independent of grammar or syntactic theory, based on centuries of contact, diglossia, and later koineization with Castilian in the region. The changes in Galician and Asturian regarding personal and personalized infinitives show how linguistic change can be motivated by sociohistorical factors – in this case, renewed pride in Galician linguistic identity, contrasted with the continuing lesser prestige of the Asturian language.

A possible direction for future research could focus on historical variation in Galician infinitives with subjects. Wireback (1994) and Harris (2013) discount the creative theory for the genesis of the personal infinitive in Portuguese, but no such claim was made of Galician by either author. As Kabatek (1996, 1997a, 1997b) notes that the Galician personal infinitive regressed in use as Castilian structures were favored, the differences in usage patterns between Galician and Portuguese as well as the presence of Galician personalized infinitives may point to a creative-theory genesis of the personal infinitive in these contexts. Such a study would depend on the availability of authentic texts from the relevant time period.

References

Alers, Hilton. 2009. Legitimación de sujeto en cláusulas infinitivas adjuntas. *Cuadernos de Lingüística/UPR Working Papers* 2(1). 5–11.

Alexiadou, Artemis & Elena Anagnostopoulou. 1998. Parametrizing AGR: Word order, V-movement and EPP-checking. *Natural Language and Linguistic Theory* 16. 491–539.

Beardsley, Wilfred A. 1921. *Infinitive constructions in Old Spanish*. New York: Columbia University Press.

Berger, Marcus. 2015. The subject position in Spanish nominalized infinitives. *Penn Working Papers in Linguistics* 21(1), article 2, https://repository.upenn.edu/pwpl/vol21/iss1/2.

Brea, Mercedes. 1985. Formación y características del gallego. *Revista de Filología Románica* 111. 121–130.

Brown, Esther & Javier Rivas. 2011. Subject-verb order in interrogatives: A quantitative analysis of Puerto Rican Spanish. *Spanish in Context* 8(1). 23–49. https://doi.org/10.1075/sic.8.1.02bro

Brown, Esther & Javier Rivas. 2019. Bringing purported black sheep into the fold: Galician inflected infinitives and Puerto Rican Spanish pre-verbal infinitival subject pronouns. *Languages* 4(2). 40. https://doi.org/10.3390/languages4020040

d'Andrés, Ramón. 2002. La llingua asturiana na sociedá. *Informe sobre la llingua asturiana*, 77–104. Oviedo: Academia de la Llinguia Asturiana.

de Mello, George. 1995. Preposición + sujeto + infinitivo: "Para yo hacerlo." *Hispania* 78(4). 825–836. https://doi.org/10.2307/345159

Fidalgo Enriquez, Francisco José. 2012. ¿Infinitivo personal en español?. In Ángel María Sainz García & Eduardo Tobar Delgado (eds.), *Actas del IV Congreso de la Enseñanza del Español en Portugal*, 60–79. Évora: Ministerio de Educación, Cultura y Deporte.

García Arias, Xosé Lluis. 2003. *Gramática histórica de la lengua asturiana: fonética, fonología e introducción a la morfosintaxis histórica*. Uviéu (Oviedo): Academia de la Llingua Asturiana.

Gorski, Edair. 2000. Variação no uso do infinitivo pessoal. *Organon–Estudos da língua falada* 14, 95–113. Porto Alegre, UFRGS–Instituto de Letras.

Graham, Lamar A. 2021. On clitic placement and gradience of strength of FP in Western Ibero-Romance. *Languages in Contrast* 21 (1). 1–27. https://doi.org/10.1075/lic.18001.gra

Haegeman, Liliane. 1994. *Introduction to Government and Binding Theory*. 2nd edn. Oxford: Blackwell.

Harris, Michael J. 2013. The origin of the Portuguese inflected infinitive through a corpus analysis. In Jennifer Cabrelli Amaral, Gillian Lord, Ana de Prada Pérez & Jessi Elana Aaron (eds.), *Selected proceedings of the 16th Hispanic Linguistics Symposium*, 303–311. Somerville, MA: Cascadilla Proceedings Project.

Kabatek, Johannes. 1996. A diferenciación diafásica do galego actual. In Rosario Álvarez Blanco, Francisco Fernández Rei, and Antón Santamarina (eds.), *A lingua galega: historia e actualidade. Actas do 1º congreso internacional, 16–20 do setembro de 1996, vol. II*, 379–388. Santiago de Compostela: Consello da Cultura Galega, Instituto da Lingua Galega.

Kabatek, Johannes. 1997a. Strengthening identity: Differentiation and change in contemporary Galician. In Jenny Cheshire & Dieter Stein (eds.), *Taming the vernacular: From dialect to written standard language*, 185–199. London/New York: Longman.

Kabatek, Johannes. 1997b. The koineization process of contemporary Galician as a field for the study of language change. In Beningno Fernández Salgado (ed.), *4th International Conference on Galician Studies Oxford*, 173–178. Oxford: Centre for Galician Studies.

Koike, Dale A. 1983. The inflected infinitive in colloquial Brazilian Portuguese as a marker of social dialects. In John J. Bergen & Garland D. Bills (eds.), *Spanish and Portuguese in social context*, 92–104. Washington, DC: Georgetown University Press.

Konvalinka, Nancy Anne. 1985. La situación sociolingüística de Asturias. *Lletres Asturianes* 16. 29–66.

Lipski, John M. 1991. In search of the Spanish personal infinitive. In Dieter Wanner & Douglas Kibbie (eds.), *New analyses in Romance linguistics: papers from the XVIII Linguistic Symposium on Romance Languages*, 201–220. Amsterdam: John Benjamins. https://doi.org/10.1075/cilt.69.17lip

Longa, Víctor Manuel. 1994. The Galician inflected infinitive and the theory of UG. *Catalan Working Papers in Linguistics* 4(1). 23–44.

Lorenzo González, Guillermo. 2016. ¿Por qué no en asturiano? El infinitivo flexionado como "apomorfia negativa" de la lengua asturiana. *Lletres Asturianes* 114. 39–58.

Martins, Ana Maria. 2001. On the origin of the Portuguese inflected infinitive: A new perspective on an enduring debate. In Laurel J. Brinton (ed.), *Historical Linguistics 1999. Selected papers from the 14th International Conference on Historical Linguistics, Vancouver, 9–13 August 1999*, 207–222. Amsterdam: John Benjamins. https://doi.org/10.1075/cilt.215.15mar

Merlan, Aurélia. 2007. Las variedades lingüísticas del noroeste peninsular: Convergencias y divergencias. *Lletres Asturianes* 96. 7–56.

Ordóñez, Francisco. 2018. Verbal complex formation and overt subjects in infinitivals in Spanish. In Ángel J. Gallego & Roger Martin (eds.), *Language, syntax, and the natural sciences*, 34–57. Cambridge: Cambridge University Press. https://doi.org/10.1017/9781316591529.003

Ortega Santos, Iván. 2002. Overt subjects of infinitives in Spanish: Abstract agreement and the EPP. *Selected proceedings of the 12th Annual Graduate and Professional Symposium on Hispanic and Luso-Brazilian Literature, Language and Culture*, 21–28. Tucson, AZ: University of Arizona.

Parafita Couto, María del Carmen. 2004. Agreement and focus in Galician inflected infinitives. *Kansas Working Papers in Linguistics* 27. 66–91.

Penas Ibáñez, María Azucena. 2014. Interferencia gramatical latina en el infinitivo flexionado iberorromance: hipótesis sintáctica. *RILCE: Revista de Filología Hispánica* 30(2). 525–558.

Pillinger, O. S. 1980. The accusative and infinitive in Latin: A refractory complement clause. *Journal of Linguistics* 16(1). 55–83. https://doi.org/10.1017/S0022226700006344

Pires, Acrisio. 2002. Cue-based change: Inflection and subjects in the history of Portuguese infinitives. In David W. Lightfoot (ed.), *Syntactic effects of morphological change*, 143–178. Oxford: University Press. https://doi.org/10.1093/acprof:oso/9780199250691.003.0008

Pöll, Bernhard. 2007. On the licensing of overt subjects in Spanish infinitival clauses. *Probus* 19. 93–120.

Purim, V. A. 2012. Alegro-me por você estar passando bem e espero que assim continue. De Reynaldo Purim para Artur Purim – 1925. *Colônia Leta do Rio Novo*, https://rionovo.wordpress.com/2012/10/12/alegro-me-por-voce-estar-passando-bem-e-espero-que-assim-continue-de-reynaldo-purim-para-artur-purim-1925/. (8 October, 2019.)

Raposo, Eduardo. 1987. Case theory and Infl-to-Comp: The inflected infinitive in European Portuguese. *Linguistic Inquiry* 18(1). 85–109.

Roseman, Sharon R. 1995. Falamos como falamos: Linguistic revitalization and the maintenance of local vernaculars in Galicia. *Journal of Linguistic Anthropology* 5(1). 3–32. https://doi.org/10.1525/jlin.1995.5.1.3

Sabatini, R. Nicholas. 1984. The Portuguese personal infinitive vs. its absence in Spanish. *Hispania* 67(2). 245–248. https://doi.org/10.2307/341733

Scida, Emily. 2004. *The inflected infinitive in Romance languages*. New York/London: Routledge. https://doi.org/10.4324/9780203488768

Sitaridou, Ioanna. 2004. On the emergence of personal infinitives in Spanish: A cue-based approach. Abstract of a talk given at the DIGS VIII, Yale University, June 24–27, 2004, http://bloch.ling.yale.edu/~digs8/digs8%20abstracts/Sitaridou%20abstract.pdf.

Sitaridou, Ioanna. 2007. Romance infinitives with subjects, subjunctive obviation and Control Theory. In Luis Eguren & Olga Fernández Soriano (eds.), *Coreference, modality, and focus: Studies on the syntax-semantics interface*, 191–219. Amsterdam: John Benjamins. https://doi.org/10.1075/la.111.10sit

Sitaridou, Ioanna. 2009. On the emergence of personal infinitives in the history of Spanish. *Diachronica* 26(1). 36–64. https://doi.org/10.1075/dia.26.1.02sit

Stowell, Tim. 1982. The tense of infinitives. *Linguistic Inquiry* 13(3). 561–570.

Thomas, Earl. 1969. *The syntax of spoken Brazilian Portuguese*. Nashville: Vanderbilt University Press.

Tuten, Donald N. 2001. Modeling koineization. In Laurel J. Brinton (ed.), *Historical linguistics 1999. Selected papers from the 14th International Conference on Historical Linguistics, Vancouver, 9–13 August 1999*, 325–336. Amsterdam: John Benjamins. https://doi.org/10.1075/cilt.215.22tut

Tuten, Donald N. 2003. *Koineization in medieval Spanish*. Berlin/New York: Mouton de Gruyter. https://doi.org/10.1515/9783110901269

Tuten, Donald N. 2021. Complexification of the early modern Spanish address system: A role for koineization? In Whitney Chappell & Bridget Drinka (eds.), *Spanish socio-historical linguistics: Isolation and contact*, 17-47. Amsterdam/Philadelphia: John Benjamins. https://doi.org/10.1075/ahs.12.co2tut

Uriagereka, Juan. 1995a. An F position in Western Romance. In Katalin É. Kiss (ed.), *Discourse configurational languages*, 153–175. Oxford: Oxford University Press.

Uriagereka, Juan. 1995b. Aspects of the syntax of clitic positions in Western Romance. *Linguistic Inquiry* 26(Winter). 79–123.

Vázquez Diéguez, Ignacio. 2012. O infinitivo (conxugado) galego e portugués e as súas correspondencias en español. *Estudos de lingüística galega* 4. 107–130.

Viejo Fernández, Xulio. 2015. Some infinitive structures in Asturian. *Dialectología* 5. 5–28.

Wireback, Kenneth. 1994. The origin of the Portuguese inflected infinitive. *Hispania* 77(3). 544–554. https://doi.org/10.2307/344992

Language variation and change through an experimental lens

Contextual modulation in the use of the Progressive in three Spanish dialects

Martín Fuchs and María M. Piñango
Utrecht University / Yale University

Spanish exhibits two markers to convey a *progressive* meaning: the Simple Present and the Present Progressive. The use of these markers is contextually biased: the Simple Present requires contexts where speaker and addressee *share perceptual access* to the situation at issue, while the Present Progressive does not require such support. We test this generalization through real-time comprehension: the Simple Present marker in contexts *without* shared perceptual access should elicit slower reading times than *within* shared perceptual access contexts. A self-paced reading study (n = 176) in three different varieties of Spanish (Mexican, Rioplatense, and Castilian) bears this prediction out. Additionally, we find that the Mexican variety appears further advanced in the *Progressive-to-Imperfective* diachronic shift than its dialectal counterparts.

Keywords: Spanish, Progressive, dialectal variation, shared perceptual access, real-time methods

1. Introduction

In a communicative situation, a speaker holds two seemingly opposing construals: (1) there is overlap between the knowledge and beliefs that she shares with her addressee(s), and (2), this overlap is only partial. These opposing considerations create a tension that constrain the form of linguistic communication and the form of the speaker's message. They underlie the well-known interplay between linguistic *economy* and linguistic *expressivity* (Zipf 1949). The interplay works as follows: the assumption that the intended meaning can be inferred on the basis of what is shared knowledge between speaker and addressee pushes the speaker towards linguistic *economy*; the awareness that this shared knowledge is always incomplete, in turn,

https://doi.org/10.1075/ahs.12.c04pin

pushes the speaker to linguistically encode all of her intended meaning, leading to linguistic *expressivity*.

In Fuchs et al. (2020a), we argued for a cognitive grounding of these two forces. Specifically, we proposed that this *economy/expressivity* tension can be seen to emerge from two constructs respectively acting in tandem in the mind of the speaker: Common Ground (Stalnaker 1978, 2002) and Theory of Mind (e.g., Premack & Woodruff 1978; Wellman 1990; Gopnik 1993). Figure 1 below shows the elements involved and how they connect with one another in the context of a communicative situation viewed from the speaker's perspective; or, in model-theoretic terms, from the set of worlds compatible with the speaker's knowledge and beliefs, at a given time, in a given world of evaluation.

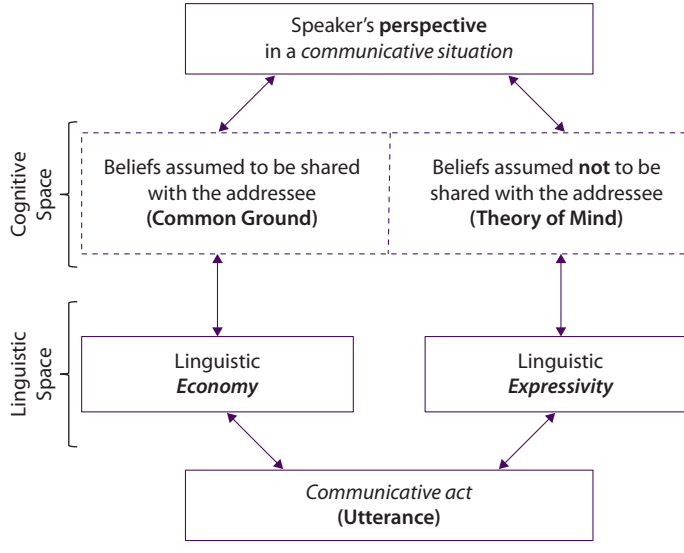


Figure 1. Elements and structure at play during a speaker's communicative act

Common Ground is the set of beliefs that the speaker brings to a communicative situation with a given addressee. This can be understood as the set of propositions that the speaker takes "for granted", which form the speaker's conversational background with the addressee before any specific communicative act has taken place (Stalnaker 1978: 321). By contrast, and in the context of a communicative situation, Theory of Mind refers to the speaker's understanding that the addressee's beliefs will not fully overlap with the speaker's own. Implicit in this understanding is the speaker's belief that the addressee might not be experiencing the context to the communicative situation in the same way as the speaker, and therefore the belief that the set of propositions that she (the speaker) attributes to the addressee before a communicative act is incomplete. The tension between these two forces in the

mind of the speaker – awareness of overlap and awareness of incompleteness – has communicative implications, leading to the opposition between linguistic *economy* and linguistic *expressivity*. Such opposition ultimately gets resolved in the speaker's communicative act. The need to resolve this tension indicates that within a given communicative situation speaker and addressee normally have distinct *perspectives*.

A measure of communicative success is whether or not the perspective of the addressee has been brought closer to that of the speaker's. This we take to be a key goal of communication: to bring to alignment the perspective of the addressee with that of the speaker's using language as a communicative device. We refer to this communicative goal as *perspective alignment* (Fuchs et al. 2020a). This process occurs every time that the addressee accepts the proposition put forth by the speaker, thus incorporating it into the Common Ground (Stalnaker 1978, 2002). A *linguistic communicative* ***situation*** is therefore a process whereby Common Ground between speaker and addressee grows as the perspective of the addressee is brought closer to that of the speaker through the addressee's gradual acceptance of the propositions presented to her by the speaker. A *linguistic communicative* ***act*** is the actual linguistic utterance by the speaker seeking to bridge the perspective gap. The choice of linguistic devices (in the *linguistic space*) can then be seen as the result of a calibration (in the *cognitive space*) between Common Ground (leading to linguistic *economy*) and Theory of Mind (leading to linguistic *expressivity*).

Here we investigate the psychological implications of this calibration between linguistic *economy* and linguistic *expressivity* during real-time sentence comprehension, and the connection between this linguistic synchronic process and larger patterns of diachronic change that are observable through semantic variation. To do this, we examine cross-dialectal variation with respect to the use of the Simple Present and the Present Progressive markers to convey an *event-in-progress* reading in three dialectal varieties of Spanish – Rioplatense Spanish, Mexican Altiplano Spanish, and Castilian Spanish – through a self-paced reading task. On the one hand, we analyze such dialectal variation as representing the synchronic manifestation of different diachronic substages in the *Progressive-to-Imperfective* shift, a crosslinguistically attested grammaticalization path in which progressive markers (such as the Present Progressive) appear in a language to express the *event-in-progress* reading, and slowly encroach into the more general domain of imperfective markers (such as the Simple Present). On the other hand, we propose that the source of the cross-dialectal variation is both the result of the sociohistorical development of these dialectal varieties and of the communicative and cognitive pressures at play in the diachronic path. In this way, we provide here an account of dialectal variation that relies both on a model of linguistic communication that is cognitively grounded, and on the historical sociolingustic background of these dialectal varieties.

The remainder of this article is organized as follows: Section 2 outlines our proposal with respect to how contextual conditions and linguistic markers interact in communicative situations to produce the necessary cognitive "fine-tuning" between Common Ground and Theory of Mind. Section 3 presents an account of the linguistic meanings of the Progressive and the Imperfective, the meanings involved in the *Progressive-to-Imperfective* shift. Section 4 looks at the diachronic and synchronic facts of Spanish, introduces the sociohistorical background of these varieties, and shows that these two proposed meanings are in fact two different *readings* of only one conceptual representation that is lexically accessed through different linguistic forms. Section 5 briefly describes the results of a previous acceptability judgment study in which we assess the role of *copresence* or *shared perceptual access* as a perceptual proxy of Common Ground. Sections 6 and 7 present the experimental study, Section 8 discusses the results in the context of diachrony and dialectal variation from a communicative and sociohistorical perspective, and Section 9 concludes the paper.

2. Communicative situations, context and linguistic markers

In language, markers and meanings are rarely in a one-to-one correspondence. Linguistic markers typically influence the sentential meaning in different ways depending on the context of utterance. That is, the meanings that are captured by lexical items usually demand contextual interaction in order to be appropriately composed with the other meanings in the utterance (e.g., Lewis 1980; Kaplan 1989). Consider the sentence in Spanish in (1):

(1) Ella bail-a.
She dance-PRS.3SG
'She dances/is dancing'

In (1), the verb *bailar* 'to dance' appears in the Spanish Simple Present form. In the translation, however, we see that this form can correspond to either the English Simple Present or the English Present Progressive. What determines its interpretation? The context. Consider the communicative situations in (2) and (3), where the sentence in (1) appears as uttered by a participant B as an answer to a question posed by a participant A:

(2) A: – ¿Qué pasatiempo tiene Ana?
B: – **Ella baila**.
A: – What is Ana's hobby?
B: – She dances.

(3) A: – ¿Qué hace Ana en su cuarto en este momento?
B: – **Ella baila**.
A: – What is Ana doing in her room right now?
B: – She is dancing.

In a context such as (2), the intended reading in Spanish is the equivalent to the English Simple Present (a *habitual* reading) while in a context such as (3), the most loyal translation is the English Present Progressive (an *event-in-progress* reading). As (2) and (3) illustrate, the Spanish Simple Present marker allows for both readings; its interpretation is determined in these cases by the nature of the question posed by A: in (2) *pasatiempo* 'hobby' suggests a *habitual* reading, whereas in (3) the addition of the temporal prepositional phrase *en este momento* 'right now' introduces a bias to an *event-in-progress* interpretation of the same linguistic marker. Alternatively, this latter reading can also be preferentially expressed in Spanish by the Present Progressive[1] marker, as in (4):

(4) Ana est-á baila-ndo.
Ana be-PRS.3SG dance-PROG
'Ana is dancing now'

As these cases illustrate, the interaction between the linguistic meanings and the non-linguistic context is manifested as the resolution of the tension between the meaning that is lexicalized and the meaning that is obtained from the contextual information in a given communicative situation. This resolution is indistinguishable from the resolution of the tension described above between Common Ground and Theory of Mind. What are the conditions under which speakers can be linguistically *economical*, rely on the situational context at issue, and use the ambiguous Simple Present form? When must they increase linguistic *expressivity* – by adding a temporal prepositional phrase or using a different construction such as the Present Progressive marker – in order to ensure communicative success?

---

1. A reviewer points out that the *estar* 'be' + *V-ndo* construction (i.e., the Present Progressive periphrasis) is in competition with other Auxiliary + *V-ndo* periphrastic constructions in Spanish (e.g., *ir* 'go' + *V-ndo*, *venir* 'come' + *V-ndo*, among others) in the expression of *progressive* aspect. In this study, we focus on the comparison between the Simple Present and the Present Progressive markers as (i) the *estar* 'be' + *V-ndo* construction is the more frequent *progressive* expression in all varieties of Spanish (Sedano 2000), and (ii) we consider that the other Auxiliary + *V-ndo* constructions contribute "non-aspectual nuances" (Torres Cacoullos 2000: 91) that make them not true alternants with the Simple Present marker in the expression of an *event-in-progress* reading. While there are studies that focus on the distribution of the different periphrastic constructions (e.g., Fafulas 2015), to our knowledge there is still no study that analyzes the potential competition between the Simple Present and these other periphrastic expressions.

We argue that speakers can rely on situational context whenever they can safely assume that the addressee already knows some specifics of the message that they will convey; that is, when there is some *shared knowledge* or *common ground* between them (Stalnaker 1978, 2002). During a communicative situation, this assumption is manifested as the speaker's awareness of *shared perceptual access* between them and the addressee (Fuchs et al. 2020b).[2]

Here we test the hypothesis that ***shared perceptual access*** is part of the calculus not only for the acceptability but also for the comprehension in real-time of the Spanish Simple Present marker when an *event-in-progress* reading is conveyed. If this hypothesis is true, this would indicate subjects' awareness during the comprehension process itself of the choice that a speaker is exercising when using a marker that necessitates reliance on context, or when using an alternative linguistic construction – the Present Progressive marker – that preferentially conveys the intended meaning. This would be in line with previous work on perspective-taking that shows that the interlocutors' awareness of each other's perspectives affects production and comprehension in real-time and it is not a *post hoc* assessment (e.g., Nadig & Sedivy 2002; Hanna & Tanenhaus 2004; Brown-Schmidt et al. 2008). We also hypothesize that it is the cognitive resolution of the tension between Common Ground and Theory of Mind that triggers the linguistic mechanisms at play in these semantic variation patterns and their underlying grammaticalization trajectories. Our test case is the Spanish Imperfective domain, which presents two markers – the Simple Present and the Present Progressive – for the expression of the different imperfective readings, markers that are also diachronically related in the *Progressive-to-Imperfective* grammaticalization path (e.g., Torres Cacoullos 2000). In what follows, we present the conceptual structure that brings those two markers together and that underlies the diachronic shift in which both currently participate.

3. The meanings of the Progressive and the Imperfective

To understand how two markers that are connected to a specific semantic domain are related diachronically, we first need to explain the relation between the meanings and readings that they express. Imperfective aspect is taken to denote a property of a situation that holds over some interval of time. This implies that a predicate under its scope presents the Subinterval Property (Bennett & Partee 1972): if a predicate *P* is true at some interval *I*, it follows that the predicate *P* is true at all (relevant) subintervals of *I*. Thus, the Imperfective domain is claimed to encompass two meanings: the *imperfective* and the *progressive* (e.g., Comrie 1976;

2. *Shared perceptual access* is a form of what Clark & Marshall (1981) discussed as *physical copresence* in their study on definite reference.

Deo 2009). The ***imperfective meaning*** is said to present three different readings: the *event-in-progress*, the *habitual*, and – with stative predicates – the *continuous* reading.[3] All these readings display the Subinterval Property. For example, consider the Spanish sentences with the Simple Present form in (5) and (6), which is considered to be an *imperfective* marker:

(5) Ana bail-a ahora.
Ana dance-PRS.3SG now
'Ana is dancing now'

(6) Ana bail-a dos veces por semana.
Ana dance-PRS.3SG two times a week
'Ana dances two times a week'

In (5), a sentence with an *event-in-progress* reading, the sentence radical (dance(Ana)) holds at every relevant subinterval of the reference interval. In the case of (6), a sentence with a *habitual* reading, the sentence radical holds at all relevant subintervals of the interval under consideration.

The ***progressive meaning***, on the other hand, is said to only present the *event-in-progress* reading. This is shown by (4), repeated below as (7), with the Present Progressive marker, which is considered to be a marker of the *progressive* meaning:

(7) Ana est-á baila-ndo.
Ana be-PRS.3SG dance-PROG
'Ana is dancing'

A unified account of these two meanings is found in Deo (2009, 2015), which provides a model-theoretic analysis for each meaning from which the different readings are derived. The account models the meaning of the *progressive* as a subset of the meaning of the *imperfective*. The *imperfective* and the *progressive* meanings are considered to be operators that apply to predicates of eventualities. These operators universally quantify over *regular partitions* of an interval; i.e., a set of collectively exhaustive, non-overlapping, equimeasured subsets of some set, against which the instantiation of a predicate is evaluated with respect to its distribution over time. This modeling captures the *regularity* in the distribution over time that is observed in sentences with imperfective aspect. Crucially for this analysis, the measure of the regular partition, which determines the value of each cell of the partition, is a free variable whose value is obtained from context. What is important to grasp is the fact that the different readings that arise in different utterances are the result of different values assigned to the regular partition measure.

---

3. For the purposes of this paper, we leave aside from analysis the *continuous* reading of the Imperfective domain, such as in *Ana vive en Tokio*, 'Ana lives in Tokyo'.

Under this account, the *progressive* and the *imperfective* operators are differentiated by their respective domains of quantification. The *progressive*'s domain of quantification is a regular partition of the reference interval, so that the predicate under its scope stands in a coincidence relation with regular subintervals of the reference interval. The *imperfective*'s domain of quantification, in turn, is a regular partition of a *superinterval* of the reference interval, so that the predicate under its scope stands in a coincidence relation with regular subintervals of a superinterval of the reference interval. The reference interval is always a subinterval of a superinterval thereof, which makes the *progressive* meaning a subset of the *imperfective* meaning.[4]

We have proposed (Fuchs et al. 2020a) that Deo's (2009, 2015) operationalized meanings are in fact two angles – reflected by each of the operators – of one underlying conceptual structure with two variables: the interval at issue and the regular partition measure. Which value is given to which variable determines the linguistic reading: *event-in-progress* or *habitual*. Part and parcel of this unification is the idea that the imperfective domain is itself not linguistic in nature, but part of a larger nonlinguistic cognitive structure to which lexicalization processes have access. This vastly simplifies the meaning–reading distribution and provides cognitive grounding for it: one conceptual structure that is accessed by different linguistic markers – Simple Present and Present Progressive – encoded as distinct lexical items, and whose interaction with context saturates the variables that give rise to the specific readings observed. From this it follows that any (dynamic) lexical item[5] that connects to this unified conceptual structure is ultimately able to convey either reading and, moreover, that the Present Progressive marker does not map to the *progressive* operator alone (to the exclusion of the *imperfective* one).

---

4. The formal representations of the operators, taken from Deo (2015), are given in (8) and (9):

(8) PROG: $\lambda P.\lambda i.\lambda w.\forall j[j \in \mathcal{R}^{c}_{i} \rightarrow \text{COIN}\ (P, j, w)]$

(9) IMPF: $\lambda P.\lambda i.\lambda w.\exists j[i \subseteq_{\text{ini}} j \wedge \forall k[k \in \mathcal{R}^{c}_{j} \rightarrow \text{COIN}\ (P, k, w)]]$

The *progressive* operator combines with a predicate of eventualities *P*, an interval *i*, and a world of evaluation *w*, and returns the proposition that every cell *j* of a regular partition of *i* **coincides** with *P* in *w*. The *imperfective* operator combines with a predicate of eventualities *P*, an interval *i*, and a world of evaluation *w*, and returns the proposition that there is some interval *j* that continues *i* (i.e., is a non-final superinterval of *i*) such that every cell *k* of a regular partition of *j* **coincides** with *P* in *w*. The coincidence relation is defined including *inertia alternatives* of the world of evaluation *w* to avoid the Imperfective Paradox (see e.g., Dowty 1977).

5. An anonymous reviewer rightfully asks what the role of lexical aspect is in this process. We argue that, while this conceptual structure is available to any lexical item, the *event-in-progress* and *habitual* readings are preferably obtained with dynamic predicates. Stative predicates, by contrast, demand that the predicate holds *both* at the reference interval and at the superinterval of the reference interval, achieving the *continuous* reading.

Figures 2 and 3 provide conceptual structure representations of the meanings proposed in Deo (2009). Figure 2 shows a predicate *P* under the scope of the *progressive* operator, which maps the predicate to regular partitions *j* of a reference interval *i* in a world of evaluation *w*, coinciding with them, thus giving rise to the *event-in-progress* reading:

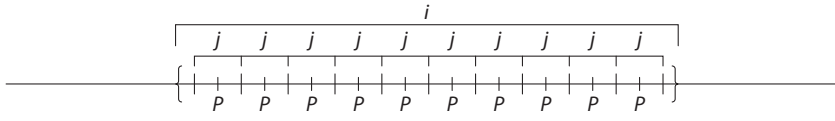


Figure 2. A representation of the Progressive meaning (with an *event-in-progress* reading)

Figure 3 shows a predicate *P* under the scope of the *imperfective* operator, which maps the predicate to regular partitions *k* of a superinterval *j* of the reference interval *i* in a world of evaluation *w*, coinciding with them, thus giving rise to the *habitual* reading:

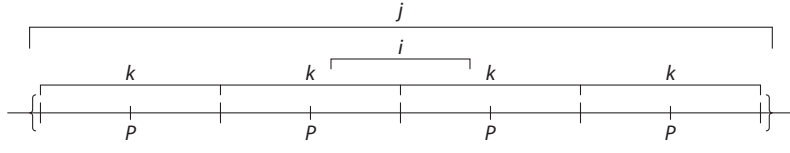


Figure 3. A representation of the Imperfective meaning (with a *habitual* reading)

These illustrations show how these two readings – the *event-in-progress* and the *habitual* – can actually arise from the same conceptual meaning structure; a structure that simply indicates that a predicate of eventualities coincides with every cell of a regular partition of an interval. The different readings result from the angle that is targeted within the same conceptual structure: while the *habitual* reading has a "wide angle" targeting both intervals of the conceptual representation, the *event-in-progress* reading has a "narrow angle" targeting instead the reference interval alone. Figure 4 shows a representation of the unified conceptual structure.

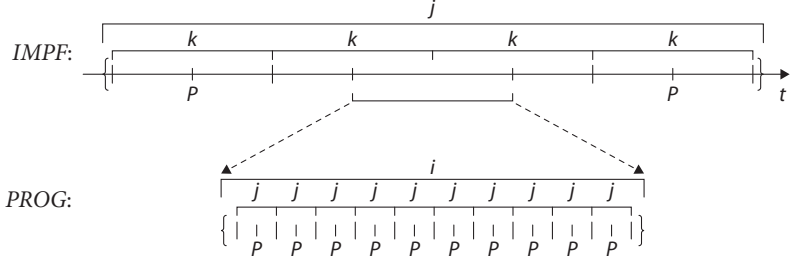


Figure 4. The underlying unified conceptual structure for the Imperfective domain

4. Spanish diachronic and synchronic facts

Markers within the Imperfective domain are related diachronically in what is known as the *Progressive-to-Imperfective* grammaticalization path (e.g., Bybee et al. 1994; Deo 2015). This diachronic shift is usually schematized as following a series of subsequent stages: (1) languages evolve from the availability of only one marker expressing both the *habitual* and the *event-in-progress* readings to recruiting a new marker to optionally express the *event-in-progress* reading (**emergence stage**); (2) the markers become restricted to mutually exclusive reading domains, such that the old marker only expresses the *habitual* reading and the new marker is the only device to convey the *event-in-progress* reading (**categoricalization stage**), and (3), the new marker gets reanalyzed as a general marker of imperfectivity and becomes able to convey both readings (**generalization stage**).

If our claim of a unified conceptual structure for the Imperfective domain is on the right track, we would expect any marker that targets any angle of this conceptual representation to be able to express both *event-in-progress* and *habitual* readings. And indeed, this is what we observe. Spanish used to have only one marker to express the *event-in-progress* and the *habitual* readings – the Simple Present marker – until the 13th century, when a previously locative construction (*estar* + location + gerund) was recruited for the expression of the *event-in-progress* reading, losing over time the need for an intervening location, and becoming the current Present Progressive marker.[6] However, since its appearance in the language, this marker has also been able to convey a *habitual* reading, as in (10) below, from the *Libro de Buen Amor* (14th century):

(10) Dixo el abutarda: "Loca, sandia, vana, sienpre **estás chirlando** locura, de mañana; non quiero tu consejo: ¡vete para villana!"
'And the great bustard said: 'Crazy, foolish, vain, you **are** always **talking** crazy in the mornings. I do not want your advice: go away, villain!'

Present-day Spanish uses these two markers to express the Imperfective domain in a two-by-two system. Both the Present Progressive marker and the Simple Present marker can express either the *event-in-progress* reading, indicating that an event is unfolding at reference time, as in (11), or the *habitual* reading, as in (12), showing that an event instantiates at regular intervals of a superinterval of the reference interval:

6. See Torres Cacoullos (2000, 2012) for a detailed explanation of the process of grammaticalization of this marker.

(11) a. Ana est-á baila-ndo (ahora).
Ana be-PRS.3SG dance-PROG (now)
'Ana is dancing now'
b. Ana bail-a ahora.
Ana dance-PRS.3SG now
'Ana is dancing now'

(12) a. Ana est-á baila-ndo todos los días.
Ana be-PRS.3SG dance-PROG all the days.
'Ana is dancing every day'
b. Ana bail-a todos los días.
Ana dance-PRS.3SG all the days
'Ana dances every day'

What the sentences in (11) and (12) show is that, as predicted by the one conceptual structure account, both markers can convey the *event-in-progress* and the *habitual* readings. The connection between these two readings is that they arise from the same conceptual meaning representation. However, Spanish speakers have preferences for the expression of these readings, suggesting a kind of *categoricalization* stage in the schematized grammaticalization path: speakers will mostly use the (newer) Present Progressive marker to express the *event-in-progress* reading, while they will use the (older) Simple Present marker to convey a *habitual* reading. Moreover, and from a variationist perspective, it could be expected that different dialectal varieties of Spanish partition the semantic space of Imperfectivity in similar, yet not identical ways. The differences in distributions across dialects would be the result of different sociohistorical trajectories. For instance, while the Mexican Altiplano dialect – triggered by the rapid development of a cultural metropolis in what is now Mexico City – quickly evolved in an independent way from the Castilian variety of the conqueror, the Rioplatense variety – under a slower colonization and population growth rate – has always been considered a more "archaic" dialect with a less independent trajectory from the Castilian variety.

Acting alongside these sociohistorical forces, there are specific contextual constraints that modulate this variation. In previous work (Fuchs et al. 2020b), we have shown that *copresence* or *shared perceptual access* plays a role in the acceptability of each of these markers to convey the *event-in-progress* reading, thus indicating the kinds of contextual factors that are involved in variable saturation for the imperfective conceptual structure. Moreover, our studies assessed the use of these markers in different Spanish varieties to show the diachronic shift at work in different 'synchronic cuts': different dialects may be at different substages within the same trajectory, just as different languages have been shown to be at different stages along the same grammaticalization path (e.g., Bybee et al. 1994). In what follows we present and discuss this previous work.

5. A questionnaire study on the *event-in-progress* reading

Fuchs et al. (2020b) present an acceptability judgment study showing that in Spanish the Present Progressive marker and the Simple Present marker are not in free variation to express an *event-in-progress* reading, opposing traditional accounts of their distribution (e.g., Marchand 1955; de Bruyne 1995; Bertinetto 2000). Conversely, Fuchs et al. (2020b) show that for Rioplatense Spanish and Castilian Spanish the distribution of these markers is governed by contextual constraints related to *copresence*: when speaker and addressee *share perceptual access* to the event described by the predicate, participants judge that a sentence with Simple Present marking that expresses an *event-in-progress* reading is more acceptable than when speaker and addressee do not share perceptual access to the event described by the predicate.[7] The use of the Present Progressive marker for the expression of this reading, on the other hand, does not show this contextual modulation, pointing to the preference of speakers for this marker for the *event-in-progress* reading. In Mexican Altiplano Spanish, in turn, Fuchs et al. (2020b) show that the Simple Present marker is no longer acceptable for the expression of the *event-in-progress* reading, regardless of whether or not *copresence* between speaker and addressee to the event described by the predicate is independently provided by the *shared perceptual access* manifested in the context. This last result is in line with analyses presented in Moreno de Alba (1978), who examines educated Mexico City oral data and shows that the Simple Present marker is rarely used to convey an *event-in-progress* reading.

Fuchs et al. (2020a) interpret these results as follows. They claim that the expression of an *event-in-progress* reading necessarily achieves the communicative goal of *perspective alignment* – the bringing of the addressee's perspective closer to that of the speaker – afforded by its reference time interpretation, which by definition points to the 'here' and 'now' and is therefore specific to the communicative situation involving speaker and addressee. Under this account, *perspective alignment* can thus be achieved either linguistically (by use of the Present Progressive marker) or non-linguistically (by *copresence* between speaker and addressee and the use of the Simple Present marker). As for the Mexican Altiplano Spanish variety,

7. Several studies point out the relevance of other factors, such as the lexical aspect of the predicate or the co-occurrence of certain adverbs, to account for the variation between Simple Present and Present Progressive marking (e.g., Torres Cacoullos 2009, 2012; Fafulas & Díaz Campos 2010, *i.a.*). In Fuchs et al. (2020b) and here, we abstracted away from these factors by (a) only considering accomplishments and achievements (evenly distributed across our stimuli), which favor the alternation between Present Progressive and Simple Present marking (Fafulas & Díaz-Campos 2010: 84) and (b) by not using accompanying adverbs in our sentences, which could have favored or made salient a specific aspectual interpretation. We argue that the remaining observable effects are thus attributable to the independent variables at play in the reported experiments.

Fuchs et al. (2020a) take this dialect's restriction to only use the Present Progressive marker to achieve *perspective alignment* to indicate that the dialect is further along the *Progressive-to-Imperfective* grammaticalization path.

In light of this pattern of results we ask: can we observe the same constraints at play in the real-time interpretation of these markers, as comprehension unfolds? To address this question, we developed a self-paced reading task that assesses whether, in the expression of the *event-in-progress* reading, the use of the Simple Present marker exerts an extra cost (compared to the Present Progressive marker) when *perspective alignment* is not independently satisfied by context. Our predictions are: (a) that the Simple Present marker will produce slower reading times when not preceded by a context that satisfies *perspective alignment*, (b) that the Present Progressive marker will *not* show such contextual modulation, and (c) that if variation is again observed across dialects, it should be in the direction predicted by the grammaticalization path – that is, the Simple Present marker will produce longer reading times than the Present Progressive marker regardless of contextual information. We present the details of the study directly below.

6. Real-time interpretation of an *event-in-progress* reading: An SPR study

To analyze the real-time processing of these markers, we developed a script with a total of 144 experimental items, consisting of 24 sextuples of items. The independent variables in the study are Context Type and Grammatical Marker, organized in a in a 2-by-3 design (two contexts, three markers). Items within each sextuple are identical aside from those two factors. Each experimental item consists of a vignette with two parts: a context – which presents a situation that involves at least a speaker and an addressee – and a sentence, which is presented as uttered by the speaker in that context. All vignettes focus on an ongoing situation, so that the sentence uttered by the speaker is always a declarative affirmative sentence that conveys an *event-in-progress* reading. The contexts are either Rich or Poor; that is, they either guarantee *shared perceptual access* between speaker and addressee to the event described by the predicate in the test sentence (Rich context) or they do not guarantee *shared perceptual access* between speaker and hearer (Poor context). The test sentences, in turn, consist of predicates with one of three possible markers: the Present Progressive, the Simple Present, or the Pretérito (Simple Past) marker. The latter represents a baseline condition, given its incompatibility with an *event-in-progress* reading. Test sentences are uniformly distributed for Grammatical Person, and they all have Singular number. The study also includes 180 fillers from an unrelated task, which results in a final script of 324 items. Table 1 provides examples of a context-sentence pair in each of the six conditions.

Table 1. Experimental Stimuli (examples). *Notes*. Prog., Present Progressive; Pres., Simple Present; Prét., Simple Past. Bold shows minimal differences between conditions

Context	Marker	Example
Poor	Prog	Ana llega a su casa de trabajar y va a buscar a su hijo a su habitación para ver cómo está. Golpea la puerta, **pero el hijo no contesta. Sin que ella llegue a abrir la puerta**, su hijo le dice: **Estoy haciendo** la tarea de Matemática. 'Anna gets home from work and goes to her son's room to see how he is doing. She knocks on the door **but her son does not answer. Before she gets to open the door**, her son tells her: I **am doing** Math homework'
	Pres	Ana llega a su casa de trabajar y va a buscar a su hijo a su habitación para ver cómo está. Golpea la puerta, **pero el hijo no contesta. Sin que ella llegue a abrir la puerta**, su hijo le dice: **Hago** la tarea de Matemática. 'Anna gets home from work and goes to her son's room to see how he is doing. She knocks on the door **but her son does not answer. Before she gets to open the door**, her son tells her: I **do** Math homework'
	Prét	Ana llega a su casa de trabajar y va a buscar a su hijo a su habitación para ver cómo está. Golpea la puerta, **pero el hijo no contesta. Sin que ella llegue a abrir la puerta**, su hijo le dice: **Hice** la tarea de Matemática 'Anna gets home from work and goes to her son's room to see how he is doing. She knocks on the door **but her son does not answer. Before she gets to open the door**, her son tells her: I **did** Math homework'
Rich	Prog	Ana llega a su casa de trabajar y va a buscar a su hijo a su habitación para ver cómo está. Golpea la puerta, **la abre, y ve al hijo sentado en el escritorio. Antes de que ella pueda decir algo**, su hijo le dice: **Estoy haciendo** la tarea de Matemática. 'Anna gets home from work and goes to her son's room to see how he is doing. She knocks on the door, **opens it, and sees her son sitting at his desk. Before she can say anything**, her son tells her: I **am doing** Math homework'
	Pres	Ana llega a su casa de trabajar y va a buscar a su hijo a su habitación para ver cómo está. Golpea la puerta, **la abre, y ve al hijo sentado en el escritorio. Antes de que ella pueda decir algo**, su hijo le dice: **Hago** la tarea de Matemática. 'Anna gets home from work and goes to her son's room to see how he is doing. She knocks on the door, **opens it, and sees her son sitting at his desk. Before she can say anything**, her son tells her: I **do** Math homework'
	Prét	Ana llega a su casa de trabajar y va a buscar a su hijo a su habitación para ver cómo está. Golpea la puerta, **la abre, y ve al hijo sentado en el escritorio. Antes de que ella pueda decir algo**, su hijo le dice: **Hice** la tarea de Matemática. 'Anna gets home from work and goes to her son's room to see how he is doing. She knocks on the door, **opens it, and sees her son sitting at his desk. Before she can say anything**, her son tells her: I **did** Math homework'

We tested participants in the three different dialectal varieties of Spanish under consideration: Rioplatense Spanish, Castilian Spanish, and Mexican Altiplano Spanish (n = 176).[8] Each participant in the study saw a unique self-paced reading script with the 324 items randomly ordered. Stimuli were presented using E-Prime software, following a standard noncumulative moving-window self-paced reading paradigm.[9] For each vignette (context-sentence pair), the participants first saw a series of dashes, each representing a letter in the words of the vignette. Participants then had to press the space bar to see the first word of the context-sentence pair. After the second press on the spacebar, the second word was displayed, and the first word went back to being represented by a series of dashes. Every spacebar press displayed the following word and replaced the previous word with dashes until the context-sentence pair was read completely. Reading times were recorded for each displayed word, and were compared across the six experimental conditions for five regions of interest: the critical word (viz., the grammatical marker) and the surrounding words (one word before and three words after it).[10]

8. Sixty participants from Buenos Aires, Argentina constitute the sample for Rioplatense Spanish (30 female, age range 19–34 years old, M_{age} = 27 years old). Sixty participants from Madrid, Spain represent the sample for Castilian Spanish (30 female, age range 20–38 years old, M_{age} = 28 years old). These participants were not bilingual speakers of any other of the official languages in Spain, such as Basque or Catalan. Finally, fifty-six participants were from Mexico City and its surroundings states (Mexico, Morelos, Tlaxcala, Hidalgo, and Puebla), composing the Mexican Altiplano Spanish sample (33 female, age range 18–37 years old, M_{age} = 27.5 years old). All participants had completed at least 12 years of formal education, by self-report had no history of neurological disease or brain injury, and had normal to corrected-to-normal vision. They all provided written informed consent in accordance with the guidelines set by the Yale University Human Subjects Committee and were compensated with $20 for their participation.

9. The script was divided in four parts, with 5 to 10 minute breaks between these parts. The whole experiment lasted approximately 2 hours, including breaks. Before the experiment started, participants were presented with six practice items to ensure that they understood the task and were familiar with the experimental paradigm. To confirm that participants were paying attention to the task, 75% of the sentences were followed by a comprehension question that could be answered with either 'yes' or 'no'. After the end of each sentence, questions were presented in their entirety on the screen and remained there until the participant chose an answer. To prevent systematic biases, half of the questions had an expected 'yes' answer, and half of the questions had an expected 'no' answer. All of the questions required comprehension of the full context-sentence pair to be answered correctly. For the sentences that did not present a comprehension question, a screen appeared that required the participant to press 'yes' or 'no' to continue with the experiment.

10. To account for word-length differences between grammatical markers, reading times were residualized; that is, a mean reading time was calculated for each region of interest, and then the reading times were centered by subtracting this mean for all values for the corresponding region (e.g., Ferreira & Clifton 1986). Given that the Present Progressive marker includes two words while the Simple Present and the Simple Past are realized by just one, reading times were also residualized by number of words. All statistical analyses were performed on this last set of residualized reading times.

7. Results

7.1 Behavioral results

The responses to the comprehension questions were analyzed, and a cutoff was set at 85% of accuracy. Under this restriction, two subjects were excluded from the Castilian Spanish sample, and only one subject was excluded from the Mexican Altiplano Spanish sample. All Rioplatense Spanish participants answered more than 85% of the comprehension questions correctly.[11]

7.2 Reading time results

Word-by-word residualized reading times were analyzed using linear mixed effects models.[12] For Castilian Spanish, one word after the grammatical marker, we find a significant main effect of grammatical marker ($\chi^2(2) = 17.681$, $p < .005$) and a significant interaction between context type and grammatical marker ($\chi^2(5) = 8.777$, $p < .05$). Post hoc tests show that the main effect of grammatical marker is explained by significantly longer reading times for the Simple Past marker – the baseline condition – over both the Simple Present ($\beta = 31.048$, $p < .001$) and the Present Progressive markers ($\beta = 38.317$, $p < .001$). The interaction between grammatical marker and context type was broken down by conducting separate analyses to assess the effect of context type on each of the grammatical markers. The interaction effect is due to sentences that display the Simple Present marker: they were read significantly slower when preceded by a Poor Context than when preceded by a Rich Context ($\chi^2(1) = 8.254$, $\beta = 20.597$, $p < .005$).[13] Figure 5 shows the residualized

11. Taking out the excluded subjects, the correct answers mean for Castilian Spanish ($n = 58$) was 95.1%, for Mexican Altiplano Spanish ($n = 55$) was 95.9%, and for Rioplatense Spanish ($n = 60$) was 97.9%.

12. All analyses were peformed in R (R Core Team 2018) with the *lme4* package (Bates et al. 2015). Separate analyses were performed for each dialectal variety. For model selection and to obtain p-values, we followed regular recommendations for linguistic analysis (Winter 2013) and performed Likelihood Ratio Tests of the full model with the effect(s) in question against a model without them. Models included context type, grammatical marker, and the interaction between them as fixed effects, while as for random effects, they included random intercepts for items and participants and by-participant random slopes for context type and for grammatical marker. Post hoc tests were run with the *multcomp* package (Hothorn et al. 2008), and p-values were corrected by Tukey. No deviations from homoskedasticity or normality were observed through visual inspection of the results.

13. No other significant interactions are found at any of the other segments under scrutiny (PreVerb: $\chi^2(5) = 3.342$, $p = .647$; Verb: $\chi^2(5) = 4.633$, $p = .462$; Verb + 2: $\chi^2(5) = 0.793$, $p = .978$; Verb + 3: $\chi^2(5) = 2.838$, $p = .725$).

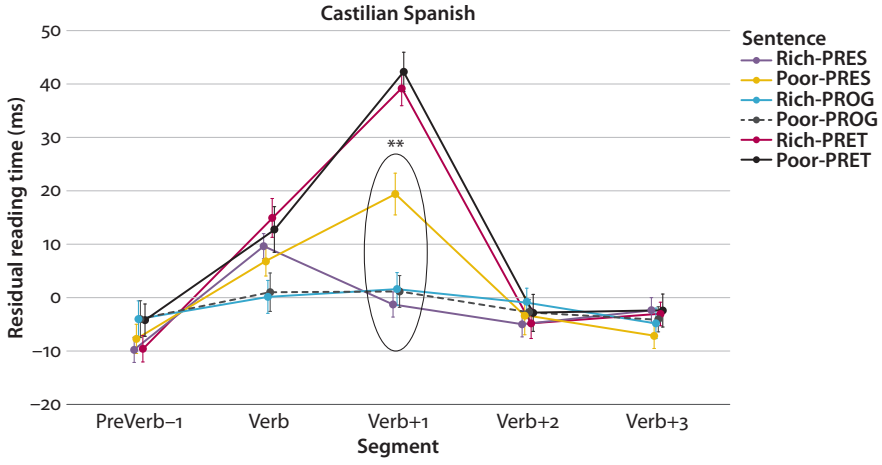


Figure 5. Word-number and letter-length corrected mean reading times for each context-grammatical marker pair at each relevant segment (Castilian Spanish)

reading times for each of the segments and the significant interactions that are found (* = $p < .05$, ** = $p < .01$; *** = $p < .001$ for all graphs).

A similar pattern is found for Rioplatense Spanish. One word after the grammatical marker, both a significant main effect of grammatical marker ($\chi^2(2) = 12.794$, $p < .005$) and a significant main effect of context type ($\chi^2(1) = 4.863$, $p < .05$) are revealed. We also find a significant interaction between context type and grammatical marker ($\chi^2(5) = 19.9$, $p < .005$). Post hoc tests show that once again the main effect of grammatical marker is explained by significantly longer reading times for the Simple Past marker over both the Simple Present ($\beta = 24.039$, $p < .01$) and the Present Progressive markers ($\beta = 35.469$, $p < .001$). The main effect of context type is due to significantly longer reading times in Poor Contexts than in Rich Contexts ($\beta = 13.983$, $p < .05$). However, given the significant interaction between context type and grammatical marker, separate analyses were conducted for each grammatical marker.

Under this assessment, only sentences that display the Simple Present marker were read significantly slower when preceded by a Poor Context than when preceded by a Rich Context ($\chi^2(1) = 5.961$, $\beta = 27.25$, $p < .05$), while no such effect is found either for the Present Progressive marker ($\chi^2(1) = 0.446$, $p = .504$) or the Simple Past marker ($\chi^2(1) = 0.431$, $p = .512$). In Rioplatense Spanish, the main effect of grammatical marker also persists two words after the verb ($\chi^2(2) = 27.756$, $p < .001$). Post hoc tests show that this effect is again explained by significantly longer reading times for the Simple Past marker over both the Simple Present ($\beta = 31.356$, $p < .001$) and the Present Progressive markers ($\beta = 34.889$,

$p < .001$).[14] Figure 6 shows the residualized reading times for each of the segments and the significant interactions that are found.

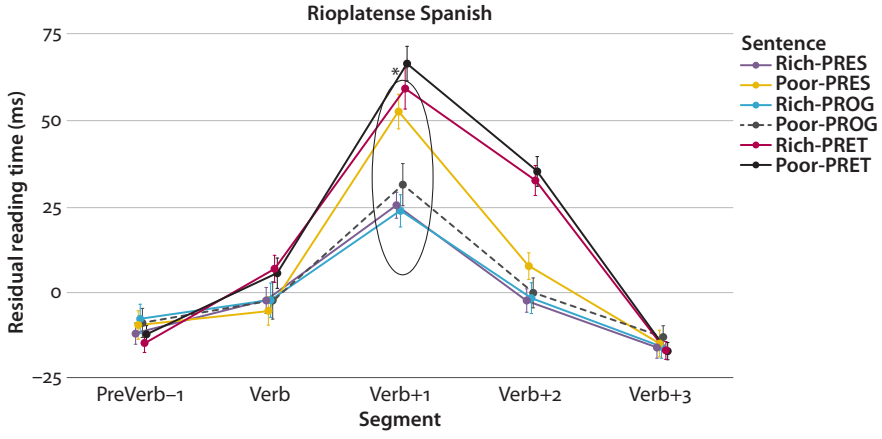


Figure 6. Word-number and letter-length corrected mean reading times for each context-grammatical marker pair at each relevant segment (Rioplatense Spanish)

The Mexican Altiplano Spanish data shows a different pattern. No significant interaction effect of the grammatical marker and the context type is found one word after the verb ($\chi^2(5) = 1.493$, $p = .684$). At that segment, we find only a significant main effect of grammatical marker ($\chi^2(2) = 7.697$, $p < .05$). Post hoc tests show that this effect is explained by longer reading times for the Simple Past marker over both the Simple Present ($\beta = 22.395$, $p < .05$) and the Present Progressive markers ($\beta = 24.901$, $p < .05$). However, different from the other dialectal varieties, we also find a significant main effect of grammatical marker at the verb position ($\chi^2(2) = 8.663$, $p < .05$). In this case, post hoc tests with Tukey correction show that the effect is explained by significantly longer reading times both for the Simple Present ($\beta = 23.914$, $p < .05$) and the Simple Past ($\beta = 24.52$, $p < .05$) over the Present Progressive marker. Moreover, there is no significant difference in reading times when comparing the Simple Present and the Simple Past regardless of contextual information ($\beta = 0.587$, $p = .998$).[15] Figure 7 shows the residualized reading times for each of the segments and the significant effects that are found.

---

14. No other significant interactions are found at any of the other segments under scrutiny (PreVerb: $\chi^2(5) = 2.583$, $p = .771$; Verb: $\chi^2(5) = 3.135$, $p = .679$; Verb + 3: $\chi^2(5) = 0.904$, $p = .97$).

15. No other significant interactions are found at any of the other segments under scrutiny (PreVerb: $\chi^2(5) = 1.238$, $p = .941$; Verb: $\chi^2(5) = 0.392$, $p = .942$; Verb + 2: $\chi^2(5) = 1.733$, $p = .885$; Verb + 3: $\chi^2(5) = 1.01$, $p = .962$).

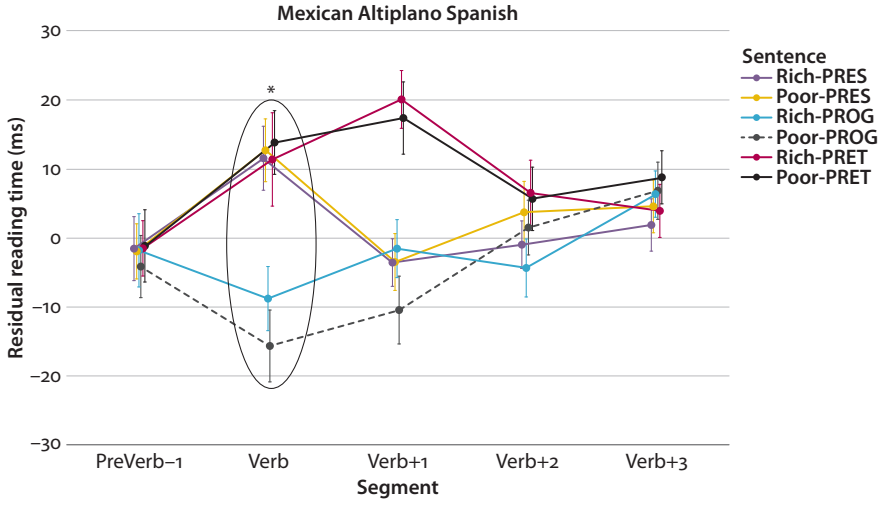


Figure 7. Word-number and letter-length corrected mean reading times for each context-grammatical marker pair at each relevant segment (Mexican Altiplano Spanish)

8. Discussion

Overall, the results from the self-paced reading task align with our predictions (see §5): (a) reading times for the Simple Present marker condition are slower when preceded by a Poor Context than when preceded by a Rich Context in Rioplatense Spanish and in Castilian Spanish, (b) no such contextual modulation is observed in any dialect for the Present Progressive marker condition, and (c) the observed variation across dialects is in accordance with the grammaticalization path – viz., in Mexican Altiplano Spanish, the reading times for the Simple Present marker condition are slower than the reading times for the Present Progressive marker condition independent of contextual bias.

Across dialects, we observe that the use of the Simple Present marker to convey an *event-in-progress* reading exerts greater cost in comparison to the Present Progressive marker. This cost is eliminated in Rioplatense Spanish and Castilian Spanish when *shared perceptual access* between speaker and addressee to the event described by the predicate is independently provided by the preceding contextual information, facilitating the reading of the Simple Present marker and its incorporation into the meaning representation of the event. We argue that *shared perceptual access* is a nonlinguistic means to increase Common Ground between speaker and addressee: it allows speakers of these dialectal varieties to take for granted that their addressees know their intended *event-in-progress* reading. This lets them use

the Simple Present marker and exert linguistic *economy*. When this contextual condition is not satisfied, Theory of Mind – i.e., their awareness of their addressee's different knowledge and belief state with respect to the event at issue – pushes the speaker towards the use of the more *expressive* construction, the Present Progressive marker, to convey the *event-in-progress* reading. This is what the reading times reflect: when participants are faced with a vignette in which a speaker is presented as having used the Simple Present marker without contextual support, they take a slightly longer time to read the utterance than when the vignette displays the Present Progressive marker, since in the former case they need to mentally construe a context that can accommodate the use of the Simple Present marker.

In Mexican Altiplano Spanish, the presence of *shared perceptual access* between speaker and addressee in the preceding context does not produce the facilitation effect in reading times. The extra cost of processing the Simple Present marker is observed regardless of the kind of contextual information. This last pattern in Mexican Altiplano Spanish suggests that *shared perceptual access* may no longer be playing a role in increasing Common Ground between speaker and addressee for the comprehension of an *event-in-progress* reading of the Simple Present marker, and that the only available marker for this reading in this variety is the Present Progressive one. This indicates, in turn, that this dialect is further ahead in the diachronic path of *Progressive-to-Imperfective*, closer to a strict *categoricalization* stage. Interestingly, this interpretation is supported by other studies that have shown that Mexican Spanish uses the Present Progressive twice as much as, for instance, varieties spoken in Spain (Fafulas 2015).

We conclude this section with a brief discussion about some sociohistorical factors involved in the differential progress along the grammaticalization path that is observed across the dialectal varieties under study. Specifically, why would a particular variety of Spanish be further along the diachronic shift? What has made Mexican Altiplano Spanish advance "faster" than the other dialectal varieties? One possibility is the influence of U.S. English – a language closer to *categorical* domains of use for the Simple Present and the Present Progressive markers – given its geographical proximity to Mexico. However, this explanation is unlikely. Torres Cacoullos (2000: 15–17) argues that if contact with English were propelling the acceleration of the change in Mexican Spanish, futurate uses of the Present Progressive should be observed (as in English *I'm leaving tomorrow*). These futurate uses of the Present Progressive are not found in her corpus data of Mexican Spanish. Moreover, the variety under analysis in this paper, in the Mexican Altiplano area, is geographically more distant from the U.S. than other varieties that might undergo more language contact with U.S. English.

It is possible, however, that a different type of contact may have played a role, and that a sociohistorical explanation is in order. This proposal relies on how the Spanish conquest proceeded in the Americas and how different linguistic *norms* or *standards* were constituted across the continent and in Spain itself. The different dialectal varieties of Spanish emerged as the result of contact situations between speakers from different regions of the Iberian Peninsula, which came together only as a consequence of the process of conquest and colonization of the American territories (e.g., Fontanella de Weinberg 1992; Rosenblat 2002). Standardization of these newly established varieties occurred at different rates across different territories, especially since urban centers tend to standardize their varieties faster. Mexico City is an exceptional case in this regard given its fast urbanization and the independent cultural development that it acquired in only a few years: conquered by 1521, eight years later it already had a cathedral, and by 1553 it had its own university (Menéndez Pidal 1962).

By contrast, the Rio de la Plata region was a marginalized zone with respect to its development until the 18th century. This territory lacked the mineral richness of Mexico or Peru and was therefore not exploited commercially by Spain until other commercial agricultural enterprises were put in motion. This lack of economic importance, together with its geographical remoteness, made its demographic growth very slow until the second half of the 18th century, when Buenos Aires finally became a relevant commercial center and the Viceroyalty of the Río de la Plata was founded in 1776. This historical isolation from the centers of power, culture and education is reflected in a variety of Spanish that has traditionally been characterized as archaic: a variety with late standardization that retained many features associated with the so-called 'pure standard' of the Castilian metropolis (cf. Schreier 2009).

As for the Castilian variety, which included monolingual speakers from Madrid and the greater Madrid region, we hypothesize that the existence of a more literate, and therefore more conventionalized, culture – together with the norms imposed by the Real Academia Española, founded in 1713 – stalled the development of the diachronic shift. This hypothesis would be in line with our data: the Rioplatense variety and the Castilian variety appear more closely related and slower in their advancement in the grammaticalization path, while the Mexican Altiplano Spanish variety appears faster in the encroachment of the Present Progressive marker into the domain of the Simple Present. We leave for future work the exploration of this possibility with the rigor that such an investigation deserves.

9. Conclusion

We have shown that the previously observed variation in an acceptability judgment task between the Spanish Simple Present and the Present Progressive markers (Fuchs et al. 2020b) can also be seen unfolding through a self-paced reading task in real-time comprehension. This synchronic variation, together with the diachronic relation between the markers in the *Progressive-to-Imperfective* grammaticalization path, appears to be constrained by cognitive principles imposed by the speaker's assessment of what is Common Ground between her and her addressees and by her assumptions about the different belief states of her addressees, or the speaker's Theory of Mind. The speaker's choice of marker within the Imperfective domain for the expression of the *event-in-progress* reading seems to be determined by whether or not they share *perceptual access* with the addressee, a proxy for the expansion of Common Ground by non-linguistic means.

The observed pattern is also in line with our proposal that the *event-in-progress* reading is a means to obtain the general communicative goal of *perspective alignment* (Fuchs et al. 2020a). When the speaker uses the Present Progressive marker, *perspective alignment* is achieved by the sole use of the linguistic marker; in contrast, when they use the Simple Present marker, they need *shared perceptual access* to increase their Common Ground with the addressee in order to attain this goal. Bearing in mind the self-paced reading data, the overall preference for the Present Progressive marker over the Simple Present marker to obtain the *event-in-progress* reading – and therefore this communicative goal – is taken as a tactic to optimize communication through linguistic means: use of the more *expressive* marker demands less computational resources than the more *economical* but more context-dependent marker, since the use of the more *expressive* but context-independent marker saves the processor the cost of integrating linguistic and contextual information to achieve the *event-in-progress* interpretation. Herein lies the source of the preference observed.

This preference for the Present Progressive marker is exacerbated in the case of Mexican Altiplano Spanish, a dialectal variety that does not show context modulation effects, suggesting that for that domain *perspective alignment* can only be achieved by linguistic means. This analysis rests on the assumption that Mexican Altiplano Spanish showed these context effects at some previous point in its diachronic development and that their current absence only indicates the outcome of a competition that the Present Progressive marker won for the expression of the *event-in-progress* reading and the achievement of *perspective alignment*. Support for this assumption lies in the crosslinguistically well-attested *Progressive-to-Imperfective* grammaticalization path (e.g., Bybee et al. 1994; Deo 2015), and on the

the effects that the different sociohistorical contexts of each dialectal variety exert on the different rates at which this diachronic development obtains across them.

Altogether, the pattern observed across dialects is consistent with a model of variation embedded in a communicative system, visible and at play during real-time comprehension, and shown to be subject to identifiable contextual and sociohistorical factors. The communicative system uses linguistic markers to optimize Common Ground and Theory of Mind cognitive pressures on the speaker and, in doing so, supports each dialectal variety's independent advancement from one stage to the following in their own larger path of change, manifesting predictable patterns of synchronic semantic variation in the process.

References

Bates, Douglas, Martin Maechler & Ben Bolker. 2015. Fitting linear mixed-effects models using lme4. *Journal of Statistical Software* 67. 1–48.

Bennett, Michael & Barbara Partee. 1972. Toward the logic of tense and aspect in English. Santa Monica, California: System Development Corporation, ms.

Bertinetto, Pier Marco. 2000. The progressive in Romance, as compared with English. In Östen Dahl (ed.), *Tense and aspect in the languages of Europe*, 559–604. Berlin: de Gruyter.

Brown-Schmidt, Sarah, Christine Gunlogson & Michael K. Tanenhaus. 2008. Addressees distinguish shared from private information when interpreting questions during interactive conversation. *Cognition* 107(3). 1122–1134. https://doi.org/10.1016/j.cognition.2007.11.005

Bybee, Joan, Revere Perkins & William Pagliuca. 1994. *The evolution of grammar: Tense, aspect and modality in the languages of the world*. Chicago, IL: University of Chicago Press.

Clark, Herbert H. & Catherine R. Marshall. 1981. Definite reference and mutual knowledge. In Aravind K. Joshi, Bonnie L. Weber & Ivan A. Sag (eds.), *Elements of discourse understanding*, 10–63. Cambridge: Cambridge University Press.

Comrie, Bernard. 1976. *Aspect. An introduction to the study of verbal aspect and related problems*. Cambridge: Cambridge University Press.

De Bruyne, Jacques. 1995. *A comprehensive Spanish grammar*. Oxford: Blackwell.

Deo, Ashwini. 2009. Unifying the imperfective and the progressive: partitions as quantificational domains. *Linguistics and Philosophy* 32. 475–521.

Deo, Ashwini. 2015. The semantic and pragmatic underpinnings of grammaticalization paths: the progressive to imperfective shift. *Semantics and Pragmatics* 8. 1–52.

Dowty, David. 1977. Towards a semantic analysis of verb aspect and the English 'imperfective' Progressive. *Linguistics and Philosophy* 1. 45–77.

Fafulas, Stephen. 2015. Progressive constructions in native-speaker and adult-acquired Spanish. *Studies in Hispanic and Lusophone Linguistics* 8(1). 85–133. https://doi.org/10.1515/shll-2015-0004

Fafulas, Stephen & Manuel Díaz-Campos. 2010. Variación morfosintáctica y lenguas en contacto: las formas analíticas y sintéticas del presente progresivo en el español monolingüe y bilingüe. *Boletín de Filología* 45(2). 71–89. https://doi.org/10.4067/S0718-93032010000200004

Ferreira, Fernanda & Charles Clifton Jr. 1986. The independence of syntactic processing. *Journal of Memory and Language* 25(3). 348–368. https://doi.org/10.1016/0749-596X(86)90006-9

Fontanella de Weinberg, M. Beatriz. 1992. *El español en América*. Madrid: MAPFRE.

Fuchs, Martín, María Mercedes Piñango & Ashwini Deo. 2020a (in press). Operationalizing the role of context in language variation: the role of perspective alignment in the Spanish Imperfective domain. In Sebastian Löbner, Thomas Gamerschlag, Tobias Kalenscher, Markus Schrenk & Henk Zeevat (eds.), *Concepts, frames, and cascades in semantics, cognition, and ontology*, chapter 10 (Language, Cognition and Mind 7). New York: Springer.

Fuchs, Martín, Ashwini Deo & María Mercedes Piñango. 2020b. The progressive-to imperfective shift: Contextually determined variation in Rioplatense, Iberian, and Mexican Altiplano Spanish. In Alfonso Morales-Front, Michael J. Ferreira, Ronald P. Leow & Cristina Sanz (eds.), *Hispanic linguistics: Current issues and new directions*, 119–136. Amsterdam: Johns Benjamins. https://doi.org/10.1075/ihll.26.06fuc

Gopnik, Alison. 1993. How we know our minds: The illusion of first-person knowledge of intentionality. *Behavioral and Brain Sciences* 16. 1–15.

Hanna, Joy. E. & Michael K. Tanenhaus. 2004. Pragmatic effects on reference resolution in a collaborative task: Evidence from eye movements. *Cognitive Science* 28(1). 105–115. https://doi.org/10.1207/s15516709cog2801_5

Hothorn, Torsten, Frank Bretz & Peter Westfall. 2008. Simultaneous inference in General Parametric Models. *Biometrical Journal* 50. 346–363.

Kaplan, David. 1989. Demonstratives. In Joseph Almog, John Perry & Howard Wettstein (eds.), *Themes from Kaplan*, 481–563. New York: Oxford University Press.

Lewis, David. 1980. Index, context, and content. In Stig Kanger & Sven Öhman (eds.), *Philosophy and grammar*, 79–100. Dordrecht: D. Reidel Publishing Company. https://doi.org/10.1007/978-94-009-9012-8_6

Marchand, Hans. 1955. On a question of aspect: A comparison between the progressive form in English and that in Italian and Spanish. *Studia Linguistica* 9. 45–52.

Menéndez Pidal, Ramón. 1962. Sevilla frente a Madrid. Algunas precisiones sobre el español de América. *Miscelánea homenaje a André Martinet: Estructuralismo e Historia*, 99–165. La Laguna: Universidad de La Laguna.

Moreno de Alba, José G. 1978. *Valores de las formas verbales en el español de México*. México: UNAM.

Nadig, Aparna S. & Julie C. Sedivy. 2002. Evidence of perspective-taking constraints in children's on-line reference resolution. *Psychological Science* 13(4). 329–336. https://doi.org/10.1111/j.0956-7976.2002.00460.x

Premack, David & Guy Woodruff. 1978. Does the chimpanzee have a theory of mind? *Brain and Behavioral Sciences* 4. 515–526.

R Core Team. 2018. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

Rosenblat, Ángel. 2002. *El español de América*. Caracas: Biblioteca Ayacucho.

Schreier, Daniel. 2009. Language in isolation, and its implications for variation and change. *Language and Linguistic Compass* 3(2). 682–699. https://doi.org/10.1111/j.1749-818X.2009.00130.x

Sedano, Mercedes. 2000. La perífrasis de gerundio en Caracas y otras ciudades hispanohablantes. *Nueva Revista de Filología Hispánica* 48(2). 253–274. https://doi.org/10.24201/nrfh.v48i2.2123

Stalnaker, Robert. 1978. Assertion. In Peter Cole (ed.), *Pragmatics*, 315–332. New York: New York Academic Press.

Stalnaker, Robert. 2002. Common Ground. *Linguistics and Philosophy* 25. 701–721.

Torres Cacoullos, Rena. 2000. *Grammaticization, synchronic variation, and language contact. A study of Spanish Progressive -ndo constructions*. Amsterdam/Philadelphia: Johns Benjamins. https://doi.org/10.1075/slcs.52

Torres Cacoullos, Rena. 2009. Variation and grammaticisation: The emergence of an aspectual opposition. In Stavroula Tsiplakou, Marilena Karyolemou & Pavlos Pavou (eds.), *Studies in language variation: European perspectives II (ICLaVE 4)*, 215–224. Amsterdam: John Benjamins. https://doi.org/10.1075/silv.5.18cac

Torres Cacoullos, Rena. 2012. Grammaticalization through inherent variability: The development of a progressive in Spanish. *Studies in Language* 36. 73–122.

Wellman, Henry. 1990. *The child's Theory of Mind*. Cambridge, MA: MIT Press.

Winter, Bodo. 2013. Linear models and linear mixed effects models in R with linguistic applications, ms.

Zipf, George K. 1949. *Human behavior and the principle of least effort*. Cambridge, MA: Addison-Wesley Press.

Adult language and dialect learning as simultaneous environmental triggers for language change in Spanish

Israel Sanz-Sánchez and Fernando Tejedo-Herrero
West Chester University / University of Wisconsin – Madison

Grammatical restructuring in contact situations is customarily analyzed under the lens of either language contact or dialect contact. In this study we argue that both processes may operate jointly in social settings where dialectal accommodation and adult L2 learning favor the same linguistic outcomes. From an evolutionary perspective, this overlap between both forms of contact may be conceptualized as a function of a common underlying process, with speakers selecting features of heterogeneous provenance, acquired at various life stages. We exemplify this joint effect by focusing on two changes in the history of Spanish: the rearrangement of the 3rd person object clitic system in medieval southern Iberian Castilian and the merging of the medieval sibilants in early colonial Spanish.

Keywords: evolutionary linguistics, contact-induced change, adult language and dialect learning, clitic pronouns, sibilant sounds

1. Introduction

This study explores the combined effects of adult language acquisition (L2 learning) and dialect accommodation (D2 learning) as possible simultaneous co-directional sociohistorical sources of influence on specific instances of language change. Although *language contact* and *dialect contact* have often been treated as separate processes (Trudgill 1986; Thomason & Kaufman 1988), we will argue that no such difference exists. Instead, the purported difference between dialect and language contact effects should be placed along a *continuum* of learnability of specific features determined by the typological distance between the features in contact and actuated in specific sociodemographic settings. From this perspective, many of the similarities between language and dialect contact identified by the literature can be understood as a function of shared universal constraints on adult language learning

https://doi.org/10.1075/ahs.12.c05san

(Lenneberg 1967; Birdsong 1999; Hudson Kam & Newport 2005; Siegel 2010: 96–100), rather than as accidental coincidences between two separate processes.

The commonalities between D2 and L2 learning as usually posited in the literature have direct implications for sociohistorical explanations of language change. These implications are most evident in situations of demographic rearrangement involving highly heterogeneous feature pools, where various types of adult learners must negotiate a socially and linguistically unsystematic landscape (cf. *unfocused* situations, Trudgill 1986; *catastrophic events*, Labov 1994; Hickey 2003). In these situations, specific solutions may present acquisitional advantages for individual learners regardless of their diachronic transmission as 'native' or 'non-native.' In turn, the agency of adults in these socio-demographic situations determines the variable primary linguistic data that new generations of children will ultimately use to construct their own grammars (cf. *cascade effect*, DeGraff 1999, 2009). This process can be seen at work both at the level of individual learners, who construct idiolects by negotiating a heterogeneous input, as well as at the level of the community, where individuals' learning patterns in a sociolinguistically unstable landscape contribute to blur pre-contact linguistic pools (whether these are seen as having involved separate languages or dialects of the same language).

We will illustrate this framework by applying it to two changes in the history of Castilian Spanish: a phonological change (the simplification of the medieval Castilian sibilant system in Early Colonial Spanish – ECS) and an historically more remote morphological change (the reformulation of 3rd-person clitics in Medieval Southern Iberian Castilian – MSIC). Both of these changes were nestled in social settings characterized by pervasive contact among native and non-native adult learners. Earlier accounts of these changes (Granda 1994; Parodi 1995; Tuten 2003; Fernández-Ordóñez 2012) have presented them largely as the result of dialect contact. In contrast to these analyses, we submit that the variants favored in each of these two historical contexts offered acquisitional advantages to adults of various backgrounds, including speakers of various Ibero-Romance varieties (i.e., traditionally thought of as D2 learners) as well as of other languages, who grew up in multilingual environments or who were exposed to Spanish as adults (i.e., L2 learners). In addition, sociohistorical evidence shows that, in both changes, non-native adult language learners played a more prominent demographic role than usually entertained by previous approaches. Therefore, a more sociohistorically accurate description of the *actuation* (Weinreich, Labov & Herzog 1968) of these changes should not place agency exclusively on dialect learners, but should also incorporate the learning experiences of other individuals along the D2-L2 learning continuum. This paper is a first step in that direction.

This study is structured as follows: Section 2 presents some of the shortcomings of the traditional theoretical schism between language contact and dialect contact for sociohistorical approaches to language change. Section 3 outlines an alternative

evolutionary-ecological framework based on the role of individual users in selecting features according to their age-specific learning abilities, regardless of their origin as features due to dialect contact, language contact, or both. Sections 4 and 5 apply this framework to the evolution of sibilants in ECS and of 3rd-person clitics in MSIC, respectively, with an emphasis on the role of adult learners in the local ecologies and the evidence for the higher learnability of the selected features vis-à-vis other available variants. Section 6 advances some of the implications of these findings for sociohistorical approaches to language change from the perspective of adult and child language acquisition and concludes this study.

2. Language contact vs. dialect contact

A well-established axiom among linguists is that it is not possible to establish categorical, qualitative linguistic differences between 'languages' and 'dialects' (e.g., Chambers & Trudgill 1998: 3–12). Traditional approaches to dialectal variation as a social phenomenon, where speakers in monolingual communities are seen as juggling a variety of contextual variables belonging to the same language, have typically been unconcerned with this distinction (Milroy and Milroy 1992; Labov 2001; Martín Butragueño 2010). By contrast, recent studies of dialect change in urban areas have had to account for the effects of learning in dialectal environments where significant amounts of speakers have acquired their variety as a second language (Sharma & Sankaran 2011; Otheguy & Zentella 2012; Kerswill 2013) and where new dialectal varieties emerge that also show the effects of non-native L2 learning.

Studies of language contact and change have historically operated under the assumption that a difference between 'languages' and 'dialects' does exist and that it is based on degrees of mutual intelligibility and structural divergence. Exactly how unintelligible or structurally different two codes need to be in order to be classified as languages is seldom made explicit. Because of this assumption, the linguistic outcomes of contact among dialects and among languages are also assumed to be qualitatively different: "There are certain language contact phenomena we can expect to find if there has been language contact […], whereas we cannot necessarily expect the same phenomena if there was dialect contact […]" (Jahr 1999: 131; see also Thomason & Kaufman 1988; Thomason 2001: 9). As a consequence, contact studies have historically developed into two separate strains, one concerned with contact among languages (e.g., Thomason & Kaufman 1988; Coetsem 2000; Thomason 2001; Winford 2007) and another one concerned with contact among dialects (e.g., Trudgill 1986, 2004; Kerswill & Williams 2000; Hickey 2003; Labov 2007). Acquisition studies are another area where the rift between the study of 'dialects' and 'languages' is prominent, with scholars studying the learning of second languages (L2s) (Ellis 2008; Mitchell, Myles & Marsden 2013; VanPatten & Williams

2020) and of second dialects (D2s) (Chambers 1992; Siegel 2010) largely as different processes. Explicit formulations of comprehensive contact frameworks are rare (e.g., *code copying*, Johanson 2002) and have usually not been tested empirically.

Despite the practice in contact studies to treat 'language contact' and 'dialect contact' separately, many shared commonalities between both have been identified in the literature. These include sensitivity to age-related effects (Chambers 1992; Siegel 2010); selectivity in learning, with adults favoring specific options from their input, rather than acquiring an unlimited number of options (cf. *salience*; Auer, Barden & Grosskopf 1998, Erker 2017); similar mechanisms, such as *transfer, simplification*, and *overgeneralization* (Trudgill 1986; Silva-Corvalán 1994); and similar results, with both yielding variable, yet systematic grammars that may differ significantly from those of L1 or D1 speakers (Chambers 1992; Labov 2007; Geeslin 2011, 2018). Throughout this chapter, we will submit that these similarities are far from coincidental. On the contrary, they are manifestations of common learning processes in L2 and D2 acquisition. In turn, these processes are a function of shared universal cognitive constraints that are active in post-critical period learning (cf. Section 3).

The practical limitations of this theoretical divide are most obvious in demographic situations that allow speakers to interact across dialectal and language lines simultaneously. In these situations, investigators struggle to attribute the observed consequences of change to either language contact or dialect contact. For instance, studies of the emergence of new varieties of Spanish in New York City assign a role to *both* contact among dialects of Spanish and between the latter and English in favoring specific changes. A clear example is the increase in the use of overt subject pronouns in New York Spanish, which the literature has attributed to a combination of the effect, on the one hand, of dialectal contact among L1 varieties of Spanish with varying rates of overt subject pronoun use, and, on the other, contact with English, where subject pronouns are categorical in most syntactic contexts (Otheguy & Zentella 2012; Orozco 2018). The possibility that dialect contact and language contact may simultaneously favor the same linguistic outcome in a given sociolinguistic environment is also raised in historical studies. For instance, Company Company (1995) argues that double possessive structures in colonial Latin American Spanish (e.g., *su casa de ella* 'her house', lit. her house of hers) have been historically motivated by both the availability of these structures in the transplanted L1 Iberian dialectal varieties and non-native transfer from similar constructions in some indigenous languages (e.g., Nahuatl).

To summarize, the theoretical isolation of 'dialect' from 'language' contact effects is problematic for the study of historical situations of language change for theoretical and methodological reasons. In order to go beyond these limitations, we must understand how individuals negotiate the available linguistic input in actual communicative settings, a point which we explore in the next section.

3. Contact among systems vs. contact among speakers: Language acquisition and language change from an evolutionary-ecological perspective

As noted in Section 2, the observed processes and results in stereotypical 'language contact' and 'dialect contact' situations often overlap. The basis for these similarities lies in the different learning abilities of children and adults, with children usually credited with much higher rates of success than adults at reproducing identifiable targets in primary linguistic data (Lenneberg 1967; DeGraff 1999, 2009; Siegel 2010: 96–100). Adults are also less successful than children when it comes to learning linguistically or socially conditioned variation (Chambers 1992; Labov 2007; Trudgill 2011), as well as building systematic, rule-based systems out of highly unsystematic input (Senghas & Coppola 2001; Hudson Kam & Newport 2005). These age-based learning differences seem to apply to individuals in language and dialect contact situations alike.

At the same time, it is undeniable that humans can learn much of the new linguistic material that they are exposed to after childhood. From a developmental perspective, therefore, the question concerns what qualifies as 'learnability' in adult language acquisition. In this paper, we define *learnability* in adult acquisition as a typological property: observable cross-linguistic typological frequencies reflect universal biases on language production, processing, and learning by adults (Thomason 2001: 52, Croft 2003; Giacalone Ramat 2003), which apply to all forms of contact. From this perspective, learnability is a matter of degree. To exemplify this concept, highly learnable (i.e., typologically frequent) options include the absence of phonological codas in syllable structure; isolating or agglutinating morphology (rather than inflectional morphology); or S-first orders in syntax (i.e., SOV, SVO), among others (Croft 2003; Clements 2009: 19–20).[1] Although learnability is a function of universal language processing constraints, the activation of these constraints is mediated by local conditions of communication that determine the

1. As accurately pointed out by one of our anonymous reviewers, the term *learnability* has received various definitions in the language acquisition literature. It has traditionally been formulated as a general property of human grammars that makes them suitable for deduction on the basis of input received by children, either on the basis of innate linguistic principles stated as a universal grammar or as a function of general learning principles (cf. Parker 1989). In this paper, as stated above, we are defining it as a *typological* property: in principle, insofar as all options in the feature pools that emerge in contact situations are produced by language users, they are learnable from a language acquisition point of view. However, in specific contact situations where individuals are presented with several competing options in various frequencies and contexts of use, some forms are more readily acquired than others. This preference is mediated by universal cognitive biases, which can be indirectly gleaned from cross-linguistic typological distributions.

degree of exposure that adults get to native and non-native features, as well as their motivation to acquire them.

From an *evolutionary-ecological* perspective, user-to-user contact is always a function of the linguistic and social environments in which specific forms interact and compete (*feature pool*), and the locus of this competition is always the mind of the individual language user (Croft 2000; Mufwene 2001, 2008; Clements 2009). Feature pools include elements of heterogeneous provenance, which, depending on the sociolinguistic environment, may originate simultaneously in different 'dialects' and 'languages' (Mufwene 2008: 19, 27). It is from these structurally and historically hybrid pools that individuals select elements to shape their personal idiolects, and these features can be selected categorically or in variation with other alternatives. Selections can also change over an individual's lifespan, as when one variant increases its frequency over others. Which specific features individuals select depends on a variety of demographic, social, structural, and cognitive factors: which features are more frequent in the specific pool that the individual has access to, which are simpler (i.e., involving fewer grammatical distinctions), which are more similar to the individual's L1, which are valued positively at the symbolic level by the community, and which are more learnable in the cognitive-typological sense mentioned above (Clements 2009: 26–27).

The resulting selections can reflect options already existing in the available feature pools, as well as innovative reinterpretations of these features (e.g., the emergence of clitic options uncharacteristic of most L1 forms of Spanish in varieties that have historically included large numbers of adult non-native speakers, such as Andean Spanish or Basque Spanish, see Klee and Lynch [2009: 47–55, 138–144] for a summary of the literature). In this process, learners' primary interest is in achieving communication by maximizing the expressive power of their linguistic resources rather than in preserving the boundaries of any assumed pre-contact communal pool, whether a 'language' or a 'dialect' (Negrão & Viotti 2014). Both children and adults play a role in this process: child acquisition ensures the diachronic survival of specific options, but this acquisition occurs on the basis of the heterogeneous linguistic input provided by adults in a *cascading* relationship between adult and child language acquisition (DeGraff 1999, 2009, cf. Section 6). Therefore, understanding how adults negotiate their input and learn specific options in a contact situation is a critical piece in the study of the long-term evolution of contact language communities.

In an evolutionary-ecological approach, individuals are seen as members of populations. Therefore, *individual selections drive observed patterns of change at the level of the community*, and language change emerges as the cumulative effect of changing individual selection patterns over time (Mufwene 2008: 120). In other words, selection via learning is, first and foremost, a process at the level

of the individual language user, with community-wide consequences when individuals are seen as members of populations. Despite the difficulties involved in relating the behavior of individuals to the patterns of change at the level of the community (Labov 2001: 34), it is in selections of individual users that innovations originate, and it is in user-to-user communication that these innovations spread across the community (Milroy & Milroy 1992; Eckert 2000; Villena-Ponsoda 2005). From a population perspective, all language change is ultimately contact-based. In evolutionary-ecological approaches to language learning and change, language has been described as a *complex adaptive system*, where the behavior of learning agents, responding to cognitive and social competing factors, gives rise to generalities and differences at the level of the population (Larsen-Freeman 2007; Blythe & Croft 2009; Bybee 2010; Kretschmar 2015).

The ultimate irrelevance of the distinction between dialect contact and language contact to an evolutionary-ecological understanding of language change is best illustrated by historical situations that involve the massive simultaneous admixture of feature pools including elements of varying degrees of typological distance coupled with the absence of clear social correlates for variation (cf. *unfocused*, *catastrophic* situations, Section 1). Far from being an historically unusual scenario, this is precisely the case of many instances of language expansion and contact. In these situations, users interact in ecologies that exhibit several typical triggers, including weak social networks; rapid rearrangement of pre-contact social structures; diffuse feature pools of both native and non-native provenance; and few social mechanisms for the enforcement of linguistic norms. In colonial or resettlement societies, these triggers are particularly active in the earliest decades, before the development of new local social hierarchies and more stable norms of use (Schneider 2009). The need for learners of various ages and ethnolinguistic backgrounds to communicate in a sociolinguistically unstable environment increases the possibility of selection irrespective of pre-contact sociolinguistic boundaries.

In what follows, we will illustrate the applicability of this approach to the study of two situations of change in the evolution of Spanish. Both of these situations were nestled in social ecologies that were conducive to pervasive contact among adult speakers across dialect and language lines, but the preceding literature has tended to put the burden of proof on dialect contact effects alone. Although the available data do not answer all our questions, the extant sociodemographic and acquisitional evidence in both cases strongly suggests that adult L2 learners were also agents of change by contributing to the selection of specific features. Our goal is to sketch a general rationale to consider the combined action of native and non-native adult language learners as a critical sociohistorical trigger for these changes. In the following sections, the literature reviewed is summarized, with more detail provided only when strictly necessary for our argument.

4. Early colonial Spanish sibilants

4.1 The internal ecology of ECS fricative sibilants

Most current varieties of Spanish exhibit only one sibilant fricative phoneme: /s/. This is in clear contrast with pre-colonial Spanish, for which several contrastive sibilant phonemes are posited. The reduction of several of these medieval sibilants to /s/ is termed *seseo* and is commonly described as characterizing Latin American Spanish since the very beginning of the colonial period (Menéndez Pidal 1962; Parodi 1995). In order to understand how this feature became general in Latin American Spanish, we need to start from a brief description of pre-colonial sibilant fricatives.

Phonological changes in late medieval Castilian Spanish (ca. 1300–1500) yielded a wide range of fricative sibilant contrast options (Menéndez Pidal 1962; Lapesa 1981; Penny 2000, 2004). Table 1 shows the traditionally posited late medieval system, and (1) exemplifies these contrasts and their canonical late medieval spelling. Tables 2 and 3 show two of the systems that emerged from this reconstructed original system, via the loss of voicing (Table 2), and of place of articulation distinctions followed by the loss of voicing (Table 3), respectively:

Table 1. Original late medieval (ca. 1300–1500) Castilian system of sibilant fricatives

Dento-alveolar		**Apico-alveolar**		**Palato-alveolar**	
[-voice]	[+voice]	[-voice]	[+voice]	[-voice]	[+voice]
/s̪/	/z̪/	/s̻/	/z̻/	/ʃ/	/ʒ/

(1) a. Dento-alveolars

<ç, c> = /ts/ (> late medieval /s̪/) — <z> = /dz/ (> late medieval /z̪/)

/de'tsiɾ/ *deçir* 'to descend' — /de'dziɾ/ *dezir* 'to say'

b. Apico-alveolars

<s> (word-initial), <ss> (intervoc.) = /s̻/ — <s> (intervoc.) = /z̻/

/es'peso/ *espesso* 'thick-MASC.SING.' — /es'pezo/ *espeso* 'spent-PARTIC.'

c. Palato-alveolars

<x> = /ʃ/ — <j, g> = /ʒ/

/'fiʃo/ *fixo* 'fixed-MASC.SING.' — /'fiʒo/ *fijo* 'son'

Table 2. Derived system A

Dento-alveolar		Apico-alveolar		Palato-alveolar	
[–voice]	[+voice]	[–voice]	[+voice]	[–voice]	[+voice]
/s̪/ ←	/z̪/	/s̻/ ←	/z̻/	/ʃ/ ←	/ʒ/

Table 3. Derived system B

Dento-alveolar		Apico-alveolar		Palato-alveolar	
[–voice]	[+voice]	[–voice]	[+voice]	[–voice]	[+voice]
/s̪/	/z̪/	/s̻/	/z̻/	/ʃ/	/ʒ/
/s̪/	/z̪/				

Note that system A, with a contrast between two front sibilants, was the precursor of the present-day northern and central Castilian contrast between /θ/ and /s/, while system B, in which dento-alveolars and apico-alveolars merged into /s/, provided a model for Latin American *seseo*. Although there is disagreement as to the reasons for the phonological readjustments in Tables 2 and 3, there is consensus that these simplified systems spread geographically in the 15th century. Much of the literature identifies system A with northern and central Castilian areas and system B with Andalusia (especially in the lower Guadalquivir valley). However, the documentary record shows many local exceptions, including cases of devoicing without place of articulation merger in Andalusia, as well as loss of place of articulation in central and northern areas (e.g., Tuten 2003: 248, Penny 2004: 65, Vázquez Balonga & Sánchez-Prieto Borja 2015). Contact with Basque, Portuguese, and Catalan in specific areas or social groups where bilingual speakers were present also resulted in various forms of merger and further contributed to the diversification of the options typical of Castilian dialects (Penny 2000: 116–117). Variation was not only geographical, but also social (Alonso 1967, 1969; Penny 2002: 98–103). Overall, we can conclude that the distribution of sibilant fricatives in Castilian Spanish varieties at the onset of the colonial period (ca. 1500) was geographically and socially heterogeneous.[2]

Demographic mixture in the early American colonies (roughly the 16th century) contributed to further complicate this sociolinguistic landscape. The newcomers included settlers from various Spanish regions (but mostly central and

2. An anonymous reviewer proposes grammatical descriptions of late medieval and early modern Spanish as a possible source of information about dialectal variation in sibilants in Iberian Spanish. Although late 16th- and 17th-century descriptions do pay attention to aspects of this variation (as studied, among others, by Alonso 1967, 1969), earlier descriptions (such as Nebrija's *Gramática de la lengua castellana* in 1492 or Juan de Valdés's *Diálogo de la lengua* in 1535) are conspicuously silent about the types of variation in sibilants attested in the archival record around the same period. Instead, they are generally reflective only of the canonical system still in vogue as a prestigious model of pronunciation (i.e., the system in Table 1).

southwestern areas, cf. Boyd-Bowman 1976), as well as several other exogenous contingents (other Europeans, African slaves), who from the start interacted with indigenous populations. This mixture meant that all the sibilant options surveyed above, and possibly others, were represented in the early colonial feature pool (Granda 1994; Parodi 1995), although with varying demographic weights. This feature pool must have included both L1 and L2 realizations. The traditional assumption in the literature has been that system B (i.e., the 'Andalusian' system) was the most prominent one demographically in the L1 input from the start (Menéndez Pidal 1962; Lapesa 1981; Penny 2000).

Recent research has questioned this assumption. For instance, Sanz-Sánchez (2019) analyzes sibilant spellings in an epistolary corpus of 16th-century Spain-born immigrants in Latin America. He shows that while some Andalusians did indeed show evidence of sibilant merger across places of articulation (i.e., system B), others exhibited a Castilian-like system (system A), with devoicing but no merger across places of articulation. For instance, the letters by Andalusian friar Andrés de Moguer include several spellings that are strongly indicative of the erosion of the distinction between dento-alveolars and apico-alveolars (*altesa* 'highness', cf. canonical late medieval *alteza*; *beza* 's/he kisses', cf. *bessa*). But letters by roughly contemporary fellow Andalusian friar Domingo de la Anunciación exhibit no such spellings – instead, all of his departures from canonical medieval orthography appear to indicate devoicing (e.g., *pasados* 'past-MASC.SING.', cf. *passados*; *favorezeran* 'they will favor', cf. *favoresceran*). By combining archival, metalinguistic, and comparative evidence, Sanz-Sánchez concludes that system A devoicers must have initially been more demographically prominent than usually assumed, and that the eventual selection of system B should be explained in terms of the advantages it offered for various types of learners (see below for more on this point). Other approaches (e.g., koinéization-based accounts: Fontanella de Weinberg 1992; Granda 1994) underscore the structural simplicity of system B, with only one rather than two front sibilants, as the reason for its eventual selection in vernacularized forms of colonial Latin American Spanish.

Although they differ in the motivations for selection (demographic prominence vs. structural simplicity), most of the approaches reviewed in this section share the same sociohistorical premise that L1 speakers of Spanish were the primary agents of selection in ECS (Sanz-Sánchez 2019 is an exception). In these accounts, speakers of L2 are assigned a marginal role, both demographically and sociolinguistically. As will be seen in the next section, this assumption is anything but straightforward, and it has serious implications for an ecologically informed account of linguistic selection in the early Spanish American colonies.

4.2 The external ecology of the early Spanish American colonies

As stated above, the new colonies featured a highly diverse sociodemographic situation, including large contingents of non-European speakers of Spanish as an L2. Contrary to what is often assumed (Escobar 2001; Lipski 2014), not all early colonial Spanish American societies were organized on the basis of impermeable layers. There is evidence that significant degrees of everyday contact occurred across ethnolinguistic lines (Salas 2000; Brain 2010), both at home and in public. Demographic data point in this direction. For instance, there was a notable imbalance between the number of men and women among Spanish colonists, with the former constituting most Spain-born settlers (Boyd-Bowman 1976: 597). Parochial records from areas where Spaniards settled intensively since the earliest decades give evidence of high rates of interethnic marriages (e.g., Schwaller 2016). In addition, interethnic mixing was the norm in non-canonical unions (Sánchez-Albornoz 1989: 85–86). Interethnic contacts were also common among adults in domestic, administrative, and labor contexts (González Ollé 1996; Brain 2010).

Despite the undeniable hierarchical nature of colonial societies, everyday life in many of these early Spanish settlements offered abundant opportunities and motivations for individuals to engage in multiple forms of learning. These acts of learning must have gone far beyond those typical of communication among Spain-born colonists. In addition to the heterogeneous, diffuse nature of the sociolinguistic landscape, there were few mechanisms to enforce a given norm as the privileged target of acquisition. Dialectal norms based on L1 speakers of Spanish must have been less than influential beyond the social networks where frequent access to such models was available. Furthermore, the lack of widespread schooling typical of pre-modern societies (Gonzalbo Aizpuru 2001) and the absence of other norm-enforcing mechanisms at work in more established colonial societies (Trudgill 2004; Schneider 2009) meant that there was little incentive to acquire any given dialectal norm. In this environment, learners must have had evidence of many possible native and non-native realizations for the sibilant systems originally imported by L1 speakers, with little to no clear sociodemographic support as to which system to acquire.

Recent work in historical sociolinguistics has stressed the degree to which reconstruction of the patterns of everyday communication in a given community is critical to the understanding of its linguistic history (Auer et al. 2015). However, by foregrounding the importance of demographics and structural criteria in determining the outcome of contact and change, literature on ECS has given little attention to the actual ecological conditions of communication and learning for speakers in this environment, especially adult learners. The following paragraphs are an attempt at addressing these shortcomings.

4.3 Acquisition of sibilants in ECS from a cross-linguistic perspective

Since ECS is no longer being acquired, the documentary record can give us clues about how adult learners processed sibilant systems.

Among adult L1 speakers, there is evidence to suggest that *interdialectal accommodation* (Kerswill 2002; Trudgill 2004) favored sibilant simplification. For instance, letters by 16th-century colonists born in northern and central Castile sometimes show spellings that are indicative of the system that appears to have been more common in western Andalusia (cf. Table 3). An example is Pedro Gómez de Montejo, a native of the town of Frías in northern Castile. He used many non-canonical spellings suggestive of system A devoicing, as seen among most other Castilians, but also a few occasional spellings reminiscent of Andalusian-like place of articulation merger (e.g., *apersebido* 'warned-MASC.SING.', cf. canonical *apercebido/apercibido*; for more examples, see Sanz-Sánchez 2019: 217–219). These spellings suggest that native sibilant contrasts were unstable even among L1 speakers (Frago Gracia 1999) and can be understood as a manifestation of the loss of phonological distinctions typical of dialect contact situations (Chambers 1992). The prevalence and direction of these D2 acts of accommodation can hardly be reconstructed from the documentary record (recall that system B merger across places of articulation is not attested only in documents from Andalusia as noted in Section 4.1 above). Despite the shortcomings of the historical records, we can expect Spain-born individuals to have been under the same motivation to accommodate to other speakers and to reformulate parts of their original phonological system as has been recorded in other situations of pervasive dialect contact (Trudgill 1986, 2004; Siegel 2010).

More definitive evidence for the effects of contact on the transplantation of Castilian Spanish sibilants to the early American colonies can be found in the documentary record that reflects forms of L2 acquisition. A common denominator of these representations is the variable, non-native interpretation of L1 contrasts. Even before the beginning of the colonial period, adult non-native speakers had found it hard to acquire several of these contrasts. For instance, literary representations of Castilian as spoken by Arabic-speaking *moriscos* (i.e., ethnic Muslims) and Romani-speaking gypsies suggest that non-native sibilant realizations among these groups were common (Plans 2004). Indeed, some of the sibilant contrasts transplanted to the colonies continued to be challenging for the new groups of non-natives speakers. Representations of the speech of African slaves, both in Spain and in the colonies, commonly record various forms of sibilant merger (Plans 2004; Lipski 2005: 75, 81, 89).

Among speakers of indigenous languages, non-etymological realizations during the early colonial period are also attested. These attestations can be found

not just in literary representations, but more importantly, in everyday archival documents produced by indigenous bilingual scribes (Karttunen & Lockhart 1976; Lockhart 1992). Representations of Spanish loanwords in these indigenous sources are particularly relevant to reconstruct the strategies followed by L2 speakers in the learning of Spanish sibilants. In these indigenous sources, a variety of spelling strategies are apparent. For instance, in colonial Nahuatl manuscripts, some sibilant spellings preserve the late medieval written Castilian contrasts, which could in turn reflect written transmission of these spellings or exposure to spoken forms of L1 Spanish where these contrasts were still active[3] (e.g., Nah. *mesa* 'table' < Sp. *mesa*; *cevara* 'barley' < *cebada*). But non-canonical spellings abound. Some appear to be approximations to dialectal articulations of sibilants present in the L1 input, including the *seseo* merger between dento-alveolars and apico-alveolars (e.g., *çiera* 'saw' < *sierra*). Yet other spellings appear to represent the more or less regular acoustic mapping of Castilian Spanish sibilants to the L1 system of Nahuatl speakers (e.g., the reproduction of apico-alveolars, perceived as "back" palato-alveolars: *xonbla* 'shadow, shade' < *sombra*). Given the heterogeneity of the sibilant feature pool in the early colonial period, the presence of these trends in the output of specific scribes can hardly be attributed only to their lack of familiarity with the spelling norms of Spanish. Rather, we submit that it was a direct reflection of the highly variable input available to L2 learners, including both L1 and L2 forms (see Sanz-Sánchez 2019: 223–225 for evidence of the ways in which L2 Spanish speakers learned the sibilants present in L1 Spanish).

In addition to the variable nature of the input stemming from this ecology of pervasive contact, the relative learnability (see Section 3) of the options contributed by L1 speakers must have proved an additional hurdle to their reproduction by adult learners. Cross-linguistic data (Maddieson 1984; Ladefoged & Maddieson 1996) strongly suggest that some of the options imported by L1 users of Spanish were highly marked. For instance, although the vast majority of languages feature at least one fricative sibilant, when more than one sibilant is present, they tend to be distributed maximally (i.e., a "front" /s/-like sound and a "back" /ʃ/-like sound). Two-way splits between front sibilants (e.g., dento-alveolar /s̪/ vs. apico-alveolar /s̻/), as well as three-way sibilant place of articulation contrasts of the type that still seem to have been in place for some pre-colonial Castilian Spanish varieties (Tables 1 and 2) are extremely rare. A study of the articulatory, perceptual, and functional factors that determine these cross-linguistic distributions is well beyond the scope of this paper. But critically for our argument, if these distributions are an

3. All examples in this paragraph are extracted from the corpus of Spanish loanwords in colonial Nahuatl manuscripts in Karttunen & Lockhart (1976).

historical function of universal acquisitional biases, we may conclude that some of these L1 sibilant systemic options were badly positioned from the start as targets of reproduction by adult learners, i.e., they were less learnable. By contrast, other solutions were less marked – most notably, the maximal distribution between a front and a back sibilant, quite common cross-linguistically.

To confirm whether these restrictions were also active for adults in the ecology of ECS, acquisition data from observed contact environments may be used. Studies of D2 learning show that, unless presented with a strong normative motivation to do the opposite, adults tend to dephonologize marked sibilant contrasts (cf. reduction of three-way to two-way sibilant contrasts in dialects of Mandarin Chinese, Polish, and Basque; Trask 1997; Zygis 2003; Shih 2012). Similar outcomes have been observed in contexts of L2 learning, with adults eliminating problematic contrasts or producing variable articulations of the relevant lexical classes (Spanish-Basque, Larraza 2009; Taiwanese-Mandarin, Shih 2012; Polish-English, Marecka et al. 2015; French-English, Shoemaker 2016). Experimental studies of L2 learning show that adult speakers have a hard time distinguishing and reproducing non-native sibilant contrasts (Chiu 2010). All of this evidence allows us to interpret the innovative solutions that adult speakers develop in contact situations as learnability strategies stemming from the acquisitional hurdle posed by marked sibilant contrasts.

4.4 Towards a new account of ECS sibilants: The role of adult language learning

We can now proceed to sketch a scenario for the simplification of sibilants in ECS that takes into account these social and cognitive triggers. In the early Spanish colonies, population mixture and everyday communication among individuals of European and non-European provenance (including both "native" and "non-native" learners) must have created an extremely diffuse Spanish feature pool. D2 and L2 learners engaged in the erosion of marked sibilant contrasts, applying the same acquisitional strategies that have been recorded elsewhere in the literature. In this environment, the reduction of sibilant variation to the front-back voiceless contrast (/s/ – /ʃ/) that system B *seseo* relies on had a built-in advantage for adults: it was available as one of the transplanted options, it preserved two segments which were phonetically present in every transplanted sibilant system, it coincided with many of the original L1 systems of non-natives, and it presented adult learners with a cross-linguistically unmarked, more learnable option.

At the same time, as observed in other situations of pervasive contact, adults in the early decades are unlikely to complete the task of creating a new community

norm. It was children born in the new environment who must have proceeded to stabilize this variable, unsystematic adult input (Kerswill & Williams 2000; Trudgill 2004), applying the same acquisitional tendencies to regularize and systematize unsystematic adult input observed in other situations (i.e., the emergence of creoles, Roberts 2000; the birth and development of new sign languages, Senghas & Coppola 2001; experimental studies of child acquisition of artificial language input, Hudson Kam & Newport 2005). Both adults and children were active in the construction of this new colonial communal norm – however, adults were instrumental in blurring the sociolinguistic correlates for specific sibilant options typical of pre-contact populations. They also shaped the frequency of specific phonetic segments in the input pool, further motivating children to favor specific phonemic alternatives as targets for acquisition (cf. *cascade effect*, DeGraff 1999, 2009; see Section 6).

Note that this is a simplified version of what must have undoubtedly been a more complicated situation sociolinguistically: not all adults had the same degree of contact with L1 Spanish pools nor the same degree of motivation to acquire them. Some adults must have been highly successful in reproducing specific L1 models, while others were less so. Different demographic proportions of D2 and L2 learners in specific colonial settlements must have shaped the social spread of merged sibilants. Continued arrivals of adult learners, from Spain and elsewhere, kept adding variability to the linguistic pool – for instance, by introducing innovations such as the interdental [θ] articulation of late medieval dento-alveolars, which started spreading in north-central Castile around 1600 (Cano Aguilar 2004: 842–843) and which at least some of the later Spain-born immigrants must have brought to the colonies. However, the demographic and sociolinguistic weight of these new arrivals diminished as time went by (Lipski 2014), and by the early 17th century, the use of system B in ECS appears to have been all but categorical (Sanz-Sánchez 2019). Undoubtedly, the historical record on how different types of adult speakers in ECS processed their sibilant fricative input is fragmentary. However, because native and non-native learners interacted on an everyday basis in their local communication ecologies, a sociohistorically realistic account of language change in this context must consider their contribution.

This hermeneutical axiom is also valid for more historically remote situations, for which linguistic and social information may be even scarcer. The southward expansion of Castilian Spanish in the Iberian Peninsula throughout the medieval period is a clear example, and is analyzed in the next section.

5. Object Pronoun paradigms in Medieval Southern Iberian Castilian (MSIC)

5.1 The internal ecology of MSIC clitics

The third-person Ibero-Romance clitic system has historically displayed a wide range of paradigms (for in-depth dialectal and historical treatments of these paradigms, see Klein-Andreu 1999; Fernández-Ordóñez 1999, 2012; Flores Cervantes 2006). In the Castilian area, this variation is clearest along the east-west continuum and in the north-central areas, where a broad variety of dialectal clitic systems exists. Among these systems, two are particularly common. One of them keeps the case-determined distinctions already active in Latin, and is characterized by morphological marking of case, gender, and number, as expressed in Table 4. Variations of this *case-based* paradigm can be found in north-western Ibero-Romance (Galician, Astur-Leonese), on the eastern fringe of the Castilian area (e.g., La Rioja) and, most relevant for this discussion, in western Andalusia.

Table 4. Case-based system (based on Penny 2000: 91)

	Singular		**Plural**	
	Masculine	**Feminine**	**Masculine**	**Feminine**
Accusative	*lo*	*la*	*los*	*las*
Dative	*le*		*les*	

The other type of 3rd-person clitic system involves virtual loss of case marking, which is replaced by *semantic features* like gender, animacy, and count/non–count distinctions, as outlined in Table 5.

Table 5. Semantically based system; northern Castilian system (based on Penny 2000: 92)

Count				**Non–count**
Singular		**Plural**		
Masculine	**Feminine**	**Masc.**	**Femin.**	
le	*la*	*les~los*	*las*	*lo*

Historical and documentary evidence shows that these systems took shape throughout the medieval period, with the semantically based paradigm emerging in the southern flanks of the north-western region (i.e., in northern Castile) and spreading southwards in the 13th century, following the repopulation efforts by the Christian

kingdoms (Fernández-Ordóñez 1994, 2001, 2012). Further south, these unstressed third-person clitic systems eventually gave way to a number of so-called *transitional* paradigms (Fernández-Ordóñez 1994), which exhibit hybrid properties between the case-based and the semantically based prototypes. Such paradigms are found in southern Castile north of Andalusia. Table 6 outlines the general properties of these transitional systems.

Table 6. Transitional/hybrid system (based on Fernández-Ordóñez 2001: 14)

	Singular					**Plural**	
	Masculine			**Feminine**		**Masc.**	**Femin.**
	Animate	**Inanimate**					
		Count	**Non–count**	**Count**	**Non–count**		
Acc.	*le*	*le*	*lo*	*la*	*la*	*los*	*las*
	(lo)	*(lo)*			*(lo)*		
Dative	*le*	*le*		*le*		*les*	*les*

Following the reincorporation of western Andalusia into the Kingdom of Castile during the 13th century, it was the case-based paradigm that appears to have gained acceptance, spreading out of the urban centers in the area (e.g., Seville) and being eventually selected as the most common system in colonial Spanish American varieties (Lapesa 1981: 565–567). Given that the southward expansion of Ibero-Romance during the medieval period originally gave rise to various forms of semantically based paradigms (of the types sketched in Tables 5 and 6), the selection of a case-based system in western Andalusia stands as an historical puzzle.

The traditional solution to this puzzle has been to accept that this selection is the result of the retention of a feature brought mainly by settlers from some Castilian and non-Castilian northern varieties (Lapesa 1981; Penny 2000). Alternative approaches attribute the selection of the case-based system to leveling effects typical in koinéization (Tuten 2003), or to a combination of demographic and structural factors (Fernández-Ordóñez 1994: 123, 2001: 30–31, 55, 2012: 100). Most accounts of this change have emphasized the role of dialect contact among the L1 speakers of Ibero-Romance that settled in the newly reincorporated territories of western Andalusia. Little to no mention has been made of the potential role of other types of speakers, despite evidence that shows that contact among diverse groups of people was part of the linguistic ecology, a point to which we now turn.

5.2 The external ecology of MSIC clitics

Painting in broad brush strokes, the ecology that emerged immediately after the Castilian seizing of western Andalusia in the mid-13th century consisted of a diverse tapestry of ethnicities and linguistic varieties. The newly arrived settler contingents from the north, speaking varieties from the Hispano-Romance continuum (Castilian, Leonese, Galician, Aragonese, and Catalan) as well as Basque, came into contact with the local residents. Relevant for our analysis is the fact that the inhabitants of these communities had been anything but linguistically homogenous under Islamic ruling, including speakers of the Arabic continuum, Berbers, and, perhaps, remnant pockets of Romance Andalusí (Gallego 2004; Sayahi 2014). In addition, the Mediterranean commercial traffic attracted individuals from neighboring countries, including Portuguese, French, Italian, and English merchants, who set up shop in the largest cities (e.g., Ladero Quesada 1989: 284–289).

Notwithstanding the massive exodus to Africa and the Emirate of Granada (eastern Andalusia) following the Christian takeover, depopulation in western Andalusia was far from complete. An initial, yet failed attempt at Christian colonization facilitated numerous opportunities for language contact generated by the *pleitos* ('surrender agreements') for those pre-conquest communities who remained under Castile's ruling. Even after the 1264 *mudéjar* revolt severed the more lenient policies of coexistence, language contact in everyday life activities never stopped completely, but fluctuated in intensity according to the overall socio-demographic urban/rural settings (cf. Miteva 2017: 165–191 for the case of Seville). Another characteristic of the resettlement process was a considerable degree of demographic mobility, which, in turn, was the seed for looser social networks. Even though a large number of Muslim enclaves remained in the Guadalquivir valley in some rural areas (Echevarría 2006), there was frequent population redistribution through forced and voluntary evacuation, and at times massive re-location, particularly in urban centers or strategic places (González Jiménez & Montes Romero-Camacho 2001–2002: 56). In short, both groups of speakers, Arabic and Romance, in all of their linguistic diversity, came into contact and contributed to the pool of features in various degrees of intensity.

As in ECS, an additional important ecological trigger was the lack of a focused dialectal norm in the decades immediately after the resettlement. In the socially and culturally changing environment of the new southern border of Castile and the remaining Muslim territories, the operation of norm-enforcing mechanisms would have been challenging at best, as illustrated by Alfonso X's frustrated establishment of a *Studium Generale* in Seville in 1254 (Sánchez Herrero 1984; Pérez González & Arboleda Goldaracena 2018). As a result, many adult learners in this period must have been under few sociolinguistic pressures to conform to any prescriptive or supra-local norms. Instead, their priority must have been to gain communicative

competence within their respective social networks, as afforded by their cognitive skills. In many of these networks, potential targets within the Castilian pool included those offered by native and non-native speakers.

Based on this evidence, we can conclude that the incorporation of western Andalusia into the kingdom of Castile gave rise to a linguistic ecology of closely related varieties in contact with more distant ones. For instance, in post-resettlement Jerez de la Frontera, as much as 40% of the Christians were non-Castilians. In addition, Christians lived along several hundred residents of Muslim and Jewish descent, whether they were *forros* 'free', captives, or a byproduct of the slavery trade (González Jiménez 2008: 227–228). Similar socio-demographics could be found in Seville (Miteva 2017: 165–191). This heterogeneous ecology offered adult D2 and L2 speakers an opportunity to learn pronominal uses from a variety of native and non-native models.

While we agree with previous approaches that dialect mixture and structural differences played an important role in the reformulation of MSIC (cf. Tuten 2003; Fernández-Ordóñez 2012), we argue that a closer analysis of the ecologies in which the case-based paradigm emerged in southern Iberia suggests that language contact among L1 *and* L2 adults provided an additional source of linguistic input for the new generations. Notice, for example, that if we restrict the analysis to dialect contact contexts alone, there might be a substantial number of cases in which western Castilian settlers constituted the absolute majority (e.g., Carmona and Jerez de la Frontera). How could the case-determined paradigm, represented by the minority of colonizers, have been selected in those cases?[4] Additionally, the fact that the semantically based paradigm was not replicated in southern Iberia correlates with the notion that the local ecology must have been considerably different from that of north-central regions. Previous scholarship has considered this ecology only partially. In the following sections we advance a proposal to start filling this gap by considering the linguistic ecology of MSIC from a more holistic perspective, one that incorporates the heterogeneity of idiolects and targets in the learning process, the role of adult learners, and their interactions in social networks with few external norm-enforcing processes in place.

4. A reviewer hints at the possibility that a semantically based clitic system could be originally spread by a socially powerful minority. However, it is difficult to envision this scenario in the early stages of the MSIC (re)settlement. As pointed out by Tuten, "The military concerns of this frontier society were especially significant in determining social status […] [T]his widespread and rapid (re)assignment of social status to significant numbers of settlers represents a tremendous fluidity of social class at the beginning of the repopulation" (2003: 218). This means that social power structures, while present, were not kept exclusively by users of the semantically based paradigm. The more rigid stratification and expansion of social influence from larger urban areas occurred much later and from Andalusian-born members of society.

5.3 Acquisition of clitics in MSIC from a cross-linguistic perspective

The study of the evolution of these clitics on the basis of written sources, as commonly preferred in historical sociolinguistic approaches to change, is challenging. The documentary record from this period is not always complete or straightforward regarding clitics: for instance, Jewish or Muslim communities in post-resettlement Andalusia often noted Romance varieties in Hebrew or Arabic script (*aljamiado* texts), but in these texts, the vowel of clitics is typically not represented (Lapesa 1981: 262–264). Somewhat clearer clues can be found in other parts of the documentary record. For instance, there is potential evidence of early D2 accommodation to the case-based system by northern L1 speakers that moved to the south, judging from the lack of semantically based pronominal uses in 13th-century Castilian notary documents from Seville (Tuten 2003: 242–245). It is admittedly difficult to assess whether this absence is representative of actual dialectal uses at the time. As it turns out, recent research on documentation from Toledo and Madrid suggests that the case-determined paradigm may have been more common around this period in central Castile than previously surmised (cf. Sánchez-Prieto Borja & Vázquez Balonga 2018). Similarly, and equally unsurprisingly, it is possible to document lingering usage of the semantically based system in 13th- and 14th-century documentation from western Andalusia (Lapesa 1968: f.n. 8, Klein-Andreu 1999: 204–206, Kauffeld 2011: 76–77). These are not counter-examples to the reorganization toward the case-based paradigm. In fact, as we will see below, they all fall within what is expected in pervasive contact situations in which some adult learners maintain minority options and contribute to the heterogeneous feature pool (Trudgill 1986, 2004; Kerswill & Williams 2000).

To compensate for the dearth of evidence emerging from written sources, we may turn to cross-linguistic data on L2 acquisition of clitic systems. Unlike the case with sibilants, which are a common phonological segment in the vast majority of languages in the world, locating cross-linguistic parallels to the acquisition of Castilian Spanish clitics is more challenging: the use of clitics is far from universal and depends instead on the specific morphology of each language. Clitics are also highly variable syntactically in languages that have them, which complicates cross-linguistic comparisons. Despite these difficulties, there are several forms of evidence that can help us reconstruct how adult learners in this environment must have processed the various pronominal systems they received as input.

As explained in Section 5.1, speakers contributed with a plethora of options from three main types of systems based on case marking, semantic features, or a combination of both. It has often been observed that, despite the loss of case, most Romance languages, byproducts of contact-induced change themselves, still maintain case marking in the pronominal system. Interestingly, the rare exceptions

to this typological trend, in which clitic systems become sensitive to semantically based distinctions (e.g., animacy, countability), have remained restricted geographically (Pescarini 2016: 745) or have emerged in different contact-induced settings (Rosenger 2005; Fernández-Ordóñez 2012: 98, Yager et al. 2015). Part of the reason behind the diffusion of the semantically based system may indeed lie in the external ecology surrounding it. Tuten has convincingly argued that the north-central hybrid paradigm, despite its emergence in a koineizing ecology, was not necessarily a more simplified outcome (2003: 261–262). This suggests that the *transitional* paradigms, when transplanted to the south, might have presented some cognitive processing difficulties in comparison to other alternatives in the initial MSIC feature pool.

Acquisition research has shown that, even if the learner's native variety encodes similar features, the fact that they might be encoded differently morphologically or syntactically makes acquisition challenging (cf. Rossi, Kroll & Dussias 2014). This task is even more complex for L2 learners as the indirect object forms encode various functions (recipient, source, experiencer, beneficiary, possessor) and the input learners receive for direct object pronoun forms also encodes the grammatical functions of definite articles, pronouns (in the case of *lo*) and complementizers in relative pronoun constructions (*lo, la, los, las*). Faced with these complex (i.e., less learnable, see Section 3) options, it is now widely accepted that acquisition goes through a stage in which learners express one meaning with one form (One-to-One Principle) before they are capable of using multiple forms for the same meaning or multiple meanings for a single form (Multifunctionality Principle) as initially proposed by Andersen (1984, 1990). Functional approaches to adult L2 language learning have also convincingly argued in favor of the scaffolding that takes place in learners' acquisition of various ways to develop form-to-function and function-to-form mappings. In their detailed analysis of L2 acquisition of object pronouns in Spanish, Malovrh and Lee (2013) proposed four stages of acquisition, moving away from the one-to-one stage toward multifunctionality. This study revealed the acquisition of non-homophonous pronominal forms before the homophonous ones. That is, learners associated *lo* with its function as direct object before *la*, *las*, *los*, which overlap in form with the definite article. The study demonstrated also the saliency of animacy, a feature expressed in the overgeneralization of *le* in direct object functions (see also Zyzik 2006), and the use of some forms in detriment of others (e.g., masculine instead of feminine). An additional finding in Malovrh and Lee's research was the increased use of pronouns (vs. full noun phrases) as the restructuring of the interlanguage system developed the L2 learners' accurate use of clitics relative to their antecedent.

The complexity in adult acquisition of object pronouns is compounded by the need to interact with input at the semantic and pragmatic levels. The higher cognitive demands involved in the acquisition of features that entail *module interfaces*

has generated considerable interest among scholars (for definitions of cognitive processability, see Hawkins 2004 and Culicover 2013). Slabakova's Bottleneck Hypothesis (2008, 2009), for instance, argues for important differences in processing linguistic mismatches at the syntax-semantics interface. Similarly, Aboh's research in creole formation (2006, 2015) has demonstrated that language transfer in L2 acquisition is likely to produce non-native results in the case of interfaces, particularly at the syntax-semantics level. The syntax-pragmatics interface has also proven to be especially vulnerable in L2 acquisition. In this respect, Geeslin and Gudmestad (2008) have shown that English-speaking learners of Spanish follow syntactic rules when using subject pronouns, but in contexts where variation in subject expression is possible, learners do not reach the same level of accuracy. This is a clear indication that adult language learners experience a considerable degree of difficulty in acquiring multifunctionality in cases where there is no one-to-one correspondence between the linguistic inventories (see Sorace & Filiaci 2006 for similar findings in the processing of pronominal subjects at the syntax-discourse interface among English L1 learners of Italian).

In the context of object pronoun acquisition, where syntactic features interact solely or partly with semantic and pragmatic features, we should expect similar learning paths, operating principles, and learning constraints to be at work. Features such as [+/− animate], [+/− count], [+/− specific], [+/− definite], which rely on semantic and pragmatic information for their interpretation, can be expected to interact in learners' minds with retrieval of (and access to) morpho-syntactically derived information encoding. In language contact situations, this means that learners might opt for the same features to encode morpho-syntactic information, but they could encounter major difficulties in processing how these features interact with semantic ones, such as 'countability'. Thus, a feminine noun such as *miel* 'honey' could be interpreted as a [−count] semantically and learners would need to select between *la* and *lo*, both of which were potential inputs fed into the contact setting's feature pool in the MSIC case (see Table 6). In this context, current studies in language contact between Haitian Creole and Dominican Spanish have identified the overgeneralization of clitic expression in connection with the feature [+animate] (Ortiz-López & Guijarro-Fuentes 2009). Borgonovo et al. (2006) report on similar findings in Brazilian Portuguese L1 learners of Spanish who overgeneralized non-native-like use of overt object pronouns when the semantic interpretation of the referent was [-specific] (e.g., *¿Habrá gente mañana? – Sí, Ø / *la habrá* 'Will there be people tomorrow? – Yes, there will be'). Other analyses on use or omission of null pronouns in L2 acquisition have focused on learners' performance depending on the verb type and topic continuity in discourse (e.g., Zyzik 2008).

Finally, diachronic approaches to object pronoun usage have underscored the importance of similar features: animacy, specificity of the subject, aspect of the verb,

and argument structure (see García 1975). Flores Cervantes (2006) has proposed additional syntax-pragmatics interface effects involving the degree of transitivity of the event and pragmatic values added to the meaning of dative and accusative clitics (see also Fernández-Ordóñez 2001: 25). With respect to the degree of transitivity, it has been widely accepted that a number of semantic factors can influence the degree of transitivity. Some of these factors include, but are not limited to, the agentivity of the subject, the dinamicity and time, aspect, mood (or TAM) characteristics of the event, the degree of individuation of the object, and the degree with which the object is affected by the event. It has also been shown that morphological exponents of TAM can correlate with the use of the clitic pronoun. Thus, a perfective, real, telic and punctual event implies a higher degree of affectedness, which in turn correlates with a higher degree of probability to use the accusative. Conversely, an imperfective, virtual, atelic and durative event assumes a lower degree of affectednes and higher probability to use a clitic in the dative case (Flores Cervantes 2006: 689).

An example of the relationship between degree of transitivity, in the terms just outlined, and the use of clitic pronouns is included in (2) and (3). In both cases the verb *matar* 'to kill' is used with different object pronouns, depending on the perception of the event as more real or abstract. While the TAM features in (2) imply that the event has taken place and the action has been brought to effect, in (3) the TAM features are kept in the more subjective domain, with the implication that the object cannot be perceived as affected by the action.

(2) endereçaron todos [al marido] et **mataron lo**
(1335, *Conde Lucanor*, apud Flores Cervantes 2006: 690)
'they attacked the husband and killed him'

(3) que si por esto **le matasen**, que nunca él tan buen día viera
(1335, *Conde Lucanor*, apud Flores Cervantes 2006: 690)
'if they had killed him for it, it would be the best day of his life'

With respect to the pragmatic values added to the meaning of the dative and accusative clitic pronouns, Flores Cervantes has also shown the extension of the dative (*le*) to express personification of and/or respect to the object and the use of the accusative (*lo*) to demote an animate object to be valued as an inanimate object or to diminish the expression of respect. Flores Cervantes illustrates these pragmatic interpretations with examples such as those in (4) and (5), taken from the same text and used by the same person; the closest to a minimal pair. In (4) Cortés, the notorious conquistador, uses the accusative in referencing his service to Moctezuma, but he uses the dative, to show a higher degree of respect in reference to his service to the King in (5).

(4) y la hubo [quietud] todo el tiempo que yo tuve preso al dicho Mutezuma, porque él estaba muy a su placer y con todo su *servicio*, según en su casa **lo tenía**
(1519–26, Cortés, *Cartas de relación*, apud Flores Cervantes 2006: 705).
'all was quiet and remained so all the time I held Moctezuma prisoner, for he was very much at his ease and kept all this household with him as before.'

(5) Suplico a vuestra cesárea majestad reciba mi pequeño *servicio*, **teniéndole** en tanto cuanto la grandeza de mi voluntad para hacerle mayor, si pudiera merecer
(1519–26, Cortés, *Cartas de relación*; apud Flores Cervantes 2006: 705).
'I beseech your Caesarean Majesty to accept my small service, bearing in account the extent of my wish to do still greater ones if fortune permits.'

It is clear from these few examples that the interplay between syntax, semantics, and pragmatics was operational among users of the semantically based clitic pronoun systems in medieval Iberia. Despite the chronological difference between the time periods to which Examples (2)–(5) belong and the historical period with which we are concerned, it is safe to assume that 13th-century speakers of the semantically based clitic pronoun system engaged in similar language interactions with members of their new linguistic ecology in southern Iberia. We submit that the social and linguistic variability of this input, a byproduct of interfaces akin to the ones illustrated above, posed a considerable barrier to L2 learners.

In sum, there seems to be a direct correlation in adult learning between ease of processing of features and the likelihood that those features will be retained in L2 varieties. All other things being equal, if learners need to interpret these features using an interface of domains (e.g., syntax/semantics), they are less likely to learn them than if no such interfaces are involved. With these findings in mind, we now proceed to explore how adult learners must have contributed to the ecology that led to the emergence of the case-based system in MSIC.

5.4 Towards a new account of MSIC clitics: The role of adult language learning

The starting point for our alternative proposal about the emergence of a case-based paradigm clitic pronoun system in MSIC is the demographic rearrangement that took place between the 13th and the 15th centuries in the southern Iberian Peninsula, resulting in the formation of heterogeneous dialectal pools of competing features. These included a variety of clitic systems, similar to those still present in Ibero-Romance dialects.

Structurally, some options in this mixture were not favored in speakers' interactions. Cross-linguistically marked features, such as the extension of dative *le* over accusative *lo*, and the processing complexities involved in the use of features at the

interface between the syntax, semantic and/or pragmatics modules, proved to be difficult targets for native and non-native adult learners to attain. From a cognitive standpoint, speakers' interactions were led by general ease-of-processing tendencies as their preferred targets, which meant attrition of local complexities (e.g., acquisition of restrictions in the use of [+/– count], [+/– human], [+/– masculine]) in favor of options closer to one-to-one mappings between form and meaning, abundantly available to them as a shared feature in the input from the majority of users of the case-based paradigm.

These structural and cognitive factors interacted with the demographic composition of the resettlement society, which appears to have included non-native speakers, as well as a higher percentage of speakers of western Castilian varieties than in the resettlement of central Castilian areas as discussed in Sections 5.1 and 5.2. Recall that these Western varieties favored the case-based system: *lo(s)* and *la(s)* to mark accusative masculine and feminine singular (plural) referents, respectively as in *lo vi* 'I saw it/him', *la vi* 'I saw it/her'; and the use of *le(s)* to mark dative case in both masculine and feminine singular (plural), as in *le(s) pidieron ayuda* 'they ask him/her (them) for help.' It was this sociodemographic landscape, coupled with the lack of any group-enforcing target norm in the initial stages of the reconfiguration and the predominant loose-knit social networks in the growing urban centers (e.g., Seville), that contributed to the transmission and eventual acceptance of the newly formed MSIC case-based system.

As in the case of ECS sibilants, in the decades immediately following resettlement, adults did not fully accommodate to any single clitic system, let alone develop a single peer-group model. Therefore, we would expect even more variability in actual spoken language than we can glean from the available archival documentation. Additionally, the socio-political instability and repopulation lag in the decades following the *mudéjar* upheaval underscore the significant role played by both adults *and* children in the subsequent stages of the clitic pronoun system reconfiguration. It was in the minds of Andalusian-born children over the following decades that we can find support for the continuation of initial triggers set up during the prior adult learning stage. Adults set this stage by eroding complex-processing features of the heterogeneous mixture which, in the case of the unstressed clitic pronoun, resulted in the attrition of semantically based triggers. Children proceeded to further reorganize the feature pool by stabilizing the patterns of variation inherited from the previous stage and advancing towards the elimination of the semantically based features, while maintaining the most prominent, cross-linguistically common, typologically learnable features of case marking in the MSIC language ecology.

It should be noted that the two changes discussed in this paper are different not just in relation to the areas of grammar that they affected (phonology in the case of ECS sibilants vs. morphology, syntax, and semantics in the case of MSIC),

but also in terms of the effects of the change. The evolution of ECS sibilants arguably implied a simplification of the input (from several sibilant contrasts to one front sibilant), while the changes in MSIC clitics implied the selection of one of the options available in the L1 input and the preservation of specific distinctions (i.e., case-based contrasts).[5] This difference, however, does not invalidate our basic premise that different types of learners may converge upon specific options in contexts of pervasive sociolinguistic contact. We do not deny that environments characterized by frequent adult D2 and L2 learning may favor specific kinds of changes (see Trudgill 2011 for an articulation of this type of argument). However, we want to stress that learners always work with the primary linguistic data that they have access to in their respective community feature pools. In some cases, the distribution of forms in these pools might motivate learners to acquire the simplest option (i.e., the one with the least formal contrasts) present in their input, as happened in the case of ECS sibilants, or to generate new simpler alternatives. In other cases, however, the most learnable form may be one that still preserves some grammatical distinctions (e.g., MSIC clitics). In both cases, however, the same principle applies: if various types of learners were significantly represented in the sociolinguistic landscape of a particular environment, a socio-historically accurate reconstruction of language change should consider all of their various motivations to acquire specific alternatives.

6. Conclusion: Individuals as agents of language change

The above pages have offered a novel account of two instances of change in the history of Spanish: the simplification of the medieval Castilian sibilant system in ECS and the reformulation of 3rd-person clitics in MSIC. It has been argued that traditional accounts of these changes, which focus on the action of D2 learners, are not consistent with the available information about the sociodemographic embedding of these changes indicating that L2 learners were also an important part of these contact ecologies. We have argued that specific options in the input presented acquisitional advantages for both types of learners. Our accounts rely on the observed shared processes and effects of adult language learning by individuals of diverse backgrounds, situated alongside a continuum between D2 and L2 learning, in turn a function of post-critical period learning. By stressing the importance of both types of learning in these two sociohistorical ecologies, this account complements rather than invalidates previous literature on the importance of dialect contact in these changes.

5. We thank an anonymous reviewer for this observation.

We have also argued that this approach to language contact may aid in improving our understanding of specific cases of actuation of language change by helping us go beyond the traditional, albeit theoretically untenable, distinction between 'languages' and 'dialects'. Instead, the focus of our proposal is on the various ways in which ecological triggers like age of acquisition, typological distance between L1/D1 and L2/D2 features, and social opportunities for communication and learning may have prompted individuals in the contact setting to access and process a heterogeneous linguistic input. Ultimately, the learning of features by adults in these contact environments is the outcome of complex forms of interactions between linguistic, cognitive, and social factors, occurring at various points across a continuum of learnability, rather than as manifestations of discrete processes of 'dialect' vs. 'language' learning.

Before concluding, we would be remiss not to reiterate that the above proposal does not imply that adults were the only agents of language change in these two sociohistorical contexts. From an evolutionary point of view, this is a moot question: *all individuals* present in a communicative ecology participate in the negotiation and selection of language features that constitute the basis of language change at the population level as allowed by their respective age-based cognitive affordances – even though, as argued above, some of these features may be more competitive for demographic, sociolinguistic, and acquisitional reasons. More specifically, young children also play a key role in the emergence of new language varieties by filtering and systematizing the variable input received from adults, as captured in DeGraff's *cascade effect* (1999, 2009). Note that the innovative role of young children envisioned in DeGraff's and the present proposal is markedly different from the generative argument whereby children actuate changes in community grammars by reanalyzing the parameters present in adult input (Lightfoot 1998; Kroch 2005) and rightly critiqued by Croft (2000: 44–49), among others. In a cascading approach, children are able to acquire adult targets when these are categorical or systematically variable. But in situations of pervasive contact, the input that children receive can be very unsystematic linguistically and socially, i.e., there is not one clear, single target to acquire. In these cases, their choices might result in more conspicuous departures from adult systems than those observed in more stable sociolinguistic environments, as proven by the literature on situations of koinéization, creolization, and the emergence of new sign languages (Roberts 2000; Senghas & Coppola 2001; Trudgill 2004, 2011). But even in these cases, children are still necessarily constructing their grammars on the basis of the input provided by older speakers, however unsystematic this input might be. Understanding how adults of varied provenance negotiate their heterogeneous input thus constitutes a key part of an ecologically informed, sociohistorically situated approach to language change.

References

Aboh, Enoch Oladé. 2006. The role of the syntax-semantics interface in language transfer. In Claire Lefebvre, Lydia White & Christine Jourdan (eds.), *L2 acquisition and creole genesis: Dialogues*, 221–252. Amsterdam: John Benjamins. https://doi.org/10.1075/lald.42.14abo

Aboh, Enoch Oladé. 2015. *The emergence of hybrid grammars: Language contact and change*. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9781139024167

Alonso, Amado. 1967. *De la pronunciación medieval a la moderna en español*. Vol. 1. Madrid: Gredos.

Alonso, Amado. 1969. *De la pronunciación medieval a la moderna en español*. Vol. 2. Madrid: Gredos.

Andersen, Roger 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

Andersen, Roger W. 1990. Models, processes, principles and strategies: Second language acquisition inside and outside the classroom. In Bill VanPatten & James F. Lee (eds.), *Second language acquisition-foreign language learning*, 45–78. Clevedon, UK: Multingual Matters.

Auer, Peter, Brigit Barden & Beate Grosskopf. 1998. Subjective and objective parameters determining 'salience' in long-term dialect accommodation. *Journal of Sociolinguistics* 2. 163–187. https://doi.org/10.1111/1467-9481.00039

Auer, Anita, Catharina Peersman, Simon Pickl, Gijsbert Rutten & Rik Vosters. 2015. Historical sociolinguistics: The field and its future. *Journal of Historical Sociolinguistics* 1. 1–12. https://doi.org/10.1515/jhsl-2015-0001

Birdsong, David. 1999. Introduction: Whys and why nots of the Critical Period Hypothesis for second language acquisition. In David Birdsong (ed.), *Second language acquisition and the Critical Period Hypothesis*, 1–22. Mahwah, NJ: Lawrence Erlbaum Associates. https://doi.org/10.4324/9781410601667

Blythe, Richard & William Croft. 2009. The speech community in evolutionary language dynamics. In Nick Ellis & Diane Larsen-Freeman (eds.), *Language as a complex adaptive system*, 47–63. Oxford: Blackwell.

Borgonovo, Claudia, Joyce Bruhn de Garavito, Pedro Guijarro-Fuentes, Philippe Prévost & Elena Valenzuela. 2006. Specificity in Spanish: The syntax/semantics interface in SLA. *Eurosla Yearbook* 6. 57–78. https://doi.org/10.1075/eurosla.6.06bor

Boyd-Bowman, Peter. 1976. Patterns of Spanish emigration to the Indies until 1600. *The Hispanic American Historical Review* 56. 580–604. https://doi.org/10.1215/00182168-56.4.580

Brain, Cecilia. 2010. Aprendizaje de lenguas indígenas por parte de españoles en Nueva España en los primeros cien años después de la conquista. *Colonial Latin American Review* 19. 279–300. https://doi.org/10.1080/10609164.2010.493686

Bybee, Joan. 2010. *Language, usage and cognition*. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511750526

Cano Aguilar, Rafael. 2004. Cambios en la fonología del español durante los siglos XVI y XVII. In Rafael Cano Aguilar (ed.), *Historia de la lengua española*, 593–612. Barcelona: Ariel.

Chambers, J. K. 1992. Dialect acquisition. *Language* 68. 673–705. https://doi.org/10.1353/lan.1992.0060

Chambers, J. K. & Peter Trudgill. 1998. *Dialectology*. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511805103

Chiu, Chenhao. 2010. Attentional weighting of Polish and Taiwanese Mandarin sibilant perception. *The Proceedings of the 2010 Canadian Linguistics Association Annual Conference*, http://homes.chass.utoronto.ca/~cla-acl/actes2010/actes2010.html. (11 January, 2016).

Clements, Joseph Clancy. 2009. *The linguistic legacy of Spanish and Portuguese: Colonial expansion and language change*. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511576171

Coetsem, Frans van. 2000. *A general and unified theory of the transmission process in language contact*. Heidelberg: Winter.

Company Company, Concepción. 1995. Cantidad vs. cualidad en el contacto de lenguas. Una incursión metodológica en los posesivos redundantes del español americano. *Nueva Revista de Filología Hispánica* 43. 305–339. https://doi.org/10.24201/nrfh.v43i2.1880

Croft, William. 2000. *Explaining language change: An evolutionary approach*. New York: Longman.

Croft, William. 2003. *Typology and universals*. Cambridge: Cambridge University Press.

Culicover, Peter W. 2013. *Grammar and complexity: Language at the intersection of competence and performance*. Oxford: Oxford University Press.

DeGraff, Michel. 1999. Creolization, language change, and language acquisition: An epilogue. In DeGraff, Michel (ed.), *Language creation and language change. Creolization, diachrony, and development*, 473–544. Cambridge, MA: MIT Press.

DeGraff, Michel. 2009. Language acquisition in creolization and, thus, language change: Some cartesian-uniformitarian boundary conditions. *Language and Linguistics Compass* 3/4. 888–971. https://doi.org/10.1111/j.1749-818X.2009.00135.x

Echevarría, Ana. 2006. La "mayoría" mudéjar en León y Castilla: legislación real y distribución de la población (siglos XI-XIII). *En la España Medieval* 29. 7–30.

Eckert, Penelope. 2000. *Language variation as social practice: The linguistic construction of identity in Belten High*. Oxford: Blackwell.

Erker, Daniel. 2017. The limits of named language varieties and the role of social salience in dialectal contact: The case of Spanish in the United States. *Language and Linguistics Compass* 11. https://doi.org/10.1111/lnc3.12232.

Ellis, Rod. 2008. *The study of second language acquisition*. Oxford: Oxford University Press.

Escobar, Ana María. 2001. Contact features in Colonial Peruvian Spanish. *International Journal of the Sociology of Language* 149. 79–93.

Fernández-Ordóñez, Inés. 1994. Isoglosas internas del castellano. El sistema referencial del pronombre átono de tercera persona. *Revista de Filología Española* 74. 71–125. https://doi.org/10.3989/rfe.1994.v74.i1/2.499

Fernández-Ordóñez, Inés. 1999. Leísmo, laísmo y loísmo. In Ignacio Bosque & Violeta Demonte (eds.), *Gramática descriptiva de la lengua española* 1, 1317–1397. Madrid: Espasa Calpe.

Fernández-Ordóñez, Inés. 2001. Hacia una dialectología histórica: reflexiones sobre la historia del leísmo, el laísmo y el loísmo. *Boletín de la Real Academia Española* 81. 389–464.

Fernández-Ordóñez, Inés. 2012. Dialect areas and linguistic change. Pronominal paradigms in Ibero-Romance dialects from a cross-linguistic and social typology. In Gunther De Vogelaer & Guido Seiler (eds.), *The dialect laboratory. Dialects as a testing ground for theories of language change*, 73–106. Amsterdam: John Benjamins. https://doi.org/10.1075/slcs.128.04fer

Flores Cervantes, Marcela. 2006. Leísmo, laísmo y loísmo. In Concepción Company Company (ed.), *Sintaxis histórica de la lengua española. Primera parte: La frase verbal*, 1. 669–749. Mexico City: Universidad Nacional Autónoma de México & Fondo de Cultura Económica.

Fontanella de Weinberg, María Beatriz. 1992. Nuevas perspectivas en el estudio de la conformación del español americano. *Hispanic Linguistics* 4. 275–299.

Frago Gracia, Juan Antonio. 1999. *Historia del español de América: Textos y contextos*. Madrid: Gredos.

Gallego, María Ángeles. 2004. The languages of Medieval Iberia and their religious dimension. *Medieval Encounters* 9. 107–139. https://doi.org/10.1163/157006703322576556

García, Érica C. 1975. *The role of theory in linguistic analysis: The Spanish pronoun system*. Amsterdam: North Holland.

Geeslin, Kimberly. 2011. The acquisition of variation in second language Spanish: How to identify and catch a moving target. In Manuel Díaz-Campos (ed.), *The handbook of Hispanic sociolinguistics*, 303–319. Oxford: Blackwell. https://doi.org/10.1002/9781444393446.ch15

Geeslin, Kimberly. 2018. Variable structures and sociolinguistic variation. In Paul Malovrh & Alessandro Benati (eds.). *The handbook of advanced proficiency in second language acquisition*, 547–565. Oxford: Wiley. https://doi.org/10.1002/9781119261650.ch28

Geeslin, Kimberly L. & Aarnes Gudmestad. 2008. Variable subject expression in second-language Spanish: A comparison of native and non-native speakers. In Melissa Bowles, Rebecca Foote, Silvia Perpiñán & Rakesh Bhatt (eds.), *Selected proceedings of the 2007 Second Language Research Forum*, 69–85. Somerville, MA: Cascadilla Proceedings Project.

Giacalone Ramat, Ana. 2003. *Typology and second language acquisition*. Berlin: De Gruyter.

Gonzalbo Aizpuru, Pilar. 2001. *Educación y colonización en la Nueva España 1521–1821*. Mexico City: Universidad Pedagógica Nacional.

González Jiménez, Manuel. 2008. *La repoblación del reino de Sevilla en el siglo XIII*. Granada: Universidad de Granada.

González Jiménez, Manuel & Isabel Montes Romero-Camacho. 2001–2002. Los mudéjares andaluces (siglos XIII–XV). Aproximación al estado de la cuestión y propuesta de un modelo teórico. *Revista d'Història Medieval* 12. 47–78.

González Ollé, Fernando. 1996. La precaria instalación del español en la América virreinal. *Anuario de lingüística hispánica* 12/13. 327–360.

Granda, Germán de. 1994. *Español de América, español de África y hablas criollas hispánicas: cambios, contactos y contextos*. Madrid: Gredos.

Hawkins, John A. 2004. *Efficiency and complexity in grammars*. Oxford: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199252695.001.0001

Hickey, Raymond. 2003. How do dialects get the features they have? On the process of new dialect formation. In Raymond Hickey (ed.), *Motives for language change*, 213–239. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511486937.014

Hudson Kam, Carla & Elissa Newport. 2005. Regularizing unpredictable variation: The roles of adult and child learners in language formation and change. *Language Learning and Development* 1. 151–195. https://doi.org/10.1080/15475441.2005.9684215

Jahr, Ernst Håkon Jahr. 1999. Sociolinguistics in historical language contact: The Scandinavian languages and Low German during the Hanseatic period. In Ernst Håkon Jahr (ed.), *Language change: Advances in historical sociolinguistics*, 119–139. Berlin: Mouton de Gruyter. https://doi.org/10.1515/9783110807653.119

Johanson, Lars. 2002. Contact-induced change in code-copying framework. In Mary Jones & Edith Esch (eds.), *Language change: The interplay of internal, external and extra-linguistic factors*, 285–313. Berlin: Mouton de Gruyter. https://doi.org/10.1515/9783110892598.285

Karttunen, Frances & James Lockhart. 1976. *Nahuatl in the Middle Years: Language contact phenomena in texts of the colonial period*. Berkeley: University of California Press

Kauffeld, Cynthia. 2011. *Andalusian Spanish: A linguistic study of 14th- and 15th-century texts from Sevilla and Córdoba*. New York: Hispanic Seminary of Medieval Studies.

Kerswill, Paul. 2002. *Koineization and accommodation*. In J. K. Chambers, Peter Trudgill & Natalie Schilling-Estes (Eds.), *The handbook of language variation and change*, 669–702. Malden, MA: Blackwell.

Kerswill, Paul. 2013. Identity, ethnicity and place: The construction of youth language in London. In Peter Auer, Martin Hilpert, Anja Stukenbrock & Benedikt Szmrecsanyi (eds.), *Space in language and linguistics: Geographical, interactional, and cognitive perspectives*, 128–164. Amsterdam: De Gruyter. https://doi.org/10.1515/9783110312027.128

Kerswill, Paul & Ann Williams. 2000. Creating a new town koine: Children and language change in Milton Keynes. *Language in Society* 29. 65–115. https://doi.org/10.1017/S0047404500001020

Klee, Carol & Andrew Lynch. 2009. *El español en contacto con otras lenguas*. Washington, DC: Georgetown University Press.

Klein-Andreu, Flora. 1999. Variación actual y reinterpretación histórica: *le/s, la/s, lo/s* en Castilla. In María José Serrano (ed.), *Estudios de variación sintáctica*, 197–220. Frankfurt: Vervuert.

Kretschmar, William. 2015. *Language and complex systems*. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9781316179017

Kroch, Anthony. 2005. Modeling language change and language acquisition. Online publication, http://www.ling.upenn.edu/~kroch/courses/lx556/lsa-forum.pdf. (5 December, 2019).

Labov, William. 1994. *Principles of linguistic change: Internal factors*. Oxford: Blackwell.

Labov, William. 2001. *Principles of linguistic change: Social factors*. Oxford: Blackwell.

Labov, William. 2007. Transmission and diffusion. *Language* 83. 344–387. https://doi.org/10.1353/lan.2007.0082

Ladefoged, Peter & Ian Maddieson. 1996. *The sounds of the world's languages*. Oxford: Blackwell.

Ladero Quesada, Miguel Ángel. 1989. Los genoveses en Sevilla y su región (siglos XIII–XVI): elementos de permanencia y arraigo. In Miguel Ángel Ladero Quesada (ed.), *Los mudéjares de Castilla y otros estudios de historia medieval andaluza*, 283–312. Granada: Universidad de Granada.

Lapesa, Rafael. 1968. Sobre los orígenes y evolución del leísmo, laísmo y loísmo. In Kurt Baldinger (ed.), *Festschrift Walther von Wartburg zum 80. Geburtstag*, 523–551. Tübingen: Max Niemeyer.

Lapesa, Rafael. 1981. *Historia de la lengua española*. 8th edn. Madrid: Gredos.

Larraza, Saioa. 2009. Acquisition of phonology and Spanish-Basque bilinguals' phonological systems. *Actes du colloque AcquisiLyon 09, Lyon, 3 et 4 décembre 2009*, http://ohll.ish-lyon.cnrs.fr/Colloques/AcquisiLyon/pageweb/Fichier/Larraza.pdf. (5 May, 2017).

Larsen-Freeman, Diane. 2007. Reflecting on the cognitive-social debate in second language acquisition. *The Modern Language Journal* 91. 773–787. https://doi.org/10.1111/j.1540-4781.2007.00668.x

Lenneberg, Eric. 1967. *Biological foundations of language*. New York: John Wiley & Sons. https://doi.org/10.1080/21548331.1967.11707799

Lightfoot, David. 1998. *The development of language: Acquisition, change, and evolution*. Oxford: Blackwell.

Lipski, John. 2005. *A history of Afro-Hispanic language: Five centuries, five continents*. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511627811

Lipski, John. 2014. The many facets of Spanish dialect diversification in Latin America. In Salikoko Mufwene, (ed.), *Iberian imperialism and language evolution in Latin America*, 38–75. Chicago: University of Chicago Press. https://doi.org/10.7208/chicago/9780226125671.003.0002

Lockhart, James. 1992. *The Nahuas after the conquest: A social and cultural history of the Indians of central Mexico, sixteenth through eighteenth centuries*. Stanford, CA: Stanford University Press.

Maddieson, Ian. 1984. *Patterns of sounds*. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511753459

Malovrh, Paul A. & James F. Lee. 2013. The developmental dimension in instructed second language learning. *The L2 acquisition of object pronouns in Spanish*. London: Bloomsbury Academic.

Marecka, Mart, Magdalena Wrembel, Agnieszka Otwinowska-Kasztelanic & Dariusz Zembrzuski. 2015. Phonological development in the home language among early Polish-English bilinguals. In The Scottish Consortium for ICPhS 2015 (eds.), *Proceedings of the 18th International Congress of Phonetic Sciences*, https://www.internationalphoneticassociation.org/icphs-proceedings/ICPhS2015/Papers/ICPHS0714.pdf. (11 January, 2016).

Martín Butragueño, Pedro. 2010. El proceso de urbanización: Consecuencias lingüísticas. In Rebeca Barriga Villanueva & Pedro Martín Butragueño (eds.), *Historia sociolingüística de México: Vol. 2: México contemporáneo*, 997–1093. Mexico City: El Colegio de México.

Menéndez Pidal, Ramón. 1962. Sevilla frente a Madrid. *Estructuralismo e historia* 3. 99–165.

Milroy, Lesley & James Milroy. 1992. Social network and social class: Toward an integrated sociolinguistic model. *Language in Society* 21. 1–26. https://doi.org/10.1017/S0047404500015013

Mitchell, Rosamond, Florence Myles & Emma Marsden (eds.). 2013. *Second language learning theories*. London: Routledge.

Miteva, Yuliya Radoslavova. 2017. Historia social de la lengua de los mudéjares. *La asimilación lingüística de la minoría islámica en la Península Ibérica en la etapa medieval*. San Cristóbal de La Laguna: Universidad de la Laguna dissertation.

Mufwene, Salikoko. 2001. *The ecology of language evolution*. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511612862

Mufwene, Salikoko. 2008. *Language evolution: Contact, competition, and change*. London: Continuum.

Negrão, Esmeralda & Evani Viotti. 2014. Brazilian Portuguese as a transatlantic language: Agents of linguistic contact. *InterDISCIPLINARY: Journal of Portuguese Diaspora Studies* 3. 135–154.

Orozco, Rafael. 2018. *Spanish in Colombia and New York City: Language contact meets dialectal convergence*. Amsterdam: John Benjamins. https://doi.org/10.1075/impact.46

Ortiz-López, Luis A. & Pedro Guijarro-Fuentes. 2009. Interfaz sintáctica-semántica en los objetos directos: el español y el criollo haitiano. In Joseph Collentine, Maryellen García, Barbara Lafford, and Francisco Marcos Marín (eds.), *Selected proceedings of the 11th Hispanic Linguistics Symposium*, 268–281. Somerville, MA: Cascadilla Proceedings Project.

Otheguy, Ricardo & Ana Celia Zentella. 2012. *Spanish in New York: Language contact, dialect leveling and structural continuity*. Oxford: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199737406.001.0001

Parker, Kate. 1989. Learnability theory and the acquisition of syntax. *University of Hawai'i Working Papers in ESL* 8. 49–78.

Parodi, Claudia. 1995. *Orígenes del español americano*. Mexico City: Universidad Nacional Autónoma de México.

Penny, Ralph. 2000. *Variation and change in Spanish*. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9781139164566

Penny, Ralph. 2004. Evolución lingüística en la Baja Edad Media: evoluciones en el plano fonético. In Rafael Cano Aguilar (ed.), *Historia de la lengua española*, 593–612. Barcelona: Ariel.

Pérez González, Silvia María & Juan Carlos Arboleda Goldaracena. 2018. El acceso a la cultura y la formación intelectual en Andalucía durante la Baja Edad Media. *Studia Historica. Historia Medieval* 36. 61–81. https://doi.org/10.14201/shhme20183626181

Pescarini, Diego. 2016. Clitic pronominal systems: Morphophonology. In Adam Ledgeway & Martin Maiden (eds.), *The Oxford Guide to the Romance Languages*, 742–757. Oxford: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199677108.003.0045

Plans, Antonio Salvador. 2004. Los lenguajes 'especiales' y de las minorías en el Siglo de Oro. In Rafael Cano Aguilar (ed.), *Historia de la lengua española*, 771–797. Barcelona: Ariel.

Roberts, Sarah. 2000. Nativization and the genesis of Hawaiian Creole. In John McWorther (ed.), *Language change and language contact in pidgins and creoles*, 257–300. Amsterdam: John Benjamins. https://doi.org/10.1075/cll.21.10rob

Rosenger, Peter. 2005. Dialect convergence in the German language islands (*Sprachinseln*). In Peter Auer, Frans Hinskens & Paul Kerswill (eds.), *Dialect change: Convergence and divergence in European languages*, 221–235. Cambridge: Cambridge University Press.

Rossi, Eleonora, Judith F. Kroll & Paola E. Dussias. 2014. Clitic pronouns reveal the time course of processing gender and number in a second language. *Neuropsychologia* 62. 11–25. https://doi.org/10.1016/j.neuropsychologia.2014.07.002

Salas, Alberto M. 2000. El mestizaje en la conquista de América. In Georges Dyby (ed.), *Historia de las mujeres: Tomo III, Del Renacimiento a la Edad Moderna*, 539–556. Madrid: Santillana.

Sánchez Herrero, José. 1984. Centros de enseñanza y estudiantes de Sevilla durante los siglos XIII al XV. *En la España Medieval* 5. 875–898.

Sánchez-Albornoz, Nicolás. 1989. Las migraciones anteriores al siglo XIX. In Birgitta Leander (ed.), *Europa, Asia y África en América Latina y el Caribe: migraciones "libres" en los siglos XIX y XX*, 61–88. Mexico City: UNESCO/Siglo Veintiuno Editores.

Sánchez-Prieto Borja, Pedro & Delfina Vázquez Balonga. 2018. Toledo frente a Madrid en la conformación del español moderno: el sistema pronominal átono. *Revista de Filología Española* 98(1). 185–215. https://doi.org/10.3989/rfe.2018.08

Sanz-Sánchez, Israel. 2019. Documenting feature pools in language expansion situations: Sibilants in Early Colonial Latin American Spanish. *Transactions of the Philological Society* 117. 199–233. https://doi.org/10.1111/1467-968X.12154

Sayahi, Lotfi. 2014. *Diglossia and language contact: language variation and change in North Africa*. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9781139035576

Schneider, Edgar. 2009. *Postcolonial English: Varieties around the world*. Cambridge: Cambridge University Press.

Schwaller, Robert. 2016. *Géneros de gente in Early Colonial Mexico: Defining racial difference*. Norman: University of Oklahoma Press.

Senghas, Ann & Marie Coppola. 2001. Children creating language: How Nicaraguan Sign Language acquired a spatial grammar. *Psychological Science* 12. 323–328. https://doi.org/10.1111/1467-9280.00359

Sharma, Devyani & Lavanya Sankaran. 2011. Cognitive and social forces in dialect shift: Gradual change in London Asian speech. *Language Variation and Change* 23. 399–428. https://doi.org/10.1017/S0954394511000159

Shih, Ya-ting. 2012. *Taiwanese-Guoyu bilingual children and adults' sibilant fricative production patterns*. Columbus, OH: Ohio State University dissertation.

Shoemaker, Elleanor. 2016. The development of perceptual sensitivity to Polish sibilants at first exposure. *Proceedings of AMP 2014*, http://journals.linguisticsociety.org/proceedings/index.php/amphonology/article/view/3770. (11 January, 2017).

Siegel, Jeff. 2010. *Second dialect acquisition*. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511777820

Silva-Corvalán, Carmen. 1994. *Language contact and change: Spanish in Los Angeles*. Oxford: Oxford University Press.

Slabakova, Roumyana. 2008. *Meaning in the second language*. Berlin: Mouton de Gruyter. https://doi.org/10.1515/9783110211511

Slabakova, Roumyana. 2009. What is easy and what is hard to acquire in a second language? In Melissa Bowles, Tania Ionin, Silvina Montrul & Annie Tremblay (eds.), *Proceedings of the 10th Generative Approaches to Second Language Acquisition Conference (GASLA 2009)*, 280–294. Somerville, MA: Cascadilla Proceedings Project.

Sorace, Antonella & Francesca Filiaci. 2006. Anaphora resolution in near-native speakers of Italian. *Second Language Research* 22. 339–368. https://doi.org/10.1191/0267658306sr271oa

Thomason, Sarah. 2001. *Language contact: An introduction*. Washington, DC: Georgetown University Press.

Thomason, Sarah & Terrence Kaufman. 1988. *Language contact, creolization, and genetic linguistics*. Berkeley: University of California Press.

Trask, Larry. 1997. *A history of Basque*. London: Routledge.

Trudgill, Peter. 1986. *Dialects in contact*. Oxford: Blackwell.

Trudgill, Peter. 2004. *New-dialect formation: The inevitability of colonial Englishes*. Edinburgh: Edinburgh University Press.

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

Tuten, Donald. 2003. *Koinéization in Medieval Spanish*. Berlin: De Gruyter. https://doi.org/10.1515/9783110901269

VanPatten, Bill & Jessica Williams. 2020. *Theories in second language acquisition: An introduction*. 3rd edn. London: Routledge. https://doi.org/10.4324/9780429503986

Vázquez Balonga, Delfina & Pedro Sánchez-Prieto Borja. 2015. ¿Seseo en el centro peninsular? *Revista de Historia de la Lengua Española* 10. 201–207.

Villena-Ponsoda, Juan Andrés. 2005. How similar are people who speak alike? An interpretive way of using social networks in social dialectology research. In Peter Auer, Frans Hinskens & Paul Kerswill (eds.), *Dialect change: Convergence and divergence in European languages*, 303–334. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511486623.014

Weinreich, Uriel, William Labov & Marvin Herzog. 1968. Empirical foundations for a theory of language change. In Winfred Lehmann & Yakov Malkiel (eds.), *Directions for historical linguistics*, 95–195. Austin: University of Texas Press.

Winford, Donald. 2007. Some issues in the study of language contact. *Journal of Language Contact – THEMA* 1. 22–40. https://doi.org/10.1163/000000007792548288

Yager, Lisa, Nora Hellmold, Hyoun-A Joo, Michael T. Putnam, Eleonora Rossi, Catherine Stafford & Joseph Salmons. 2015. New structural patterns in moribund grammar: Case marking in heritage German. *Frontiers in Psychology*, https://www.frontiersin.org/articles/10.3389/fpsyg.2015.01716/full. https://doi.org/10.3389/fpsyg.2015.01716

Zygis, Marzena. 2003. Phonetic and phonological aspects of Slavic sibilant fricatives. *ZAS Papers in Linguistics* 32(1). 175–212.

Zyzik, Eve. 2006. Learners' overgeneralization of dative clitics to accusative contexts: Evidence for prototype effects in SLA. In Carol A. Klee & Timothy L. Face (eds.), *Selected proceedings of the 7th conference on the acquisition of Spanish and Portuguese as first and second languages*, 122–134. Somerville, MA: Cascadilla Proceedings Project.

Zyzik, Eve. 2008. Null objects in second language acquisition: Grammatical versus performance models. *Second Language Research* 24(1). 65–110. https://doi.org/10.1177/0267658307082982

SECTION II

Socio-historical varieties in isolation and contact

Searching for the sociolinguistic history of Afro-Panamanian *Congo* speech

John M. Lipski
The Pennsylvania State University

Among the surviving Afro-Hispanic linguistic manifestations, one of the most difficult to trace historically is the speech of the *Congos* of Panama's Caribbean coast, who maintain a series of folkloric manifestations occurring during Carnival season that includes a special language. According to oral tradition, *Congo* speech was devised among captive and maroon Africans in colonial Panama as a means of hiding their speech from their colonial masters. Putting together the contemporary variation in *Congo* speech and what diachronic developments can be extrapolated, a complex picture emerges that cannot be easily resolved with the notion that this dialect developed exclusively as a cryptolect in contact with Spanish colonists. The present study offers a plausible scenario, based on synchronic variation and available historical documentation.

Keywords: Panama, Congos, Afro-Hispanic language, slavery, maroons, cryptolect, creolization

1. Introduction

Tracing the sociolinguistic history of ethnolects and contact-induced languages is a challenging task, since in most cases the time depth of reliable information is quite shallow, and matters may be exacerbated by popular stereotypes and parodies as well as simply inaccurate portrayals. This is nowhere more notable than in the case of language varieties emerging from the African diaspora in the Americas. For example, Afro-Atlantic creole languages have been dismissed in popular opinion as "broken" or improper versions of their historical lexifiers, and scholars struggle to separate fact from fiction in reconstructing the historical trajectories (cf. Lipski 2005).

Among the surviving Afro-Hispanic linguistic manifestations, one of the most difficult to trace is the speech of the *Congos* of Panama. The following sections provide an overview of *Congo* speech, including popular accounts as well as more detailed studies, and offer a triangulation of linguistic and socio-historical factors as the most plausible scenario for the rise and spread of this phenomenon.

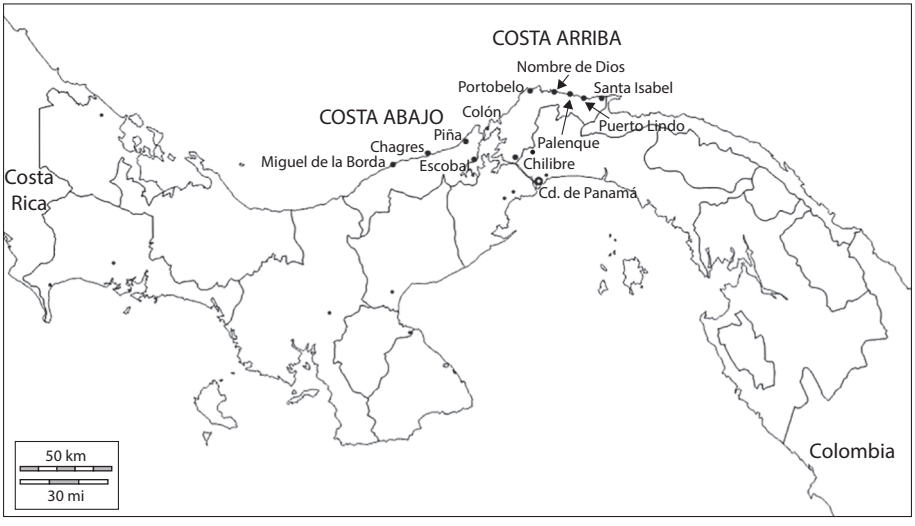
https://doi.org/10.1075/ahs.12.c06lip

2. Afro-Panamanian *Congo* speech and culture

One of the most unique cultural manifestations of Panamanian society is the culture of the *Congos* of the Caribbean coast. The *Congos* represent an Afro-colonial tradition, whose participants engage in dramatic representations and speech acts that provide unique manifestations of ethnic identity and cultural resistance through language and theatre. *Congo* culture predominates among Panama's Afro-Colonial population along the Caribbean coast as well as in some inland locations. The *Congo* ceremonies, which take place during Carnival season each year, combine a historical component (the reference to the history of enslaved Africans in Panama), and an African aspect (the dancing and drumming and the costumes). The social, musical, and theatrical aspects of contemporary *Congo* culture have been analyzed by Alexander (2006), Alexander Craft (2008a, 2008b, 2015), Bettelheim (2004), De la Rosa Sánchez (1988), Laribe (1968, 1969), Luque de Pérez (2001), Smith (1975, 1994), and Tejeira Jaén (1974), among others.

A key cultural manifestation during *Congo* season is the use of a special dialect mode, distinct from regional popular Spanish. *Congo* speech is all but unintelligible to the uninitiated, and *Congo* dramatic personages use this speech in addressing members of the community during dances and ritualized aggression. *Congo* speech and the accompanying rituals are found from Miguel de la Borda on the westernmost portion of the Costa Arriba, to the west of the Panama Canal, to Santa Isabel, the easternmost *Congo* community on the Costa Arriba, to the east of the canal. *Congo* language and culture vary between the Costa Abajo and the Costa Arriba, and this variation is crucial to reconstructing the social and linguistic history of the *Congos*. Map 1 shows the approximate location of the *Congo* communities.

According to oral history, the *Congo* rituals represent a re-enactment of historical events affecting enslaved Africans in the colonial port of Portobelo, through which both South American treasure and Spanish merchandise passed. There is however no consensus among *Congos* as to the origins of the special language, nor have historians addressed the issue. In interviews conducted by the author in the 1980s, some *Congos* asserted that *Congo* speech was a continuation of the attempts to speak Spanish by African-born captives (known as *bozales*), but given the deliberately distorted nature of much of the language, a strict socio-historical explanation cannot be taken at face value. Other commonly-heard assertions about *Congo* speech refer only to talking "backwards" (*hablar al revés / revesinas*) in "una mezcla de francés, portugués y castellano" [a mixture of French, Portuguese, and Spanish] (e.g., Racero 2005, quoting the historian Apolonio Acosta and former INAC director Krishna Menacho). In fact, *Congo* language has no French, Portuguese, or other xenoglossic elements, but the deviations from canonical Spanish – including some "backwards" talk – together with the largely improvised nature of contemporary



Map 1. Approximate location of *Congo* communities in Panama

Congo speech, result in an exotic and barely intelligible aura that creates the impression of a jumble of languages.[1] Samples of *Congo* language from several regions and time periods are found in the Appendix. As of 2018, UNESCO has included Afro-Panamanian *Congo* culture in the Representative List of the Intangible Cultural Heritage of Humanity, but with no mention of the special language used during *Congo* events (https://ich.unesco.org/en/RL/ritual-and-festive-expressions-of-the-congo-culture-01383).

The most significant obstacle in reconstructing the sociolinguistic history of *Congo* speech is the absence of any known documentation of *Congos* dating back to the colonial slaveholding period.[2] In fact, the first mentions of *Congos* come in the second half of the 20th century, with reference to earlier decades of the same century. As a result, any historical reconstruction must depend crucially on known facts about the presence of Africans and Afro-descendants in early colonial Panama, together with a comparison with other Afro-Iberian linguistic phenomena, past and present. These caveats circumscribe the hypotheses presented in the present essay.

---

1. Despite the many "African"-sounding words that result from *Congos*' distortion of Spanish words, there is only one lexical item that has so far defied etymological searches and which may be of African origin: *mojongo* (occasionally *mojonga*) meaning 'woman, wife'.

2. Alexander Craft (2008b: 23) states that "In Colonial times, the Congo carnival performances were done under the watchful gaze of enslavers who saw the reversals of clothing, heard the Congo dialect and participated in the joke not knowing that they were its objects." While this is a likely scenario, there is no documentation of *Congo* language or ceremonies from this time period.

The earliest published descriptions of the Panamanian *Congos* e.g. Béliz (1959) did not dwell on *Congo* speech, with the first known explicit account found in Franceschi (1960). This author analyzed *Congo* speech as stemming from acts of resistance to enslavement during the colonial period: "Fue por ese proceder del hombre blanco, por lo que dicen los negros que ellos visten, hablan y actúan en revesina. Por ello contestan *No* cuando es *Sí* y viceversa. La intención del blanco fue siempre convencer al negro de que era bruto e incapaz de igualarse a los rubios" [it was because of this behavior on the part of the whites that the blacks dress, act, and speak backwards. For this reason, they answer *no* when it is *yes* and vice versa. It was also the intention of the whites to convince the blacks that they were brutes incapable of equaling the whites]. Franceschi (1960: 103) reiterates the *Congo* tradition of semantic reversals: "Otro tanto podemos decir en cuanto a la jerga que hablan, cortando las palabras y haciéndose difíciles de entender por el mismo hecho de decir las cosas al revés" [we can say the same about the jargon they speak, cutting off the words and making themselves hard to understand by the same strategy of saying things backwards]. The first known published examples of *Congo* speech were presented by Zárate & Pérez de Zárate. (1962: 126–129), who attempted both a translation and a suggestion for the origin of *Congo* speech during the colonial period. A few other early brief examples were given by Drolet (1980a, 1980b) and Olmos (1980, 1984).

Linguistic analysis of *Congo* speech began with Joly (1981, 1984, 2001, 2006), who studied samples obtained in Miguel de la Borda, in the westernmost Costa Abajo. Joly found some creole-like constructions, particularly the apparent pre-verbal particle *mi*, in addition to general traits such as consonant cluster reduction, reduplication, and the inevitable semantic reversals. At almost the same time, and at first without knowledge of Joly's work, the present author collected and analyzed *Congo* speech from the historically prominent Costa Arriba communities of Portobelo, Nombre de Dios, and Palenque (Lipski 1985, 1986a, 1986b, 1989, 1997). The Costa Arriba communities exhibited some different lexical items from those found on the Costa Abajo, did not use preverbal *mi*, and were characterized by a series of systematic phonetic modifications that are rare or nonexistent in the Costa Abajo communities. Castañeda Álvarez (1997) gives some more recent examples of *Congo* speech from Escobal, also on the Costa Abajo, while González (1988) and Villarreal (2009) give some examples of *Congo* language from Chilibre, a town in the interior between Panama City and the Caribbean port of Colón. *Congo* speech has also been used in literature, most extensively by Góndola Solís (2005a, 2005b, 2005c, 2007), a writer and community leader from Santa Isabel, the easternmost community on the Costa Arriba. Leis (1992) and Marín Zuñigan (1999) include examples of *Congo* language representative of Portobelo. Comprehensive studies

incorporating fieldwork carried out in all *Congo* communities are found in Lipski (2009, 2011, 2017), and inform the following linguistic analysis of *Congo* speech.

Although many *Congo* practitioners affirm that the traditions and speech are the same across the Costa Abajo and Costa Arriba communities, comparative fieldwork carried out by the present author reveals many systematic differences as well as numerous shared traits. All *Congo* speech combines phonetic distortion of Spanish words, frequent semantic displacements – most frequently reversals of meaning – a set of Spanish-derived lexical items that form a common core of *Congo* language, and some linguistic traits that link *Congo* speech to other Afro-Hispanic linguistic manifestations, past and present. The latter will be analyzed in detail in a following section.

All regional varieties of *Congo* language exhibit numerous phonetic distortions of common Spanish words, some of which are essentially lexicalized while others appear to be spontaneous improvisations. Along the Costa Abajo, most phonetic modifications involve vowels, sometimes giving the appearance of vowel harmony (also noted by Joly 1981, 1984); examples include *compañero* > *cumpañede* 'companion, friend,' *bueno/buena* > *güene* 'good,' *bronzo* < *bronce* 'brass,' *entonso* < *entonces* 'then,' *vacíe* < *vacío/vacía* 'empty,' *múquino ajeno* < *máquina ajena* 'machine belonging to someone else,' *huesa* < *huesos* 'bones,' *suavo* < *suave* 'soft,' *gurupata* < *garrapata* 'tick [insect],' *guyine* < *gallina* 'chicken.' In the Costa Arriba communities, in addition to vocalic modifications, intrusive [ɾ] or [ɾj] is sometimes added: *Pringamá* < *Panamá*, *pogriá* < *pagar* 'to pay', *poquitria* < *poquito* 'a little bit,' *diabria* < *diablo* 'devil.' The most striking phonetic modification found in the Costa Arriba communities, and analyzed below as a possible Afro-Hispanic trait, is the neutralization of prevocalic /ɾ/, /r/, and /l/ to [d]: *padase* < *pararse* 'to stand up,' *tieda* < *tierra* 'earth, dirt,' *cudo* < *culo* 'buttocks.'

Although phonetic modifications may render *Congo* speech nearly unintelligible upon first hearing it, for many outside observers, the most characteristic trait of *Congo* speech is the use of frequent semantic reversals, the *revesinas* noted by Franceschi (1960), Zárate & Pérez de Zárate (1962), Joly (1981, 1984) Olmos (1984), and others. Typical examples include *mara* < *malo* 'bad' for *bueno* 'good,' *culo* 'buttocks' for *cara* 'face,' *ponese vivi* < *ponerse vivo* 'to become alive' for *morirse* 'to die,' *ponese entedo* < *ponerse entero* 'to become whole' for *romperse* 'to break.' Partial semantic displacement also occurs, for example *Papa di yo* 'my father' is used to mean 'my mother,' *yodá* (Sp. *llorar*) 'to cry' is 'to sing,' *pretañá/potoñá* (Sp. *pestañar*) 'to blink' is 'to sleep' and *sepedín/sopodín* (Sp. *Zeppelín*) 'Zeppelin' is used for any motorized vehicle, including motor launches and airplanes. The latter word illustrates the ongoing versatility of *Congo* language, which also includes *aeroprango* < *aeroplano*, the somewhat antiquated Spanish word for 'airplane.'

The reversal of meaning is an inherent component of the *Congo* tradition, which includes putting old clothes on inside out, in imitation of the colonial practice of wealthy colonists giving the slaves castoff and even new clothing to wear during the Carnival festivities. According to oral tradition, the slaves put the clothing on *al revés* (which in Spanish can mean 'backwards' or 'inside out') as a gesture of scorn. Contemporary *Congos* also greet one another by shaking feet instead of hands, and blacken their faces with soot from burned-out coconut hulls to 'whiten' themselves. They also collect "fines" (typically alcoholic drinks) for supposed transgressions, and engage in mock land "measurements," all parodies of real-life activities carried out contrary to the usual practice.

In terms of its relation to Spanish, Afro-Panamanian *Congo* speech can be considered as an anti-creole in the sense of Couto (2002), namely a mixed language that combines the morphosyntax of the dominant language and a significant quantity of lexical items from one or more substrate languages. Panamanian *Congo* speech is the only documented Spanish-based anti-creole. Better known anti-creoles are the Afro-Brazilian cryptolects *Cafundó* (Andrade Filho 2000; Vogt & Fry 1996), *Tabatinga* (Queiroz 1998), and *Calunga* (Byrd 2006, 2007; Byrd & Moraes 2007). Some examples are in (1); non-Spanish/Portuguese elements are in italics, including deformed Spanish words in *Congo* speech.

(1) a. Cafundó
Vimbundo está *cupopi*ando na *marrupa* (Vogt & Fry 1996: 39)
black man COP speak-GER in the sleep
'the black man is talking in his sleep'

b. Calunga
Camanu num tá *apum*anu a *nanga* mai, não (Byrd 2006: 64)
man NEG COP put on-GER the pants more NEG
'The man doesn't put on [nice] pants any more'

c. Tabatinga
Os *camonim* chega no *conjolo do cuete* (Queiroz 1998: 95)
the.PL child arrive in the house of the man
'The children arrive at the man's house'

d. *Congos* of Panama
*mi m*engo a *chukero* a *jaraná* (Lipski 2011: 153)
I come to pigsty (*chiquero*) to urinate (*orinar*)
'I am going home to urinate'

The Afro-Brazilian languages contain at least some words of legitimate African origin, while the nominally xenoglossic elements in *Congo* speech are most frequently distortions of Spanish words, but in all cases the mixed languages were created deliberately by fluent speakers of the dominant language (Portuguese or Spanish), as acts of identity and resistance, and involve acknowledgment of African origin.

Unlike Panamanian *Congo* speech, the Afro-Brazilian languages are spoken in geographically delimited communities, have contemporary communicative functions, and are not confined to particular events or seasons.

3. When and where did *Congo* language first emerge?

Although much is known about the linguistic traits of contemporary *Congo* speech, no published descriptions antedate the examples given by Zárate & Pérez de Zárate (1962). All knowledge of earlier stages depends on the questionably reliable recollections of the oldest *Congo* practitioners, whose memories extend to the early 20th century, leaving many questions unanswered. Given the lack of reliable historical documentation on earlier stages of the *Congo* language (Drolet 1980a: 158) – and only sketchy information on maroon communities and other Afro-colonial groups (e.g. De la Guardia 1977: chap. 3, Díez Castillo 1968; Molina Castillo 2011: chap. 2, Tardieu 2009; Vila Vilar 1987) – the only recourse is historical reconstruction based on a combination of synchronic data on *Congo* language and behavior together with comparisons with other Afro-descendent speech communities in Latin America. The following sections will briefly survey the principal questions surrounding the origin and development of *Congo* language, and will propose some initial solutions.

Since any creole-like Spanish dialect would have disappeared from Panama probably by the first decades of the 18th century, at which time the proportion of African-born L2 speakers of Spanish was very small, it is tempting to suggest that *Congo* dialect is a post-slavery creation, coupled to the *Congo* games in order to celebrate freedom, and to hark back to earlier periods when enslaved Africans spoke an incomplete Spanish. This hypothesis is not likely, however, due to a number of considerations. First, the *Congo* dialect shares remarkable similarities among a number of villages whose mutual geographical isolation until only a decade or two ago would have precluded any conscious planning for restoration, imitation, or invention of Afro-Hispanic language. Also significant is the Afro-Panamanians' lack of easily obtainable information regarding Afro-Hispanic speech of earlier centuries; quasi-illiterate residents of isolated coastal villages would have no access to historical materials that could suggest patterns upon which a modern imitation of *bozal* language might be based. Finally, given the known social prejudice against Afro-descendent groups in Panama, there would be no impetus to simply invent a speech form that serves to reinforce racist stereotypes.

The most probable scenario for the creation of *Congo* speech, supported both by historical evidence and by some beliefs within the Afro-Panamanian coastal communities, is that this cryptolect was invented by Afro-descendants during the colonial period, using a combination of recurring pidginized Spanish elements and deliberate morphophonological distortions, all with the purpose of creating a form

of speech that would be unintelligible to the uninitiated while supplying the speakers with a viable communication medium. Although *Congo* language may have been propagated by maroons, it is most reasonable to assume that it arose within the colonial communities in which enslaved Africans were faced with the need for safe communication that could not be easily decoded by Spanish colonists. This *Congo* talk would have served two functions. The first and arguably most essential was to provide captives with a means of freely communicating with one another and to plan acts of rebellion (e.g. maronage, sabotage, etc.) without being understood by the colonial masters. At the same time the deliberate distortion of patrimonial Spanish items constituted a mockery of all things colonial, a manifestation of scorn and repudiation against a society that had enslaved Africans and forced them to live and work far from their ancestral homelands. Given that any African languages had long since disappeared from coastal Panama, Afro-descendants pieced together a language of resistance using the linguistic resources at their disposition, namely the speech of the colonists, but also an awareness of the colonists' anti-African prejudice and the widely held belief that Africans were inherently incapable of learning "proper" Spanish. Both Franceschi (1960: 8) and Zárate & Pérez de Zárate (1962: 126–127) have suggested that *Congo* language originated during the slaveholding period in colonial Panama and exploited Europeans' racist prejudice to create an "African"-sounding language reminiscent of – but not directly derived from – newly arrived Africans' attempts at speaking Spanish.

Previous studies of *Congo* speech had concentrated on individual varieties spoken in a single community, with little thought given to using multi-region data to trace the history of *Congo* language. Given the partially improvised nature of *Congo* language there is considerable idiolectal variation, but in general the varieties of the Costa Arriba (Portobelo, Nombre de Dios, Palenque, etc.) are more Spanish-like and easier to understand than the Costa Abajo varieties found in communities such as Escobal, Chagres, and Miguel de la Borda, which in addition to phonetic distortion and semantic shifts exhibit more morphosyntactic deviations. The first systematic comparison of a broad spectrum of *Congo* varieties is found in Lipski (2009, 2011), together with an attempt to determine whether *Congo* language emerged from a single source or arose spontaneously in more than one location. A phonetic and morphological analysis of lexical samples taken from twenty-four *Congo* communities, ranging from Miguel de la Borda in the west to Santa Isabel in the east, and also including Chilibre along the trans-oceanic corridor[3]revealed a

3. Along the Costa Abajo, from west to east: Miguel de la Borda, Gobea, Río Indio, Palmas Bellas, Chagres, Piña, Cuipo, Escobal. Along the Costa Arriba, from west to east: Pilón, María Chiquita, Portobelo, Cacique, José del Mar, Puerto Lindo, Isla Grande, Nombre de Dios, Viento Frío, Palenque, Miramar, Santa Isabel. Along the central corridor: Puerto Escondido and Cativá (Colón area), and Chilibre. Also the Curundú neighborhood in Panama City.

variation that is both systematic and regionally clustered, precisely the distribution that one expects of a natural language evolving in mutually isolated communities over a period of many generations. The proposed patterns of evolution place the geographical locus around Chagres, Escobal, and Piña along the Costa Abajo. In many other respects this cluster of dialects exhibits the highest proportion of independently documented Afro-Hispanic items (summarized below), together with most of the key lexical items, as compared with invented or distorted modifications of Spanish words. The Costa Abajo was a prime location for the fomentation of resistance movements, from which *Congo* speech apparently emerged. With a small Spanish population and numerous fugitive Africans (Martínez Marín 2004: 126–128), the Costa Abajo was a viable staging ground for more extensive acts of rebellion, which could spread *Congo* speech as far east as Santa Isabel, on the border of the non-Hispanized Kuna redoubt.

Little information on *Congo* theatrical ceremonies and dances is known prior to the 20th century, i.e. well after slavery was abolished in Panama (in 1820). Before this time any overt celebration of African resistance would not have been permitted (although mocking parodies during Carnival season did take place); to the extent that *Congo* language originally emerged as a potentially subversive cryptolect, this linguistic manifestation must antedate the modern *Congo* activities.

4. To what extent does *Congo* language reflect earlier Afro-Hispanic pidginized speech?

Although in essence a deliberately manipulated variety of contemporary Spanish, *Congo* language as produced by traditional speakers – especially in the communities along the Costa Abajo – has a number of elements found in other Afro-Hispanic varieties past and present, a fact which suggests at least some carryover of earlier contact-influenced language phenomena. These items are not found in any contemporary variety of Panamanian Spanish (or elsewhere in the Caribbean), and most probably disappeared in Panama prior to the 19th century. The most salient features found in *Congo* language as well as in other Afro-Hispanic manifestations are summarized below.

a. The gender-invariant third-person pronoun *elle*, previously documented for Afro-Caribbean Spanish and still found vestigially among some elderly Afro-Cubans (Ortiz López 1998; Perl 2001). All Afro-Hispanic and Afro-Portuguese creoles and semi-creoles have gender-invariant pronouns; this includes Afro-Colombian Palenquero, traditional Afro-Bolivian speech (Lipski 2008), Papiamentu, and the Portuguese-lexified creoles of Cape Verde, São Tomé and Príncipe, and Annobón, all of which have some variant of *ele*.

b. The *Congo* personage *Juan de Dioso* shows a paragogic vowel added to *Dios*; the same form was found in literary attestations of Afro-Hispanic speech (Lipski 2002, 2005), as well as in the Portuguese-derived creoles of São Tomé, Príncipe, and Annobón, and (with the variant *Rioso*) in Palenquero. Some *Congos*, particularly along the Costa Arriba, use the expression *mi señodo* 'my lord,' from *mi señor* (but referring to the *Congo* queen), also with paragogic vowel, e.g. *¿y Juan de Dio y mi Señodo que no lo he bristo trohavía?* 'And Juan de Dios y la Reina, whom I haven't seen yet?' (Olmos 1980). *Señoro/señodo/sioro* appears repeatedly in imitations of Afro-Hispanic language from the 16th-18th centuries (Lipski 2005). These words may also reflect vowel harmony since the added vowel is a duplicate of the preceding vowel, as found, for example, in São Tomé and Annobón Portuguese-lexified creoles. Other *Congo* words suggestive of vowel harmony are *entonso* < *entonces* 'then,' *óyomo* < *óyeme* 'listen to me,' *cumpuñede* < *compañero* 'companion,' *nocho* < *noche* 'night.' A few *Congos* use *ayere* < *ayer* 'yesterday,' for example in a song that begins *ayere hablé con un hombre* 'yesterday I spoke with a man.' The form *ayere* is also found in the Afro-Bolivian semicreole dialect.

c. The *Congo* word for 'child' is *jurumingue/juduminga*, presumably derived from *hormiga* 'ant' (as in the *Congo* song "Jurumingue me pica" [the ants are biting me]). The same word is found in Afro-Hispanic songs from Cuba (Feijóo 1987: 67), Puerto Rico (Alegría 1954: 74), and Venezuela (Aretz 1970; Sojo 1986: 86).

d. The invariant copula *so(n)*, also found in Afro-Cuban speech, recurs in *Congo* language (Fuentes Guerra & Schwegler 2005; Lipski 1999), as in the following examples recorded by the author:
a yo *son* nengre 'yo soy Congo' (Piña) ['I am Congo']
elle me *son* negremacha 'soy Congo' (Pilón) ['I am Congo']
ese *son* mi mojonga 'ella es mi esposa' (Cacique) ['she is my wife']
ese *so* la chakere di a yo 'esa es mi casa' (Escobal) ['this is my house']

e. The second-person (singular and plural) *Congo* pronoun *utene* (with occasional variant *antene*) is almost identical to the Afro-Bolivian *otene* (Lipski 2008) and is similar to Palenquero *utere*. The *Congo* pronoun *(ma)soto* < *nosotros* 'we' is very close to the Palenquero *suto*.

f. Along the Costa Abajo *Congo*s are known as *nengres*, a variant of *negro* with intrusive nasal consonant. This item was once attested in Afro-Cuban and Afro-Peruvian Spanish (Lipski 2005) and the same word is found in the Suriname creoles Sranan Tongo (a language once known as *nengre tongo*) and Saramaccan (in the word *nenge*). Several *Congo* dance groups have incorporated *nengre* into their names, and *Congo* ceremonies are punctuated by the cry *¡Viva nengre!* This item is unlikely to have arisen spontaneously among Afro-Panamanians, but rather stands as a vestige of earlier Afro-Hispanic language as used in several colonial settings.

g. Characteristic of the Costa Arriba *Congo* dialects but found only infrequently along the Costa Abajo is the realization of Spanish /ɾ/, /r/ and /l/ as occlusive [d], e.g. *cadeteda* < *carretera* 'road,' *ete dao* < *este lado* 'over here,' *cumpuñede* < *compañero* 'companion.' The loss of the opposition /l/-/ɾ/ and the merger with /d/ are suggestive of the phonological systems of Bantu languages, in which there is a single liquid phoneme, and in which [l] and [d] are usually in complementary distribution, and similar phenomena have been documented for other Afro-Hispanic linguistic remnants in Latin America (Lipski 2007). Some *Congos* from the Costa Abajo indicate that it was once more common to pronounce both /d/ and /l/ as [ɾ], particularly in the article *la*. Both pronunciations are very infrequent nowadays, suggesting a diachronic shift in the direction of modern Spanish.

h. *Congo* speech on the Costa Abajo has several instances of onset cluster reduction, most noticeably in *oto* < *otro* 'other' and *(ma)soto* < *nosotros* 'we.' Onset cluster reduction is a common denominator in much Afro-Hispanic language across several centuries (Lipski 2002). This pronunciation is infrequent in the Costa Arriba dialects, but in all likelihood it was once part of the *Congo* language of that region. Since full onset clusters represent a closer approximation to modern Spanish patterns, it is possible to extrapolate to an earlier period of Afro-Panamanian speech in which more onset cluster reduction was occurring.

A few vestiges of earlier *bozal* Spanish also survive in the lyrics of Panamanian folk dances, particularly the Zaracundé (Rhodes 1998), which suggests that awareness of Afro-Hispanic speech may have been more extensive in earlier times. This dance, also known as *El Cuenecué* or *Danza de los negros bozales*, is currently performed in the town of Los Santos (with a very small population of African origin), but was once performed during Carnival season in other parts of Panama. One of the characters of this ritual dance is the *Negro bozá*, a pronunciation reflecting the truncation of final consonants in Afro-Hispanic speech; final /ɾ/ is frequently deleted in vernacular Panamanian Spanish, but final /l/ almost never falls in contemporary speech. Other characters' names also reflect *bozal* confusion of Spanish morphological endings: *Pajarité* [*pajarito* 'little bird'], *Fransisqué* [*Francisco*]. The *Negro bozá* chants phrases which include Afro-Hispanic *bozal* language, including *yo tené* [*yo tengo* 'I have'], *la huerté* [*la huerta* 'the garden'], *yuqué* [*yuca* 'yucca'], *tamarindé* [*tamarindo* 'tamarind'], *papayé* [*papaya*]. The frequent replacement of Spanish final *-o* and *-a* by *-e* is similar to phenomena attributed in literature to Haitian L_2 speakers of Spanish in the Dominican Republic (Lipski 2004), and was actually verified by Ortiz López (1999a, 1999b, 2001, 2010).

Putting together the contemporary variation in *Congo* speech and what diachronic developments can be extrapolated, a complex picture emerges that cannot be easily resolved with the notion that this dialect developed exclusively as a

cryptolect in contact with Spanish colonists. Despite many regional differences, there is a clear dichotomy between the Costa Arriba and Costa Abajo varieties: the Costa Abajo *Congo* language is more "basilectal," i.e. exhibiting more traits that can be traced to the pidginized Spanish spoken by African-born *bozales*. The Costa Arriba speech contains a much higher proportion of phonetic modifications and formulaic semantic inversions (e.g., simply adding *no* to affirmative clauses), which may reflect a later arrival of *Congo* language in this region and a more rapid shift to Spanish.

5. To what extent was – and is – *Congo* language used for effective communication?

Given the humorous and largely improvised nature of contemporary *Congo* speech, the question arises as to whether such a strategy could have been a viable means of subversive communication during the colonial period, together with the possibility that modern *Congo* language is less systematic and therefore less suitable for efficient communication than during earlier times. Nowadays only a relatively small group of residents in each community routinely use the dialect, but in previous generations, when *Congo* speech was used widely during Carnival season, other residents had some passive knowledge. The rapid decline in the number of *Congo* speakers (described in Lipski 2017) has resulted in a scarcity of knowledge of *Congo* language among younger residents, some of whom "re-invent" *Congo* speech in a fashion that bears little resemblance to the production of traditional speakers.

The *Congo* queen Alejandrina Lan (born 1930) relates that as an adolescent her father (a former *Congo* king) took her to visit all the *Congo* communities along Panama's Caribbean coast, to learn the traditions and meet the leaders of each group. She affirms from memory that *Congo* speech was essentially identical from one community to another, although given substantial differences documented only a few years later and still found today, this assertion must be regarded with caution. It is likely, however, that communication between *Congos* from widely separated villages was efficient and effective, given that most practitioners used the language frequently and in a variety of circumstances. A picture in Franceschi (1960: 95) shows a meeting between *Congos* from the Costa Abajo and the Costa Arriba. This inter-community communication rarely takes place today, and *Congos* from the Costa Arriba and the Costa Abajo may not actually understand one another, and may resort to simple verbal posturing without any clear communicative intent. In 2008 the author traveled with Marcia Rodríguez, the daughter of Alejandrina Lan and currently the reigning *Congo* queen, to Nombre de Dios, where Marcia (who speaks the Escobal *Congo*

variety from the Costa Abajo) attempted to speak to a young *Congo* who clearly did not comprehend her language, and who responded with *Congo*-like gibberish that in no way responded to Marcia's questions and observations.

Describing the situation in the 1970s, Drolet (1980a: 181) noted that "Individuals who do not participate understand much of what is being said, but they do not respond in the dialect. Congo dialect is used by male role performers as a means of trickery or bribery […] an individual skilled in the dialect is able to speak in a fast and convoluted manner to confuse and entangle the other person into doing or saying something which is interpreted by the Congo male to his advantage […]" She also noted that many of her informants were unable to "translate" their own utterances after the fact. Her interpretation was that the informants were making excuses based on their own inebriation; her subjects could not speak the dialect unless they had reached a threshold level of intoxication, which leads to the conclusion that the dialect itself is the result of drunken deformations of the Spanish language, perhaps to the exclusion of any authentic African base. In this author's experience, the most cohesive specimens of *Congo* language are produced in a state of total sobriety. The present author has recorded very long *Congo* interchanges during Carnival season, noting the progressive linguistic deterioration that accompanied the increasing consumption of alcohol, including the repetition of stereotyped phrases, as well as the loss of the songs, poems and unique *Congo* elements that characterize virtuosic dialect performances.

Describing *Congo* speech in other communities during the same time period, Joly (1981: 17) observed that "Although certain morphophonemic changes are made, the basic canonic shape of words and the syntactic structure of Spanish remains essentially the same so that the symbolic meaning of the words and sentences prevails […] It is plausible, therefore, that some words are memorized rather than generated by morphophonemic rules. The players, however, do not memorize verses or dialogues […] but improvise their performance according to behavior expected of their roles."

Some young *Congo* practitioners have become adept at largely meaningless strings of syllables and jumbled words approximating the traditional *Congo* speech, as in the case of the young Nombre de Dios speaker described above. Most authentic *Congo* speakers, while perhaps possessing a small repertoire of expressions and formulaic topics in the dialect, are nonetheless able to decipher their own speech and that of others, even when unwilling (or unable) to parse and describe every individual word. True *Congo* speech from the Costa Arriba is considerably easier for the uninitiated to decode than the varieties found along the Costa Abajo, as long as speakers are actually using accepted *Congo* language and not simply improvising by randomly distorting Spanish and creating nonsense syllables and words.

The inclusion of earlier Afro-Hispanic elements and the more systematic usage of the oldest surviving *Congos* are consistent with a language that once had greater communicative value and therefore greater mutual intelligibility among practitioners than in contemporary Panama, where *Congo* traditions are rapidly being repurposed to attract tourists as well as to participate in more wide-ranging celebrations of the African diaspora in the Americas.

6. Conclusions: In search of the *Congo* sociolinguistic trajectory

As noted at the outset, the history of Afro-Panamanian *Congo* practices is hampered by the lack of reliable data prior to the mid-20th century, but, based on the comparative analyses reviewed in the present study, some preliminary conclusions can be put forth. First, given the steep decline in *Congo* language usage over the past generation and the increasingly haphazard and meaningless approximations to *Congo* speech by the youngest practitioners, it is likely that the original form of this speech mode had a considerably higher effective communication value. Earlier *Congo* language evidently also contained a higher proportion of authentic Afro-Hispanic elements and perhaps more creative semantic reversals instead of the overly facile pre-posing of *no* to otherwise unremarkable utterances, as frequently occurs among contemporary *Congos* (cf. the reference to the Portobelo recording of 2015 found in the Appendix). This in turn would place the emergence of *Congo* language within the colonial slaveholding period. In view of the rapid decrease in the proportion of African-born Spanish speakers after the middle of the 17th century (cf. Figure 1), memories of Afro-Hispanic *bozal* language among colonial-born Afro-descendants would rapidly fade in succeeding generations.

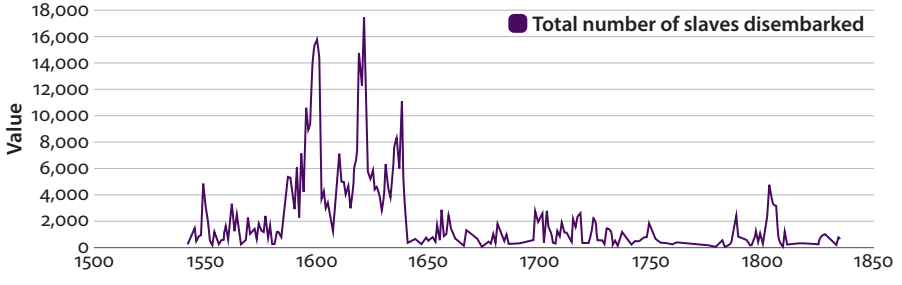


Figure 1. Arrival of enslaved Africans along the Caribbean coast of South America (from the Trans-Atlantic Slave Trade Database (https://www.slavevoyages.org)

Even the Afro-Panamanian urban *cofradías* were by the end of the 17th century largely composed of American-born Afro-descendants, both free and enslaved (Mena García 2000; Molina Castillo 2011: 193–204), and moreover these organizations would have little contact with any remaining rural maroon communities or small coastal villages.

The largest number of maroon revolts took place in Panama during the 16th and early 17th centuries (e.g., Diez Castillo 1968; Gallup-Díaz 2010; Pike 2007; Rodríguez 1979), and under the assumption that *Congo* language emerged as a subversive strategy, the combined facts converge on the hypothesis that *Congo* speech arose no later than the middle of the 17th century (cf. Zárate & Pérez de Zárate 1962: 120, Franceschi 1960: 98, citing the researcher Felicia Santizo). Even when rebellion and revolt were not at stake, *Congo* language survived among coastal Afro-Panamanian communities long past the time when African-born *bozales* could be heard, and with the abolition of slavery the *revesinas* and other verbal jousting could become the primary function of *Congo* speech, largely confined to the Carnival season and no longer serving a primary communicative function. The names of some *Congo* personages also reflect true maroon leaders, for example Juan de Dioso (Zárate & Pérez de Zárate 1962: 122) and Pajarito (De La Guardia 1977: 104). Although some *Congo* speakers still use the language in non-ludic contexts (the *Congo* queens Alejandrina Lan and Marcia Rodríguez routinely speak of daily activities in *Congo*), most of the remaining speakers are adept only at the mock-aggressive confrontations in which *Congo* speech is used to accuse and demand tribute.

The rapid shift away from the use of *Congo* language usage, especially along the Costa Arriba, is tied to the improved infrastructure along the entire Caribbean coast, including paved roads leading to communities that were previously reachable only on foot or by small boats on dangerously choppy waters. The current revival of *Congo* ceremonies for tourists, especially in Portobelo and Nombre de Dios (Alexander Craft 2008b) does not frequently include *Congo* speech, and when it is present (as in the 2015 transcription sample from Portobelo in the Appendix) there are none of the traditional phonetic modifications or Afro-Hispanic carryovers that once characterized the speech of this region. Most accounts of *Congo* culture that regularly appear in the Panamanian press every carnival season mention only the costumes and dancing, rarely referring to *Congo* language. In Costa Arriba communities not frequented by tourists (e.g., Cacique, Puerto Lindo) the remaining *Congo* speech adheres more closely to traditional patterns, while on the Costa Abajo, where tourism during Carnival is quite limited, more authentic *Congo* language is still to be heard.

The history of *Congo* speech may never be fully elucidated, but comparative linguistic analysis together with an examination of the cultural and demographic

history of Afro-descendants in colonial Panama can provide a plausible reconstruction of the sociolinguistic trajectory leading from maroon resistance to Afro-Panamanian street theater and the accompanying language usage. Only in this fashion can the confusing thicket of misleading and inaccurate characterizations of *Congo* speech be traversed, in order to do justice to a language born of resistance to enslavement.[4]

References

Alegría, Ricardo. 1954. *La fiesta de Santiago Apóstol en Loíza Aldea*. Madrid: ARO Artes Gráficas, Colección de Estudios Puertorriqueños.

Alexander, Renee. 2006. *Art as survival: The Congo tradition of Portobelo, Panama*. Evanston, IL: Northwestern University dissertation.

Alexander Craft, Renee. 2008a. Una raza, dos etnias: The politics of be(com)ing/performing 'Afropanameño.' *Latin American and Caribbean Ethnic Studies* 3(2). 123–147. https://doi.org/10.1080/17442220802080519

Alexander Craft, Renee. 2008b. "¡Los gringos vienen!" (The gringos are coming!): Female respectability and the politics of Congo tourist presentations in Portobelo, Panama. *Transforming Anthropology* 16(1). 20–31. https://doi.org/10.1111/j.1548-7466.2008.00003.x

Alexander Craft, Renee. 2015. *When the devil calls: The Congo tradition and the politics of blackness in twentieth-century Panama*. Columbus: Ohio State University Press.

Andrade Filho, Silvio Viera de. 2000. *Um estudo sociolingüístico das comunidades negras do Cafundó, do Antigo Caxambu e de seus arredores*. Sorocaba: Secretaria de Educação e Cultura.

Aretz, Isabel. 1970. *El tamunangue*. Barquisimeto: Universidad Centro-Occidental.

Béliz, Anel. 1959. Los Congos: Afro-Panamanian dance-drama. *Américas* 11(11). 31–33.

Bettelheim, Judith. 2004. Carnaval of los Congos in Portobelo, Panama: Feathered men and queens. In Geneviève Fabre & Klaus Benesch (eds.), *African diasporas in the New and Old worlds: Consciousness and imagination*, 287–309. Amsterdam/New York: Rodopi.

Byrd, Steven. 2006. Calunga: An Afro-Brazilian speech. *Papia* 16. 62–80

Byrd, Steven. 2007. Calunga: uma fala afro-brasileira de Miras Gerais, sua gramática e história. *Revista Internacional de Lingüística Iberoamericana* 5(1). 203–222.

Byrd, Steven & Daniela Bassani Moraes. 2007. Calunga and *Calungadores:* An Afro-Brazilian speech community of Minas Gerais. *Afro-Hispanic Review* 26(2). 27–45.

---

4. Many individuals have contributed to the success of my research into *Congo* language. During the 1980s Luz Graciela Joly and Patricia Drolet graciously shared their data with me, and my field expeditions were aided by Dr. Soledad Epifanio of the University of Panama and Aminta Núñez of the INAC. My return to *Congo* territory in the 21st century was greatly aided by *Congo* princess (now Reina) Marcia Rodríguez Lan and inspired by the activism of her mother and Queen, Alejandrina Lan. Gustavo Lawson was a faithful companion during much of the fieldwork and Nolis Boris Góndola and Ernesto Polanco shared their deep knowledge of *Congo* culture and language. And to all the *Congos* who have taught me their language and acquiesced to my inquiries, *¡Viva nengre!*

Castañeda Álvarez, Otilia María. 1997. *Dramatugria y jerga del congo en Escobal*. Panamá: Universidad de Panamá Tesis de licenciatura,

Couto, Hildo do. 2002. *Anticrioulo: manifestação lingüística da resistência cultural*. Brasília: Thesaurus Editora.

De la Guardia, Roberto. 1977. *Los negros del istmo de Panamá*. Panamá: Instituto Nacional de Cultura.

De la Rosa Sánchez, Manuel Antonio. 1988. El juego de los tambores congos (tradición oral afromestiza de Panamá). In Guillermo Quartucci (ed.), *Segundo Congreso Nacional Asociación Latinoamericana de Estudios Afroasiáticos, Universidad Veracruzana, Jalapa, Veracruz, 3 al 5 de julio de 1985*, 153–177. México: El Colegio de México.

Díez Castillo, Luis. 1968. *Los cimarrones y la esclavitud en Panamá*. Panamá: Editorial Litográfica.

Drolet, Patricia. 1980a. *The Congo ritual of northeastern Panama: An Afro-American expressive structure of cultural adaptation*. Urbana-Champaign: University of Illinois dissertation.

Drolet, Patricia. 1980b. *El ritual congo del noroeste de Panamá: una estructura afro-americana expresiva de adaptación cultural*. Panamá: Instituto Nacional de Cultura.

Feijóo, Samuel (ed.). 1987. *El negro en la literatura folklórica cubana*. Havana: Editorial Letras Cubanas.

Franceschi, Víctor. 1960. Los negros congos en Panamá. *Lotería* 51. 93–107.

Fuentes Guerra, Jesús & Armin Schwegler. 2005. *Lengua y ritos del Palo Monte Mayombe: dioses cubanos y sus fuentes africanas*. Frankfurt/Madrid: Vervuert/Iberoamericana. https://doi.org/10.31819/9783964565464

Gallup-Díaz, Ignacio. 2010. A legacy of strife: Rebellious slaves in sixteenth-century Panama. *Colonial Latin American Review* 19(3). 417–435. https://doi.org/10.1080/10609164.2010.520951

Góndola Solís, Nolis Boris. 2005a. *El revellín de los bumbales*. 2nd edn. Colón: Biblioteca Siglo XXI, Imprenta Colón.

Góndola Solís, Nolis Boris. 2005b. *La gloria del revellín de los bumbales*. 2nd edn. Colón: Biblioteca Siglo XXI, Imprenta Colón.

Góndola Solís, Nolis Boris. 2005c. *Herencia del revellín de los bumbales*. Panamá: Imprenta Xtreme Solutions.

Góndola Solís, Nolis Boris. 2007. *Golpe congo del revellín de los bumbales: poemas y cuentos de un negro Congo*. Panamá: Imprenta ARTICSA.

González, Natanael. 1988. *Los diablos sucios de Chilibre*. Panamá: Universidad de Panamá tesis de licenciatura.

Joly, Luz Graciela. 1981. The ritual play of the Congos of north-central Panama: Its sociolinguistic implications. *Sociolinguistic Working Papers* (Southwest Educational Development Laboratory, Austin, Texas) 85.

Joly, Luz Graciela. 1984. Implicaciones sociolingüísticas del juego de congos en la Costa Abajo de Panamá. *Revista Lotería* 338–339. 22–55.

Joly, Luz Graciela. 2001. El lenguaje del juego ritual de los Congos. *Revista electrónica Istmo* 7, http://collaborations.denison.edu/istmo/n07/articulos/lenguaje.html.

Joly, Luz Graciela. 2006. *El lenguaje del juego ritual de los congos*. David, Panama: Luz Graciela Joly Adames.

Laribe, Lucette. 1968. *Nombre de Dios et les "regnes de Congos."* Panama: Alliance Française Panamá.

Laribe, Lucette. 1969. *Les "regnes de Congos" de Nombre de Dios*. Panama: Alliance Française Panamá.

Leis, Raúl. 1992. *Mundunción*. Panamá: Instituto Nacional de Cultura.

Lipski, John. 1985. The speech of the Negros Congos of Panama: Creole Spanish vestiges? *Hispanic Linguistics* 2. 23–47.
Lipski, John. 1986a. The Negros Congos of Panama: Afro-Hispanic creole language and culture. *Journal of Black Studies* 16. 409–428. https://doi.org/10.1177/002193478601600405
Lipski, John. 1986b. El lenguaje de los *negros congos* de Panama. *Lexis* 10. 53–76.
Lipski, John. 1989. *The speech of the Negros Congos of Panama*. Amsterdam: John Benjamins. https://doi.org/10.1075/cll.4
Lipski, John. 1997. El lenguaje de los *negros congos* de Panamá y el *lumbalú* palenquero: función sociolingüística de criptolectos afrohispánicos. *América Negra* 14. 147–165.
Lipski, John. 1999. Evolución de los verbos copulativos en el español *bozal*. In Klaus Zimmermann (ed.), *Lenguas criollos de base lexical española y portuguesa*, 145–176. Frankfurt: Vervuert.
Lipski, John. 2002. Epenthesis vs. elision in Afro-Iberian language: A constraint-based approach to creole phonology. In Teresa Satterfield, Christina Tortora, & Diana Cresti (eds.), *Current issues in Romance languages*, 173–188. Amsterdam/Philadelphia: John Benjamins. https://doi.org/10.1075/cilt.220.13lip
Lipski, John. 2004. Nuevas perspectivas sobre el español afrodominicano. In Sergio Valdés Bernal (ed.), *Pensamiento lingüístico sobre el Caribe insular hispánica*, 505–552. Santo Domingo: Academia de Ciencias de la República Dominicana.
Lipski, John. 2005. *A history of Afro-Hispanic language: Five centuries and five continents*. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511627811
Lipski, John. 2007. El cambio /r/ > [d] en el habla afrohispánica: ¿un rasgo fonético 'congo'? *Boletín de Lingüística* 19(27). 94–114.
Lipski, John. 2008. *Afro-Bolivian Spanish*. Frankfurt/Madrid: Vervuert/Iberoamericana. https://doi.org/10.31819/9783865279026
Lipski, John. 2009. Tracing the origins of Panamanian *Congo* speech: The pathways of regional variation. *Diachronica* 26. 380–407. https://doi.org/10.1075/dia.26.3.08lip
Lipski, John. 2011. *El habla de los congos de Panamá en el contexto de la lingüística afrohispánica*. Panamá: Instituto Nacional de Cultura.
Lipski, John. 2017. The *Congo* language of Panama: Born of resistance, surviving through activism. *Afro-Hispanic Review* 36. 97–118.
Luque de Pérez, Colomba. 2001. The Congo of Costa Arriba. In *Afro-American cultural identity in language and literature*, 51–55. Panama: Universidad de Panamá; Sociedad de Amigos del Museo Afroantillano (SAMAAP); Kellogg Foundation.
Marín Zuñigan, Maricel. 1999. *Paloma reina de los Congos: el orgullo de una raza*. Portobelo: Fundación Interamericana del Desarrollo/Centro de Estudio para la Promoción del Desarrollo/Comité Distrital "Amalia Garay Vda. de Luque."
Martínez Marín, José. 2004. *Dominicos en Panamá*. Panamá: Imprenta ARTICSA.
Mena García, Carmen. 2000. Religión, etnia y sociedad: cofradías de negros en el Panamá colonial. *Anuario de Estudios Americanos* 57(1). 137–169. https://doi.org/10.3989/aeamer.2000.v57.i1.262
Molina Castillo, Mario José. 2011. *La tragedia del color en el Panamá colonial 1501–1821*. Panama: Imprenta Impresos Modernos.
Olmos, Domingo. 1980. *La canta de los congos de la Costa Arriba*. Panama: Universidad de Panamá tesis de licenciatura.
Olmos, Domingo. 1984. La música de los negros congos de la Costa Debajo de Colón. *Revista Lotería* 338–339. 56–86.

Ortiz López, Luis. 1998. *Huellas etno-sociolingüísticas bozales y afrocubanas*. Frankfurt: Vervuert. https://doi.org/10.31819/9783865278845

Ortiz López, Luis. 1999a. El español haitiano en Cuba y su relación con el habla bozal. In Klaus Zimmermann (ed.), *Lenguas criollos de base lexical española y portuguesa*, 177–203. Frankfurt: Vervuert.

Ortiz López, Luis. 1999b. La variante hispánica haitianizada en Cuba: otro rostro del contacto lingüístico en el Caribe. In Amparo Morales (ed.), *Estudios de lingüística hispánica: homenaje a María Vaquera*, 428–456. Río Piedras, PR: Editorial de la UPR.

Ortiz López, Luis. 2001. El sistema verbal del español haitiano en Cuba: implicaciones para las lenguas en contacto en el Caribe. *Southwest Journal of Linguistics* 20(2). 175–192.

Ortiz López, Luis. 2010. *El español y el criollo haitiano: contacto lingüístico y adquisición de segunda lengua*. Frankfurt/Madrid: Vervuert/Iberoamericana. https://doi.org/10.31819/9783865279071

Perl, Matthias. 2001. Nelle/cunelle–afrospanische Pronomen in kubanischen Theatertexten im 19. Jahrhundert. In Bruno Staib (ed.), *Linguistica romanica et indiana. Festschrift für Wolf Dietrich zum 60. Geburtstag*, 347–355. Tübingen: Gunter Narr.

Pike, Ruth. 2007. Black rebels: The Cimarrons of sixteenth-century Panama. *The Americas* 64(2). 243–266. https://doi.org/10.1353/tam.2007.0161

Queiroz, Sônia Maria de Melo. 1998. *Pé preto no barro branco: a língua dos negros da Tabatinga*. Belo Horizonte: UFMG.

Racero, Amada. 2005. El grito de los cimarrones. *La Prensa (Panama), 23 febrero 2005*, https://impresa.prensa.com/grito-cimarrones_0_1409859155.html.

Rhodes, Elizabeth. 1998. Saracundé of Panama: A look at the dance on the street and on the stage. In Nancy Stokes (ed.), *The art of the moment: Looking at dance performance from inside and out* (Proceedings of the 31st Annual Conference Congress on Research in Dance), 133–140. Columbus: Ohio State University, Department of Dance.

Rodríguez, Frederick Marshall. 1979. *Cimarron revolts and pacification in New Spain, the Isthmus of Panama and colonial Colombia, 1503–1800*. Chicago: Loyola University of Chicago dissertation.

Smith, Ronald. 1975. *The society of los Congos of Panama*. Bloomington, IN: Indiana University dissertation.

Smith, Ronald. 1994. *Arroz colorao: Los Congos* of Panama. In Gerard H. Béhague (ed.), *Music and black ethnicity: the Caribbean and South America*, 238–266. New Brunswick, NJ/London: Transaction Publishers/North-South Center, University of Miami.

Sojo, Juan Pablo. 1986. *Estudios del folklore venezolano*. Los Teques: Biblioteca de Autores y Temas Mirandinos, Instituto Autónomo Biblioteca Nacional y de Servicios de Bibliotecas.

Tardieu, Jean-Pierre. 2009. *Cimarrones de Panamá: la forja de una identidad afroamericana en el siglo XVI*. Frankfurt & Madrid: Vervuert/Iberoamericana. https://doi.org/10.31819/9783954872763

Tejeira Jaén, Bertilda. 1974. Los Congos de Chepo. *Patrimonio Histórico* 1(3). 129–148.

Vila Vilar, Enriqueta. 1987. Cimarronaje en Panamá y Cartagena. El costo de una guerrilla en el siglo XVII. *Cahiers du monde hispanique et luso-brésilien* 49. 77–92. https://doi.org/10.3406/carav.1987.2341

Villareal, José Bolívar. 2009. Los diablos sucios de Chilibre. *Revista Lotería* 486. 73–93.

Vogt, Carlos & Peter Fry. 1996. *Cafundó: a Africa no Brasil, linguagem e sociedade*. São Paulo: Editora da UNICAMP.

Zárate, Manuel & Dora Pérez de Zárate. 1962. *Tambor y socavón*. Panama: Imp. Nacional.

Appendix. Transcribed examples of *Congo* speech

From Zárate & Pérez de Zárate (1962); probably Piña:

Oye, Micé, pónete y-ejo da y-o. A … y-o sao Micé Putó de Borrachate, de Pinga. Yo tene acutose pabresa en Guinea. Como sa ma consigue la buena muette de c'abrá la mutación de la cutú mantenía, sei mía, y sa cutú ya no cutú, ya na juntao. Antonce a … y-o a puse con toa mi nengre a que adarara comino hacé un puceo puqui, purú curutera, putó cumino ata que he llega Cunumá. Cuando y-o te lleve en Cundumá y-o consigue araranao que so a utene. Son araranao de Saro Brujo. Saguro. Y ahora utene son la minga que y-o tene ma malo, sao, son la que tene mi colasón competamente cutío, sao, son la que tene la mala muette de mi alma. Como a y-o no le guta utene y-o tene una cunción que le va cuntá que se mama "lame la teta cupala". O-yo?

The Zárates' rough translation, from memory: 'Oye, Mercé, ponte cerca de mí. Yo soy Mercé, la legítima representante de Piña. Yo tengo mis dominios en Guinea. Como yo considero, es una buena suerte el de la costumbre mantenida de que seas de los míos, costumbre que debe ser siempre costumbre, para que estemos juntos. Por eso dispuse con todos mis negros que hiciéramos el camino y un paseo hasta aquí, por la carretera y por todos los caminos hasta llegar a Panamá. Cuando he llegado a Panamá he sido recibido por el Reinado de Uds. Nosotros somos del Reinado de Cerro Brujo, seguro. Y ahora veo que Uds. son los amigos más buenos que yo tengo, ¿saben? Son los que tienen mi corazón completamente rendido, los que dan la buena vida a mi alma. Y como Uds. me gustan tanto, yo tengo una canción que les voy a cantar, que se llama "Dale teta, chúpala"'.

From Olmos (1980), Nombre de Dios:

ombe, ahoda mismo está too brien
'man, everything is fine now'

¿y Juan de Dio y mi Señodo que no lo he bristo trohavía?
'and Juan de Dioso and the Queen, that I haven't seen yet?'

…bruscamo do grolluneta, pa poneno en movimienta de mujede y de mojongo y too lo juduminga tambrién
'let's get the drummers and get moving with all the women and children'

brien, si tú no quiede no agade el sopodín y no te vaya, que yo me quedo aquí confriomio
'well if you like take the airplane and leave, I'll stay here just the same'

From Joly (1981), Miguel de la Borda:

Pujurete: Güene, poque ya eso mi meno sempe, cuando ya mi so la nuevecite, que mi meno útimo. Ya ete mi so lo rurumento delle. Y entonso masoto mi sigo la mime rurumento que mi meno tora la sumana.	***Pajarito***: Bueno, porque ya eso viene de siempre; Cuando ya eran los nuevecitos (viejos) que vinieron de último (primero). Ya este es el reglamento de ellos. Y entonces nosotros seguimos el mismo reglamento, cuando venimos todas las semanas (años).
Po lo meno ahoro, masoto mi mene la otu sumana. Mi co Gubé, mi mene otu. Ya elle mimito como masoto; mi weyo.	Por lo menos ahora, nosotros venimos la otra semana (año). Lo mismo Gobea, vendrán otra vez. Ya ellos son mismito como nosotros; ellos juegan.

From Lipski (1989):

Buena, también hay machas que se ponen sobre ehte dao, a jubriyá, y cuntá, po e padenque de negremasha hay mucho zancudo 'Hay mujeres que vienen aquí a hablar, porque en el palacio de los Congos hay muchos zancudos' (Portobelo, 1984).

tiene que pogriá una … una cuntrubución … de cododao, de bianco de … de do que sea, agua de sodiya 'Tienes que pagar una contribución, de ron añejo, ron blanco, lo que sea, aguardiente' (Palenque, 1984).

tú como gruyuneda, y tú como congo de … de triempo antrigu, como negro macha de triempu antrigu, uhtede tiene que sabé … lo que es un botije llene 'tú como tambolero, como Congo de los tiempos antiguos, debe saber lo que es una botella vacía' (Nombre de Dios, 1984).

From González (1988), Chilibre:

Ño Matía tiene gana de friuma	El señor Matías tiene ganas de fumar,
puqui llega su mojongo	por aquí va su mujer
que e su cucuñera	que es su compañera
y lleva el sogodilla.	y le trae el cigarrillo.
Juduminga, Juduminga	Muchachos, muchachos,
vamo a godobotea	vamos a bailar
sino tiene compañede	si no tiene compañero
cudiá paa ete da.	ponte para este lado.

From Lipski (2011):

la biarín, se angara uno marote, si pone tirao pram bajala qui puera quedá con la canida dentre di la chakere se angara la puyeja, la brajuco la brajuco munó, si angara si pone, si angara y si mete con ra cruña y con una caca chibre se va tumprando 'el tambor, se agarra un palo, se tumba fuera de la casa, se le quita la corteza, se amarra un bejuco, si meten las cuñas y se templa' (Palmas Bellas, 2007).

yo no me acuda; bueno sabo lo que va cambrasá, yo tene como cinco aña, como cinco aña yo tene que yo no quere crabatiá má pero yo quero que no se jaya la tradisió tu sabo que no se pone jayao porque si se jaya entonce yo no conocí mi papo yo no conocí mi abuelo yo no cunuquiao ninguno pero cua mayo, mi papa é nengre, mi mama é nengre, mi hermano nengre, tó nengre, yo tumpoco nengre; ¿sabo cómo yo me mama en nengre? Revellín. Cuando ta nueve como yo se va jayando todo lo que tiene aquí afuere. 'No me acuerdo; sabes lo que te voy a decir, tengo como cinco años que no bailo, pero no quiero que se pierda la tradición, porque si se pierde, entonces no conocí a mi padre, a mi madre, a mi abuelo, pero ahora mi padre es Congo, mi madre es Conga, mi hermano, todos Congos, yo soy Conga también. ¿Sabes cómo me llamo en Congo? Revellín. Cuando uno es viejo como yo se le va olvidando todo lo que tiene adentro' (Escobal, 2007).

A young Congo from Portobelo, recorded 2015 (https://www.youtube.com/watch?v=xMZhvdgsq84):

Muy bueno nocho ne(g)ro macho y saluda a toa la mojonga que no me quieren tá viendo en este Carnal; ahora mimo ne(g)ro macho viene una susasón crítica y económicamenta que no sabo; e negro macho é cómo se llamo mi nome sume Candeliye; a tó mundo que no tá conociendo la Candeliye lo telnen que tá viene internacionalmente en Corombia Panamaso, lo que tú no quiero

[…] *Purtubeya donde viene lo verdadera Conga y la raí culturá que significa Conga*. Buenos días señores y saludos a todas las mujeres que me quieren ver en este Carnaval; ahora este hombre [yo] les ofrece una presentación crítica y económica, ¿sabe? este hombre se llama Candelilla, a todas las personas que vienen conociendo a Candelilla que vienen internacionalmente de Colombia, Panamá, dondequiera […] Portobelo, de donde vienen los verdaderos *Congos* y la raíz cultural que significa *Congo*

A socio-historical perspective on the origin and evolution of two Afro-Andean vernaculars

Sandro Sessarego
University of Texas at Austin

This paper casts light on the sociohistorical background of two Afro-Andean vernaculars: Yungueño Spanish (Bolivia) and Choteño Spanish (Ecuador). Contrary to what has been suggested in the literature (Lipski 2008; Perez 2015; Schwegler 1999, 2014), results indicate that a concomitance of sociodemographic factors significantly reduced the possibility of Spanish creoles forming in the colonial Andes. For this reason, this study provides new data that contribute to the long-lasting debate on the evolution of the Afro-Hispanic varieties of the Americas (McWhorter 2000; Lipski 2005; Sessarego 2019a). In particular, the evidence here reported appears to cast serious doubts on proposals suggesting that these and other Afro-Hispanic varieties may be conceived of as the result of a previous (de)creolization phase (Granda 1968 *et seq*.).

Keywords: Yungueño Spanish, Choteño Spanish, Afro-Andean Spanish, Spanish Creole debate, slavery, (de)creolization

1. Introduction

The analysis of the origin and evolution of the Afro-Hispanic vernaculars of the Americas has long been a debated topic in the field of creole studies. Different authors have offered contrasting accounts on the formation of these contact varieties and no common consensus has yet been achieved (Granda 1968; Otheguy 1973; Schwegler 1991a, 1999; Lipski 2005; Perez 2015; Sessarego 2019a). The present article is aimed at casting some light on the evolution of two specific Afro-Hispanic dialects: Yungueño Spanish (YS) and Choteño Spanish (CS). YS and CS are spoken in two rural regions of the Andes (in the Yungan Valleys, Department of La Paz, Bolivia and Chota Valley, Departments of Imbabura and Carchi, Ecuador), by the descendants of the slaves taken to these regions in colonial times to work on plantations.

https://doi.org/10.1075/ahs.12.c07ses

YS and CS, in line with the vast majority of the other Afro-Hispanic vernaculars spoken across the Americas, share a number of morpho-syntactic features that diverge systematically from those found in standard Latin American Spanish (Sessarego 2013a). For this reason, in some cases, such grammatical deviations from the standard norm have been taken as indicators of a creole origin for all of these varieties, which in a subsequent phase (after the abolition of slavery in the 19th century) would have approximated to standard Spanish and thus decreolized (Decreolization Hypothesis, Granda 1968 *et seq*.).

As I have indicated elsewhere, all the features that have been ascribed to such a hypothetical creole language can also be explained as the result of advanced second-language acquisition processes, which do not necessarily imply any previous creole stage (Sessarego 2019b). Due to space limitations, I will not re-propose such linguistic analyses here, but the interested reader may want to consult Sessarego (2013b, 2014) for a more detailed elucidation of the problem. The present paper will offer a sociodemographic analysis of slavery in colonial Bolivia and Ecuador to show that the historical conditions did not appear to be in place for Spanish creoles to form and eventually decreolize.

In order to provide the reader with a look at YS and CS, I present here two brief speech excerpts extracted from my Afro-Andean corpus (1–2). They are meant to illustrate how, besides certain morphological and phonological reductions, current YS and CS approximate quite closely to Spanish; thus, they do not present the traces of the more intense grammatical restructuring that is generally associated with creole languages.

(1) *YS*
Todo[toda] la comunidad participaba; mucha gente
all-M.SG the-F.SG community-F.SG participated much-F.SG people-F.SG
venía, mucho[mucha] gente venía desde lejos. todas, toditas
came much-M.SG people-F.SG came from far all-F.PL all-F.PL
las personas se reunían. Muy bonito[bonita] la
the-F.PL people-F.PL REFL met very nice-M.SG the-F.SG
fiesta era . .. (Sessarego 2014: 127)
party-F.SG was
"All the community used to participate, many people used to come, many people used to come from far away. All, all the people gathered. The party was very nice . . ."

Map 1. Yungueño Spanish and Choteño Spanish (indicated by the red circles) in the context of the Afro-Hispanic vernaculars of the Americas (adapted from Klee & Lynch 2009: 6)

(2) CS
Todo[toda] la familia se iba con los
all-M.SG the-F.SG family-F.SG REFL. go-PAST with the-M.PL
amigo[amigos], con toda la gente de Concepción; cuando
friend-M.SG with all-F.SG the-F.SG people-F.SG of Concepción when
yo era pequeño mucha[muchas] persona[personas] Rezaba[rezaban],
I be-PAST small much-F.SG person-F.SG pray-PAST
mucho[mucha] devoción tenían. (Sessarego 2013b: 71)
much-M.SG devotion-F.SG have-PAST
"The whole family used to go with friends, with all the people from Concepción; when I was young many people used to pray, they used to be very devoted."

Given the lack of standardization and the stigmatization attached to the traditional speech of these Afro-Hispanic groups, much linguistic variation can be detected across speakers. As reported in other studies (Sessarego & Gutiérrez-Rexach 2012; Gutiérrez-Rexach & Sessarego 2014), extra-linguistic factors such as education and age may significantly affect the way the members of these communities speak. The two aforementioned Examples (1)–(2) proceed from the speech of two elderly and illiterate informants; younger and more educated speakers tend to use more standard-like forms. It is also important to highlight that, overall, YS is more grammatically restructured than CS (Lipski 1987, 2006). Indeed, the presence in the Bolivian dialect of certain grammatical features (i.e., genderless pronoun system, verbs consistently inflected for tense but not person/number, verbal constructions of the type *ta* + invariant verb, etc.) has fueled an animated debate on its potential creole past (Lipski 2008; Sessarego 2013c, 2016; Schwegler 2014; Perez 2015). In more recent studies, some authors have claimed that the reason for the contrasting views on the origin of YS may be due to different data collection methodologies (Perez 2016; Perez, Sessarego & Sippola 2017). Moreover, in some cases, answers to these questions have also been sought beyond the analysis of morphosyntactic data, such as in an examination of prosodic patterns in YS and CS (Rao & Sessarego 2016, 2018).

Contact-driven restructuring operates on a cline (Thomason & Kaufman 1988). As for the Afro-European contact vernaculars spoken in the Americas, some varieties that have traditionally been labeled as "creoles", such as Haitian Creole, Sranan Tongo and Palenquero, have gone through more intense processes of grammatical restructuring, and thus their grammars diverge quite significantly from their European lexifiers, to the point that mutual intelligibility is greatly reduced. Other varieties, such as CS and YS, are not so grammatically divergent from Spanish, as shown in (1–2), and for this reason, some authors would claim that, in their current state, they align more closely with Spanish "dialects" than with Spanish "creoles"

(McWhorter 2000; Sessarego 2019a, 2019b). The issue, therefore, is to understand if they have always been like this, or if they used to be creoles and then decreolized. To address this question, it is of fundamental importance to cast light on the social histories of these black communities.

In the following sections, I will show that, based on the available sociodemographic evidence for the colonial Andes, YS and CS should not be analyzed as the result of a decreolization process. On the contrary, this historical analysis appears to corroborate the linguistic findings reported in Sessarego (2013b, 2014), suggesting that these dialects were never creoles. Rather, they may be better conceived as the result of advanced L2 processes, which were conventionalized at the community level in two isolated rural zones, far away from the pressure exerted by schooling, urban society and the standardizing norm.[1]

These two Afro-Andean communities, in fact, until well into the 20th century were quite isolated and surrounded by indigenous populations that spoke little Spanish: in the case of YS, the surrounding indigenous group spoke almost exclusively Aymara; as for CS, the population of the Chota Valley primarily consisted of Quichua speakers (Lipski 1987, 2008). Given this peculiar context, it may well be said that YS and CS bear sociohistorical and linguistic similarities to Judeo Spanish in the eastern diaspora (i.e., Sephardic Spanish in Istanbul) (Romero & Sessarego 2018). Indeed, all of these dialects have survived for centuries as Spanish enclaves surrounded by non-Spanish speakers. All three of the varieties, as a result, present parallel universal contact-induced features, which should not necessarily be ascribed to the specific languages with which they have been in contact (i.e., Aymara, Quichua and Turkish) (Sessarego 2019b).

This paper offers an analysis of the social dynamics that characterized the life of black communities in the Andes across the colonial phase up to the present. It demonstrates how certain external factors may have contributed to shaping the contact varieties they speak today.

---

1. An anonymous reviewer wonders what the effects of the substrate African languages (L1s) might have been on YS and CS. Given a concomitance of historical and logistic factors, it is almost impossible to reconstruct the exact linguistic composition of the slave gangs that were taken to the Andes during the colonial phase. Historical records suggest that captives proceeded from many different regions across the western African coast; thus, they presumably spoke a variety of African languages (i.e., Bran, Mandinga, Fula, etc.) (Sessarego 2013d). Nevertheless, as I have suggested elsewhere, besides certain lexical borrowings from African languages (Sessarego 2011, 2013b), most of the features currently encountered in CS and YS should not necessarily be ascribed to a specific substrate; rather, they consist of universal, contact-induced, grammatical patterns, which tend to occur in most Spanish contact varieties (e.g., high rates of overt subjects, reduced phi-agreement, etc.) (Sessarego 2019b).

Section 2 offers an overview of the hypotheses that have been proposed to account for the origins and evolutions of these vernaculars. Section 3 is an analysis of sociohistorical data showing that, overall, the Andean region was not suitable for the formation and/or preservation of creole languages. Section 4 focuses on the two Afro-Andean communities under analysis to show that, in these specific cases, a (de)creolization hypothesis is also not sustainable. Finally, Section 5 summarizes this work and provides the concluding remarks.

2. Decreolization and Afro-Andean Spanish

Granda's (1968) Decreolization Hypothesis had a vast impact on the field of Afro-Hispanic linguistics. In fact, this model has been proposed over and over to account for a number of the linguistic features found in almost every contemporary Afro-Hispanic vernacular (Otheguy 1973; Schwegler 1991a, 1991b; Megenney 1993, Álvarez & Obediente 1998, among many others). The latest instance I have encountered of this long-lasting claim on the origin and evolution of the Afro-Latino varieties of the Americas has been provided by Guy (2017), who selects YS to exemplify this hypothetical (de)creolization process, which, in his view, affected many other Afro-Hispanic and Afro-Lusophone vernaculars, including Popular Brazilian Portuguese (Guy 1981, 2004). Guy (2017: 72) states:

> Its history of linguistic isolation implies that [YS] must be more basilectal, closer to the speech of the earliest generations of Africans in the Americas, than Brazilian Portuguese and Caribbean Spanish. This in turn implies a historical trajectory by which all of these varieties started out as creoles, or at least restructured varieties tending toward the creole end […], and then acquired their present form through differing degrees of standardization.

The explanations proposed in the literature to account for the evolution of YS and CS exemplify this half-century-long (de)creolization approach to the formation of the Afro-Latino varieties of the Americas. Lipski (2008: 186), for example, indicates that "in the absence of any other viable scenario, Afro-Yungueño Spanish must be viewed as the descendant of a colonial Afro-Hispanic pidgin". Perez (2015) also hypothesizes that YS derived from a restructured pidgin/creole, and claims that this vernacular would have originated from a Portuguese-based contact variety. The Portuguese-creole view on the origins of several Afro-Hispanic vernaculars has also been suggested on numerous occasions by Schwegler (1991a, 1991b, 1993, 1999), who, in a recent study (Schwegler 2014) has re-proposed the same evolutionary path for several contact varieties, including both YS (2014: 426–429) and CS (2014: 420–426).

It has to be said that a Portuguese creole origin for CS has been envisioned by Schwegler since the late 1990s (Schwegler 1999). In that study, as in Schwegler (2014), he states that the presence in this dialect of the apparently Portuguese third-person pronoun, *ele*, would represent evidence of an Afro-Lusophone creole substrate involved in the formation of this variety. Conversely, Lipski (2009) appears to be less convinced of a potential decreolization process for this particular Afro-Andean vernacular. He provides a different account to explain the presence of *ele* in CS by suggesting that this element should be seen as the result of a paragogical process of final *-e* insertion (i.e., *él* + *e*), quite widespread across the CS lexicon (2009: 113).

Against the (de)creolization model for CS is McWhorter (2000), who claimed that this Afro-Andean variety was never a creole, since, in line with his Afrogenesis Hypothesis, Spanish never creolized in the Americas. In fact, in his much-debated book, *The Missing Spanish Creole*, McWhorter (2000) claims that creole languages developed out of pidgins, which originally formed in Africa in the castle factories where slaves were imprisoned before being shipped to the Americas. Given that Spain did not have slave castles along the western African coast, an Afro-Hispanic pidgin never formed in Africa. Consequently, a Spanish creole could not possibly develop on the other side of the Atlantic, in Spanish America, even though in certain Spanish colonies, such as Chota Valley (2000: 10–11), the sociodemographic conditions for creole formation would have been perfect: (a) thousands of African-born slaves speaking a variety of African languages; (b) minimal presence of white people speaking Spanish; (c) harsh working conditions in sugarcane plantations; (d) a highly isolated region, far away from Spanish-speaking urban centers.

In the following sections we will see how the available sociohistorical information for the colonial Andes does not support either the Decreolization Hypothesis or the creolizing conditions envisioned by McWhorter's Afrogenesis Hypothesis. Quite conversely, historical data indicate that the living conditions of the slaves working in the communities under analysis allowed the captives to obtain relatively good access to the colonial language, thus explaining why YS and CS do not show the traces of the more intense grammatical restructuring characterizing creole grammars.

3. A sociohistorical sketch of black slavery in the Andes

Black slavery was introduced to the Andes from the very beginning of the Spanish colonization of the region (in the early decades of the 16th century) and lasted for more than 300 years, until the middle of the 19th century. Nevertheless, even after the abolition of slavery, Afro-Andeans did not obtain real freedom. On the contrary, in most cases, their living and working conditions did not improve, since – in

practice – many of them did not find other occupations and had to keep working for their former masters as unpaid peons. They had to wait until the Land Reforms, which took place at different points during the second half of the 20th century (1953 in Bolivia and 1964 in Ecuador), to see certain improvements in their living and working conditions (De Janvry, Sadoulet & Wolford 1998). This meant becoming owners of small lots of land that belonged to the haciendas in which they worked, obtaining access to education and the possibility to move without being bound to the plantation. Even if over time the social conditions of Afro-Andeans have improved, their situation is far from being optimal, since they still represent the most discriminated sector of Andean society (Hernández 2013).

3.1 First arrivals (16th century–mid-17th century)

The black presence in the Andes has been reported since the earliest Spanish colonial expeditions in the region, when Afro-descendants participated side-by-side with the Spaniards in the first exploring and conquering missions. Blacks played an active role in this early colonial period. Bryant (2005: 1), in his dissertation on the African Diaspora to Ecuador, points out how Afro-descendants were among the first founders of what would become Ecuador's capital city: Quito. He approaches the topic in this way:

> Inscribed on the outside walls of Quito's Cathedral is a list of the men who invaded, occupied and ultimately renamed the Inka city of Quito for Saint Francis on August 28, 1534. Listed among these individuals are two blacks – Juan (*de color negro*) and Antón (*negro*). Juan and Antón represented the hundreds (and perhaps thousands by this date) of black explorers, conquistadors, slaves and squires who had come to the Americas during the age of conquest (1492–1550). Although the biographical records for Juan and Antón are sketchy, sources indicate that they were freemen.

Brockington (2006: 130) concurs on the active involvement of several Afro-descendants in this phase. When describing the black presence in Bolivia, the author states the following:

> The African slaves and people of African descent were active participants – voluntary or otherwise – in a militaristic, conquering/pacifying, horse-and-gun culture here, as elsewhere in the Spanish Americas at that time. I am convinced that some of them remained […] as, among other things, slave and free cowboys and ranch hands.

In order to understand what languages these people might have spoken, it is important to keep in mind that, during this early colonial phase, the slaves entering the Andes did not usually proceed directly from Africa (Bowser 1974; Mellafe

1984). Rather, they came for the most part from already-settled colonies (in the Caribbean) or from Spain. On this point, Restall (2000: 190–219) reminds us how the Spanish Caribbean, and in particular Hispaniola, became a place where – as indicated by a local royal official at that time – slave traders could make "a living by buying Africans, teaching them some trade and then selling them at a profit on the mainland" (see Aguirre Beltrán 1946: 20). This implies that before being re-sold on the mainland, they often spent long periods of time (sometimes a number of years) acquiring job skills (along with the Spanish language) in the Caribbean.

This observation is backed by demographic records. Bowser (1974: 72–73), for example, shows how the majority of the slaves sold in Lima, Peru during this early phase (1560–1650) were not *bozales*[2] proceeding directly from Africa. Rather, out of a total sample of 444 black captives, 369 individuals were brought from other Latin American colonies, where they had resided for some time (possibly several years), and thus had learned the Spanish ways before being sold in Peru.

Even though black people were present and visible in the Andes from the beginning of this early colonial phase, their introduction was not massive, since they were expensive to acquire. As Lockhart (1994) points out, big cargos of black slaves were not common. Single transactions concerning more than ten slaves were rare (1994: 177); captives were generally acquired one at the time, while sales of two or three captives were more sporadic (1994: 178). The reasons for these trading limitations in the region were multiple (Sessarego 2015: Chapter 5). Three main factors may be seen as particularly crucial: logistic constraints, the Crown's monopoly of the slave trade, and the availability of an indigenous workforce.

As for the logistic constraints, it should be kept in mind that the countries now known as Ecuador and Bolivia do not have any direct access to the Atlantic Ocean. This means that taking African-born slaves to these lands implied a particularly strenuous trip, which systematically resulted in many casualties among the captives. As a consequence of such a geographical barrier, the price of a slave, who could be purchased for some 200–400 pesos in the ports of Cartagena (present-day Colombia) or Buenos Aires (present-day Argentina) during the 16th–17th centuries, could easily reach 500–600 pesos in Lima (present-day Peru), and the price would tend to increase in the Ecuadorian and Bolivian highlands, which were located further away from the major slave markets (Bowser 1974; Colmenares 1997; Brockington 2006).

An additional reason for the high price of slaves had to do with the fact that black captives were not traded on a free market; rather, this was a highly constrained business. The Spanish Crown held a monopoly on the introduction of slaves to its

2. African-born slaves.

American colonies and granted only a limited number of *asientos* 'licenses' to trading companies. In addition, for each slave transaction, the Crown charged *almojarifazgos* 'import taxes' and *alcabalas* 'sales taxes' (Bryant 2005). Such restrictions on the introduction of an African workforce to Spanish America resulted in the black population being a relatively small group in several colonies, especially when compared to the native population (Rosenblat 1954).

The relative availability of a native workforce, in fact, is certainly a factor that further determined the non-massive introduction of black captives to the Andes. Indeed, even though Indians could not technically be enslaved according to the Spanish legal system of the time (Sessarego 2018), they were forced to work in precarious conditions for minimal compensation. To achieve this, the Spaniards adopted a pre-Columbian working system implemented by the Incas, the *mit'a*. According to the *mit'a*, each man of a given Indian community had to carry out compulsory duties for a certain period of time. This system worked on a rotation basis according to which nobody would be required to work again until everyone covered his own shift (Rostworowski Tovar de Díez Canseco 1999).

It should also be pointed out that the Spanish Crown was not much interested in the development of a large-scale plantation business of the kind seen in the French and English colonies in the Caribbean. Rather, in the Andes – as in most of Spanish America – the main concern of the Spanish administration had to do with the extraction of precious metals, primarily silver and gold. The Andean region was particularly rich in this sense. In 1545, the biggest silver deposit of the Americas was discovered in the mountains of Potosí (current-day Bolivia). The high costs of black slaves (especially in the highlands) induced the Spanish administration to extensively rely on the native population to extract the precious metal. In 1578, under the administration of Viceroy Toledo, it was established that 14,248 native men would serve their *mit'a* shift in the silver deposit of Potosí, the *cerro rico* 'rich mountain' (Bowser 1974: 40).

The reduced presence of an Afro-descendant group in the highlands is further supported by the available demographic information of the time. Crespo (1995: 26–29) reports data concerning the evolution of the black population for the city of Potosí for the 17th–19th centuries (see Table 1). As it can be observed, the Afro-descendant group never represented more than a fraction of the total population.

Table 1. Demographic figures for the city of Potosí (1611–1832)

Year	Afro-descendant population	Total population	Afro-descendant population (percentage)
1611	6,000	160,000	3.75%
1719	3,206	70,000	4.58%
1832	1,142	224,000	0.51%

While in the highland mining regions the presence of the black population was minimal during this first phase, in the main urban centers, and especially in cities like Lima and Quito, the Afro-descendant group was much more visible. The vast majority of these people were in all likelihood Spanish speakers, since many were *ladinos* and *criollos*, and only a smaller part were *bozales* (Bowser 1974). The increasing presence of locally born *criollos* is further attested by Bryant (2005: 78), who shows how between 1580 and 1600 only 25.8% of the slave transactions in Quito concerned African-born captives. He states that, given the geographical and financial constraints affecting the importation of slaves into the Andes, it was common practice among slave owners to incentivize local reproduction. Providing indirect evidence of this fact, the price of black women on recorded transactions was often established according to their potential capability of "producing a good number of Afro-Creole children" (2005: 79).

During this first phase of conquest and colonization, the agricultural sector, which was not based on large-scale plantations, relied primarily on the native workforce. However, over time, Afro-descendants came to populate the warmer regions of Ecuador and Bolivia, and to complement the indigenous population, which had been shrinking during the 16th and 17th centuries due to wars, European diseases, and the harsh working conditions imposed by the *mit'a* system (Kenneth 1995; Brockington 2006).

3.2 The second wave (mid-17th century-last decades of the 18th century)

As indicated above, three main factors constrained the introduction of an Afro-descendant workforce in the Andean region: geographical barriers, the Crown's monopoly of slave trafficking, and the relative availability of indigenous workers. Over time, the native population shrank and the perceived need for black labor became more significant. Nevertheless, the logistic impediments and the financial restrictions related to the lack of a free market for slave trading did not change during this second phase, and the introduction of a black workforce in the Andes never achieved massive dimensions.

Even though the agricultural sector gradually developed and some plantations achieved large-scale dimensions in certain Spanish colonies, for the most part the economy of the territories now known as Ecuador and Bolivia was based on small and middle-sized haciendas, which did not rely on the extensive exploitation of an enslaved black workforce. The reasons behind the limited use of black captives had to do, once again, with their high cost. In fact, laymen did not usually have access to enough capital to purchase slaves in large numbers; the only institution that disposed of such financial resources at that time was the Catholic Church and, in particular, the Company of Jesus.

Macera (1966) provides a detailed analysis of how Jesuit haciendas were run. He shows that the Jesuits were able to obtain high production rates from the enslaved workforce while creating a working and living environment that would minimize the risks of revolt by generating a feeling of community among the captives. They supported the creation of nuclear families and incentivized local births, encouraging marriages by keeping an almost equal number of men and women in each plantation. Slaves were provided with high caloric foods so that they could tolerate the intense workload typically related to sugarcane growing. Each family was also assigned a small piece of land, called a *chacra*, on which they could grow their own crops and raise animals. This custom, besides providing slave families with extra food, created a sense of ownership and belonging to the local community.

Captives were also systematically Christianized. The Jesuits devoted much energy to this task. Macera (1966: 30) describes several activities that had this purpose: *bozales* were methodically catechized, children had to attend daily religion classes, everyone was required to go to Mass on Sundays and during sacred festivities (i.e., Christmas, Easter, etc.), and religious ceremonies (i.e., baptisms, weddings, funerals, etc.) were regularly celebrated. It appears that the emphasis placed on Christian indoctrination also served as a means by which captives had significant exposure to the Spanish language. As a result, it may be inferred that language acquisition was facilitated.[3]

3.3 The gradual path to emancipation (last decades of the 18th century–present)

Even though the Jesuit management was family-oriented and tried to create a sense of community among the captives as a way of reducing the potential risk of rebellions, it has to be said that the working conditions in these plantations could be harsh and that corporal punishment was often applied to the slaves who did not follow the rules (Cushner 1980: 89–90). Unfortunately, after the expulsion of the Company of Jesus from Spanish America in 1767, the lives of Afro-descendants working on those plantations did not improve. At that point, in fact, the Jesuit

3. An anonymous reviewer wonders whether the Jesuits would use African languages to Christianize slaves in the Andes, since it is well-known that they learned and used indigenous languages to teach native populations. Given the multitude of African languages that arrived to Spanish America and the fragmentation of the African groups that were taken to the Andes, it is likely that the primary language used to interact with the slaves was Spanish. Nevertheless, it is also attested that the Jesuits would often rely on captives capable of speaking Spanish as well as African languages (*ladino* slaves) to communicate with *bozales* and to teach them the Spanish language over time (Sessarego 2013d).

haciendas (and the captives working on them) passed under the control of the *Temporalidades* council, a public institution in charge of selling the properties confiscated from the Company of Jesus to private buyers.

Historical records show that the new administration did not care much about the well-being of the slaves and the preservation of the traditional community norms implemented by the Jesuits. On many occasions the *Temporalidades* attempted to take away the *chacras* from the captives, and in some cases the integrity of slave families was put at serious risk, since the new administration tried to sell family members individually, as single tokens. The new management under the *Temporalidades* led to a series of revolts across the former Jesuit plantations. These violent episodes, combined with the *Temporalidades*' general lack of business skills, resulted in a disastrous effect on the agricultural sector of the colonies. The Jesuits had built a well-administrated economic empire, which fell apart as soon as the *Temporalidades* started managing it (Andrien 1995).

The following decades were characterized by tensions between Spain and its Latin-American colonies, which entered into a series of independence wars. After years of conflicts between different Latin-American armies and the Spanish forces, the territories now known as Ecuador and Bolivia managed to become independent in 1822 and 1825, respectively (Luna Desola 1978: 88–89). Subsequently, these new countries went through convoluted phases of slavery abolition (Clementi 1974). In principle, slavery was abolished in Bolivia in 1826 and in Ecuador in 1851; however, in practice, for more than a century the Afro-descendant population of the Andes did not enjoy the same rights and freedoms as the white and *mestizo* citizens living in those regions (Hernández 2013).

This situation is well-explained by Hassaurek (1867: 328–329), who visited Ecuador from 1861–1865, and used the following words to describe the living conditions of the Afro-descendants of Chota Valley:

> The Indians have entirely disappeared from the valley. The Negroes, who have taken their places, are *concertados* […]. They are slaves in fact, although not slaves in name. Their services are secured by a purchase of the debts which they owe. As long as they remain in debt, which state, thanks to the skillful management of their masters, almost always lasts till they discharge the great debt of nature, they must either work or go to prison. Like the Indians, they are ignorant of their legal rights. They are hardly ever able to pay their debts, which, on the contrary, continue to increase, as their wages of one half real to one real are insufficient to satisfy their wants. When slavery was abolished in Ecuador, the owners of the Negroes in the sugar district immediately employed them to work for wages, and managing to get them into debt, secured their services as debtors. Thus it may almost be said that they profited, instead of losing by, the abolition of slavery. They pocketed the

> compensation which the law provided for the slave owners and at the same time retained the slaves. It is true that the blacks do not work so much now as when they were bondmen, nor can their masters beat them as unmercifully as they did before; but, on the other hand, it must be considered that it is much cheaper now to purchase a Negro than it was then. Now, by paying a debt of fifty or seventy dollars which the poor fellow owes to somebody, his services may be secured, while formerly it took, perhaps, ten times that amount to purchase a slave.

The preceding quotation offers a good description of the living conditions of most Afro-descendants after the abolition of slavery, not only in Ecuador, but also in Bolivia. Unfortunately, until the advent of the Land Reforms, which took place in the second half of the twentieth century, the vast majority of the blacks living in rural areas kept working on haciendas as peons receiving wages under the level of basic sustentation. Legally, they were not slaves, but in practice, they still had to work almost for free and were not allowed to leave the hacienda without the owner's permission.

4. A closer look at YS and CS

Having provided the general sociohistorical background that characterized the African Diaspora to the Andes, we will now turn our attention to these two specific Afro-Andean communities: Yungas (Bolivia) and Chota Valley (Ecuador). These specific analyses will offer a view of two different, but to a large extent parallel, rural settings, and thus help us cast light on the potential origin of the Afro-Hispanic vernaculars spoken in these regions.

4.1 Yungas, Bolivia

The Yungas are tropical valleys located in the Department of La Paz, Bolivia. It is still not completely clear when the first black people arrived to this region; however, in a report of 1805, the overseer of the local Mururata hacienda, Mr. Francisco Xavier de Bergara, mentions that the owner, the Marquesa de Haro, was the first landlord relying on a black workforce in the area (Crespo 1995: 105). This note, therefore, may be taken as a clue suggesting that the first significant presence of an enslaved black population in these valleys might date to the end of the 18th or the beginning of the 19th century.

This hypothesis appears to also be indirectly supported by the existence of several documents of that period concerning purchase transactions of just a few slaves at a time. For example, Portugal-Ortiz (1977: 78) reports the sale of a married

couple from Angola,[4] who were purchased by a priest living in Chulumani in 1761, as well as the purchase of a young girl who was employed by a Yungan landowner in 1773 to carry out agricultural work in a local hacienda. These small sales, including only a few captives in each transaction, are perfectly in line with the overall historical records concerning the Black Diaspora to Bolivia. Several historians, in fact, as we saw, concur that the logistic and financial constraints on slave trading did not allow for a massive introduction of a black workforce to the agricultural sector of the colony. Rather, slave transactions usually involved no more than one or two slaves at a time (Lockhart 1994; Bridikhina 1995). In line with this constrained black presence in the region are the data provided by Crespo (1995: 96), who reports on a document written by a local priest working in the village of Chirca in 1802. He provides the number of Afro-descendants and natives living in the local haciendas. Of these small plantations, only four used a black workforce. They were Guayraoata, with 15 Afro-descendants (blacks and mulattoes) and 50 Indians, San Agustín (17 Afro-descendants and 28 Indians), Yacata (23 Afro-descendants and 128 Indians), and Collpar (28 Afro-descendants and 142 Indians).

Not much more demographic information is available for the Yungan haciendas until 1883. For that period, Portugal-Ortiz (1977: 87–91) provides a more detailed picture of two local villages: Pacallo and Mururata. In Pacallo there were 67 whites, 63 *mestizos*, 340 Indians and 56 Afro-descendants; Mururata counted 55 whites, 183 *mestizos*, 236 Indians and 324 Afro-descendants. He also reports, without categorizing the inhabitants into specific ethnic groups, the villages of Chulumani, with 14 Afro-descendants out of a total population of 220; Tajmo, Calupre, Chigno, Chimasi, Tolopala, Suquillo with 49 out of 902; Coroico with 113 out of 5,335; Impata with 252 out of 2465; Coripata with 315 out of 3,867; Chupe with 240 out of 1212; and Lanza with 102 out of 8,255.

When we analyze this demographic information for the 18th and 19th centuries, what emerges is certainly not a large-scale plantation society that relied extensively on a black labor force. On the contrary, the Afro-descendant group appears to be a quite constrained minority. Even in Pacallo and Mururata, the haciendas with the highest percentages of blacks and mulattoes, there are no large disproportions between the Afro-descendants and the white/*mestizo* group (Pacallo: 56 blacks/mulattoes v. 130 whites/*mestizos*; Mururata: 324 blacks/mulattoes v. 233

4. An anonymous reviewer wonders whether these slaves came directly from Angola or if Angola was their ethnic name (also see Álvarez López 2015 on this point). Portugal-Ortiz (1977: 78) simply indicated that in their official sale records they were enlisted as 'Angola', without specifying if that noun referred to the slaves' ethnicity or to the region/port from which they had been shipped from Africa.

whites/*mestizos*). Data indicate that only in Mururata the black/mulatto group outnumbered the whites and *mestizos*. If we consider that the mulattoes were probably Spanish speakers (since, by definition, one of their parents was white) and that even the blacks at this point had probably not arrived to the Yungas directly from Africa, the chances of language creolization in such a context are quite slim.[5]

The analysis of 19th-century Yungas provided by Busdiecker (2006: 38) is consistent with the aforementioned demographic data. She describes a system of small and medium-sized haciendas, where slaves were not the only source of labor, but rather they were complemented by both Indians and freed blacks, who received different forms of compensation for their work. In addition, data indicates that manumission was not rare. In fact, even though the working conditions were harsh, some slaves managed to save capital and buy their own freedom over time (Portugal-Ortiz 1977: 69). The most famous case of self-manumission is the story of King Bonifaz. He became well-known for helping other captives buy out their own freedom. For this reason, in the 19th century, he was symbolically proclaimed King of the Afro-Bolivian population by the members of the local community. This symbolic monarchy is still in place today, and the current King, Julio Pinedo, is an inhabitant of the local Mururata village (Sessarego 2011: 41).

4.2 Chota Valley, Ecuador

Chota Valley is a tropical region located on the frontier between the Province of Imbabura and Carchi, in the northern Ecuadorian Andes. The first record of an Afro-descendant presence in this area dates back to a letter written in 1582 by a local priest, Antonio de Borja. In this document, he stated that in the valley, which by that time was inhabited for the most part by native people, there were six Spaniards, who owned a few slaves working on vineyards.

5. An anonymous reviewer wonders what would be needed in demographic terms to have a context leading to language creolization. There is no clear answer to such a question. The first scholar to explicitly try to quantify the demographic disproportion needed for pidginization in plantation societies was Bickerton (1981: 4), who suggested a 4:1 ratio as an ideal scenario for a 'radical' creole to develop. Language restructuring operates on a cline in which the demographic factors play a major role. For this reason, I would suggest that the 324:233 demographic figure reported for Mururata is not likely to generate a highly restructured Spanish variety. Moreover, we should not confuse the category 'blacks' with 'Africans'. These are two very different groups. In fact, given the history of the slave trade to the Andes, it is likely that only a fraction of the people classified as 'blacks' in Mururata actually came directly from Africa. In line with historical research, most of them were probably locally born or from other Spanish colonies, thus they were not necessarily speakers of African languages. They might have been black people, born in Spanish America, who spoke vernacular varieties of Spanish (Sessarego 2011, 2013c, 2014).

As for this initial phase, historical data appear to suggest that blacks occupied an intermediate position between the Spaniards and the natives, thus they might have had relatively good access to the Spanish language. They were a minority group, often employed by the settlers to exert violence against the local native populations, as indicated in a report of the time quoted by historical Coronel Feijóo (1991: 69):

> Mayordomos, esclavos negros y propietarios a punta de piedras, látigo, rejo y palo, comienzan a imponer nuevas normas de distribución del agua sobre las antiguas reglas indígenas […] andan con palos, rejones, perros, rodando y aguardando el agua (Supervisors, black slaves, and owners impose new water distribution rules on old Indian norms; they use stones, whips, spikes, and sticks […] they walk around with sticks, big spikes, and dogs to control water resources).

Coronel Feijóo (1991) has also shown how by the beginning of the 17th century the Company of Jesus managed to acquire properties across the valley and to gradually implement a shift in labor force from native workers to Afro-descendants. Such a shift was primarily driven by the fact that Indians systematically fled. This progressive introduction of black workers, nevertheless, was not massive, due to the financial constraints that affected slave trade, as Coronel Feijóo (1991: 81) clearly points out:

> Hablar de importación masiva de negros, para la época, parece sobredimensionado; difícil resulta atribuir a los estancieros de la zona un negocio de tal magnitud (talking about massive black importation, by that time, seems to be overstated; it is difficult to ascribe such a big business to local settlers).

By the end of the 17th century, the Jesuit agricultural enterprise expanded and, as a result, it began to rely more significantly on enslaved workers. Historical analyses of the Jesuit haciendas suggest that the majority of the captives used on these farms were locally born *criollos* (Peñaherrera de Costales & Costales Samaniego 1959: 215, Coronel Feijóo 1991: 93). This statement may be further supported by the data provided by Colmenares (1997) on the number and origins of the captives sold during the same period in Popayán (current-day Colombia), the most important slave market of the northern Andes (Bryant 2005). Colmenares (1997: 35) shows that, even during the peak of the slave trade in the 1730s, 60% of slave transactions involved captives who were born in Spanish America.

According to data reported by Cushner (1982: 136), by 1767, the year in which the Jesuits were expelled from their Chota Valley haciendas, there were a total of 1,364 slaves, of whom 488 were children under the age of ten and 94 were classified as "too old to work". This demographic picture suggests, on the one hand, that birth rates were probably high (since small children were in all likelihood locally born and not shipped directly from Africa) and, on the other hand, it indicates that life expectancy was probably quite long (given that many slaves achieved an elderly age).

As described for many other Jesuit haciendas across Latin America, in Chota Valley the Company of Jesus also incentivized local slave reproduction and self-maintenance (Bouisson 1997: 48–51) by ensuring an equal number of men and women in each plantation and support for nuclear families through marriage. Additionally, each family had an independent house and a *chacra*, slaves could grow their own products during their time off and sell them at the local market, and they could accumulate capital and thus purchase their own freedom over time. Coronel Feijóo (1991: 110) even suggests that by the end of the 18th century there is evidence of black people renting out their *chacras* to white and *mestizo* workers.

5. Final remarks

This paper has analyzed aspects of the Black Diaspora to the Andes to explore whether YS and CS formed in a context that was suitable for Spanish creolization, as indicated by several authors (Schwegler 1999, 2014; McWhorter 2000; Lipski 2008; Perez 2015). Results indicate that a concomitance of sociodemographic factors significantly reduced the possibility of Spanish creoles forming or being preserved in the colonial Andes. The agricultural sector consisted for the most part of small and middle-sized estates, as geographical and economic constraints did not favor the introduction of a massive black workforce in these territories. Even when larger agricultural complexes existed, workers were for the most part *ladino* and *criollo* captives, who presumably spoke vernacular varieties of Spanish. Private investors did not usually have enough capital to purchase slaves in significant numbers. The largest slave-owner institution which could actually rely more systematically on an enslaved workforce was the Company of Jesus, which, due to several social and religious practices, favored the acquisition of Spanish in its haciendas (Macera 1966).

A closer look at black slavery in Yungas indicates that Afro-descendants came to populate these valleys gradually, in all likelihood since the 18th century. Data show that a large-scale plantation society was never in place in this region: the black/mulatto group never represented the majority of the population and the workers of these farms were not just slaves; rather, the enslaved workforce was systematically complemented by paid workers. Working conditions were probably not as harsh as in other American colonies. In fact, it appears that black captives had chances to accumulate some capital and eventually buy their own freedom. All in all, this scenario suggests that Afro-descendants probably had relatively good access to the Spanish spoken by the white/*mestizo* population. Moreover, given the geographical and financial constraints that characterized slave trafficking in the Andes, in all likelihood a good percentage of Yungan slaves were *criollos* rather than African-born *bozales*.

The historical information concerning Chota Valley also indicates that those haciendas diverged quite significantly from the typical plantation societies documented for the English and French Caribbean. Unlike those societies, colonial Chota Valley was made up of small farms in which a certain degree of social flexibility and chances to improve one's living conditions were provided to the members of the Afro-descendant group. These elements, in addition to the fact that most of the captives were in all likelihood locally born, suggest that the (de)creolization hypothesis is not the most feasible explanation for the nature of CS grammar.

All of these factors significantly reduce the possibility of a creole language forming in colonial Yungas and Chota Valley and therefore decrease the likelihood of YS and CS being the result of a (de)creolization process. The present study not only sheds light on the nature of these two Afro-Andean dialects; more broadly, it helps provide a new perspective on the evolution of the Afro-Hispanic languages of the Americas. It does so by showing that the long-lasting Decreolization Hypothesis should not be taken at face value. Rather, it should be systematically tested against the available sociohistorical evidence we have for other black communities in Latin America in order to better understand the origins of the contact varieties they speak.

Acknowledgments

Netherlands Institute for Advanced Study, Institut Universitari de Drets Humans de la Universitat de València.

References

Álvarez, Alexandra & Enrique Obediente. 1998. El español caribeño: antecedentes sociohistóricos y lingüísticos. In Pearl Matthias & Armin Schwegler (eds.), *América negra: panorámica actual de los estudios lingüísticos sobre variedades hispanas, portuguesas y criollas*, 40–61. Madrid: Iberoamericana.

Álvarez López, Laura 2015. Who named slaves and their children? Names and naming practices among enslaved Africans brought to the Americas and their descendants with focus on Brazil. *Journal of African Cultural Studies* 27(2). 159–171. https://doi.org/10.1080/13696815.2014.992396

Aguirre Beltrán, Gonzalo. 1946. *La población negra en México: estudio etnohistórico*. México DF: Ediciones Fuente cultural.

Andrien, Kenneth. 1995. *The Kingdom of Quito, 1690–1830: The state and regional development*. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511529054

Bickerton, Derek. 1981. *Roots of language*. Ann Arbor: Karoma.

Bouisson, Emmanuelle. 1997. Esclavos de la tierra: Los capesinos negros del Chota-Mira, siglos XVII–XX. *Procesos, Revista Ecuatoriana de Historia* 11. 45–67.

Bowser, Frederick. 1974. *The African slave in colonial Peru*, 1524–1650. Stanford: Stanford University Press.

Bridikhina, Eugenia. 1995. El tráfico de esclavos negros a La Paz a fines del siglo XVIII. *Estudios Bolivianos* 1. 183–191.

Brockington, Lolita. 2006. *Blacks, Indians, and Spaniards in the eastern Andes*. Lincoln & London: University of Nebraska Press.

Bryant, Sherwin K. 2005. *Slavery and the context of ethnogenesis: African, Afro-Creoles, and the realities of bondage in the Kingdom of Quito, 1600–1800*. Columbus, OH: The Ohio State University dissertation.

Busdiecker, Sara B. 2006. *We are Bolivians too: The experience and meaning of blackness in Bolivia*. Ann Arbor, MI: University of Michigan Dissertation.

Clementi, Hebe. 1974. *La abolición de la esclavitud en América Latina*. Buenos Aires: Editorial La Pléyade.

Colmenares, Germán. 1997. *Historia económica y social de Colombia–II Popayán una sociedad esclavista 1680–1800*. Cali: TM editores, Universidad del Valle, Banco de la República, Colciencias.

Coronel Feijóo, Rosario. 1991. *El valle sangriento de los indígenas de la coca y el algodón a la hacienda cañera jesuita, 1580–1700*. Quito: ABYA-YALA.

Crespo, Alberto. 1995. *Esclavos negros en Bolivia*. La Paz, Bolivia: Librería Editorial Juventud.

Cushner, Nicholas. 1980. *Lords of the land: Sugar, wine, and the Jesuit estates of coastal Peru*. Albany: State University of New York Press.

Cushner, Nicholas. 1982. *Farm and factory: The Jesuit and the development of agrarian capitalism in colonial Quito, 1600–1767*. Albany: State University of New York Press.

de Janvry, Alain, Elisabeth Sadoulet & Wendy Wolford. 1998. The changing role of the state in Latin American land reforms. *CUDARE Working Papers* 852. 1–35.

Granda, Germán de. 1968. La tipología criolla de dos hablas del área lingüística hispánica. *Thesaurus* 23. 193–205.

Gutiérrez-Rexach, Javier & Sandro Sessarego. (2014). Morphosyntactic variation in three Afro-Andean dialects: The evolution of gender agreement in DP. *Lingua* 151. 142–161. https://doi.org/10.1016/j.lingua.2014.04.011

Guy, Gregory. 1981. *Linguistic variation in Brazilian Portuguese: Aspects of the phonology, syntax, and language history*. Philadelphia, PA: University of Pennsylvania dissertation.

Guy, Gregory. 2004. Muitas linguas: The linguistic impact of Africans in colonial Brazil. In José C. Curto & Paul Lovejoy (eds.), *Enslaving connections: Changing cultures of Africa and Brazil during the era of slavery*, 125–137. New York: Humanity Books.

Guy, Gregory. 2017. The African diaspora in Latin America. In Cecelia Cutler, Zvjezdana Vrzić & Philipp Angermeyer (eds.), *Language contact in Africa and the African diaspora in the Americas*, 49–78. Amsterdam: John Benjamins. https://doi.org/10.1075/cll.53.03guy

Hassaurek, Friedrich. 1867. *1831–1885: Four years among Spanish-Americans*. New York: Hurd and Houghton.

Hernández, Tanya. 2013. *Racial subordination in Latin America: The role of the state, customary law, and the new civil rights response*. Cambridge: Cambridge University Press.

Kenneth, Andrien. 1995. *The Kingdom of Quito, 1690–1830*. Cambridge: Cambridge University Press.

Klee, Carol & Andrew Lynch. 2009. *El español en contacto con otras lenguas*. Washington, DC: Georgetown University Press.

Lipski, John. 1987. The Chota Valley: Afro-Hispanic language in highland Ecuador. *Latin American Research Review* 22(1). 155–170.

Lipski, John. 2005. A history of Afro-Hispanic language. *Five centuries, five continents*. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511627811

Lipski, John. 2006. El dialecto afroyungueño de Bolivia: En busca de las raíces del habla afrohispánica. *Revista Internacional de Lingüística Iberoamericana* 3. 137–166.
Lipski, John. 2008. *Afro-Bolivian Spanish*. Madrid & Frankfurt: Iberoamericana. https://doi.org/10.31819/9783865279026
Lipski, John. 2009. Afro-Choteño speech: Towards the (re)creation of "Black Spanish". *Negritud* 2. 99–120.
Lockhart, James. 1994. *Spanish Peru, 1532–1560: A social history*. Madison: University of Wisconsin Press.
Luna de Sola, David. 1978. Algunos aspectos ideológicos de la independencia latinoamericana. *Anuario de Estudios Centroamericanos* 4(1). 79–92.
Macera, Pablo. 1966. *Instrucciones para el Manejo de las haciendas jesuitas del Perú, ss. XVII–XVIII*. Lima: Universidad Nacional Mayor de San Marcos.
McWhorter, John. 2000. *The missing Spanish Creoles. Recovering the birth of plantation contact languages*. Berkley: University of California Press.
Megenney, William. 1993. Elementos criollo-portugueses en el español dominicano. *Montalbán* 15. 3–56.
Mellafe, Rolando. 1984. *La introducción de la esclavitud negra en Chile: tráfico y rutas*. Santiago: Editorial Universitaria.
Otheguy, Ricardo. 1973. The Spanish Caribbean: A creole perspective. In Charles-James Bailey & Roger Shuy (eds.), *New Ways of Analyzing Variation in English*, 323–339. Washington, DC: Georgetown University Press.
Peñaherrera de Costales, Piedad & Alfredo Costales Samaniego. 1959. *Coangue o historia cultural y social de los negros del Chota y Salinas*. Quito: Llacta.
Perez, Danae. 2015. Traces of Portuguese in Afro-Yungueño Spanish? *Journal of Pidgin and Creole Languages* 30(2). 307–343. https://doi.org/10.1075/jpcl.30.2.04per
Perez, Danae. 2016. Response to Sessarego's response. *Journal of Pidgin and Creole Languages* 31(1). 213–216. https://doi.org/10.1075/jpcl.31.1.09per
Perez, Danae, Sandro Sessarego, & Eeva Sippola. 2017. Afro-Hispanic varieties in comparison: New light from phylogeny. In Peter Bakker, Finn Borchsenius, Carsten Levisen & Eeva Sippola (eds.), *Phylogenetic approaches to creole languages*, 269–292. Amsterdam: John Benjamins. https://doi.org/10.1075/z.211.12per
Portugal-Ortiz, Max. 1977. *La esclavitud negra en las épocas colonial y nacional de Bolivia*. La Paz: Instituto Boliviano de Cultura.
Rao, Rajiv & Sandro Sessarego. 2016. On the intonation of Afro-Bolivian Spanish declaratives: Implications for a theory of Afro-Hispanic Creole Genesis. *Lingua* 174. 45–64. https://doi.org/10.1016/j.lingua.2015.12.006
Rao, Rajiv & Sandro Sessarego 2018. Intonation in Chota Valley Spanish: Contact-induced phenomena at the discourse/phonology interface. *Studies in Hispanic and Lusophone Linguistics* 11(1). 163–192. https://doi.org/10.1515/shll-2018-0006
Restall, Matthew. 2000. Black conquistadors: Armed Africans in early Spanish America. *The Americas* 57(2). 171–205. https://doi.org/10.1353/tam.2000.0015
Romero, Rey & Sandro Sessarego. 2018. Hard come, easy go: Linguistic interfaces in Istanbulite Judeo-Spanish and Afro-Ecuadorian Spanish. In Jeremy King & Sandro Sessarego (eds.), *Language variation and contact-induced change: The Spanish language across space and time*, 63–81. Amsterdam: John Benjamins. https://doi.org/10.1075/cilt.340.04rom
Rosenblat, Ángel. 1954. *La población indígena y el mestizaje en America*. Buenos Aires: Editorial NOVA.

Rostworowski Tovar de Díez Canseco, María. 1999. *Historia de los Incas*. Lima: Editorial Bruño.
Schwegler, Armin. 1991a. La doble negación dominicana y la génesis del español caribeño. *Lingüística*, 3: 31–88.
Schwegler, Armin. 1991b. El español del Chocó. *América Negra*, 2: 85–119.
Schwegler, Armin. 1993. Rasgos (afro-)portugueses en el criollo del Palenque de San Basilio (Colombia). In Carmen Díaz Alayón (ed.), *Homenaje a Jose Perez Vidal*, 667–696. La Laguna, Tenerife: Litografía A. Romero S. A.
Schwegler, Armin. 1999. Monogenesis revisited: The Spanish perspective. In John Rickford & Suzanne Romaine (eds.), *Creole genesis, attitudes and discourse*, 235–262. Amsterdam: John Benjamins. https://doi.org/10.1075/cll.20.16sch
Schwegler, Armin. 2014. Portuguese remnants in the Afro-Hispanic diaspora. In Patrícia Amaral & Ana Maria Carvalho (eds.), *Portuguese-Spanish interfaces: Diachrony, synchrony, and contact*, 403–441. Amsterdam: John Benjamins. https://doi.org/10.1075/ihll.1.21sch
Sessarego, Sandro. 2011. *Introducción al idioma afroboliviano: una conversación con el awicho Manuel Barra (with CD)*. Cochabamba: Plural Editores.
Sessarego, Sandro. 2013a. Afro-Hispanic contact varieties as conventionalized advanced second languages. *IBERIA* 5(1). 96–122.
Sessarego, Sandro. 2013b. *Chota Valley Spanish*. Madrid: Iberoamericana. https://doi.org/10.31819/9783954871995
Sessarego, Sandro. 2013c. On the non-creole bases for Afro-Bolivian Spanish. *Journal of Pidgin and Creole Languages* 28(2). 363–407. https://doi.org/10.1075/jpcl.28.2.04ses
Sessarego, Sandro. 2013d. Chota Valley Spanish: A second look at creole monogenesis. *Revista Internacional de Lingüística Iberoamericana* 22(2). 129–148.
Sessarego, Sandro. 2014. *The Afro-Bolivian Spanish determiner phrase: A microparametric account*. Columbus, OH: The Ohio State Press.
Sessarego, Sandro. 2015. *Afro-Peruvian Spanish*. Amsterdam: John Benjamins. https://doi.org/10.1075/cll.51
Sessarego, Sandro. 2016. A Response to Perez. *Journal of Pidgin and Creole Languages* 31(1). 200–212. https://doi.org/10.1075/jpcl.31.1.08ses
Sessarego, Sandro. 2018. *La schiavitù nera nell'America spagnola*. Genova: Marietti Editore.
Sessarego, Sandro. 2019a. *Language contact and the making of an Afro-Hispanic vernacular: Variation and change in the Colombian Chocó*. Cambridge: Cambridge University Press. https://doi.org/10.1017/9781108661782
Sessarego, Sandro. 2019b. Universal processes in contact-induced syntactic change: Evidence from the Afro-Hispanic varieties. In Jeroen Darquennes, Joseph Salmons & Wim Vandenbussche (eds.), *Language contact. An international handbook*, 24–37. Berlin: Mouton de Gruyter.
Sessarego, Sandro & Javier Gutiérrez-Rexach. 2012. Variation, universals, and contact-induced change: Language evolution across generations and domains. In Melvin González-Rivera & Sandro Sessarego (eds.), *Current formal aspects of Spanish syntax and semantics*, 251–270. Newcastle: Cambridge Scholars Publishing.
Thomason, Sarah & Terrence Kaufman. 1988. *Language contact, creolization, and genetic linguistics*. Berkeley: University of California Press.

Vamos en Palma 'we are going to Palma'

On the persistence (and demise) of a contact feature in the Spanish of Majorca

Andrés Enrique-Arias
Universitat de les Illes Balears

This study analyzes historical and contemporary data of Spanish produced by Catalan-dominant bilinguals in Majorca to explore the origin and historical evolution of a contact feature of Majorcan Spanish: the preposition *en* to express direction of movement. The transgenerational survival of this feature of Majorcan Spanish for over three centuries points to a scenario of intra-community recycling of a structure that emerged as the result of incomplete grammatical competence in Spanish and was then perpetuated by limited contact with monolingual Spanish speakers. The changing linguistic ecology of Majorca, however, with wider access to canonical varieties of Spanish in the last half century, has facilitated the demise of this trait among younger urban generations.

Keywords: language contact, Spanish and Catalan in Majorca, spatial relations, prepositions, intra-community recycling

1. Introduction

Since the seminal work of Moll (1961) on interference phenomena in the Spanish of Catalan-dominant speakers in Majorca (Spain) there has been growing interest in the study of this contact variety. In recent years, the creation of new resources, such as sociolinguistic interviews, dialectological surveys, laboratory recordings for phonetic analysis and historical corpora, provides an empirical basis for the understanding of the mechanisms of language variation and change in this contact situation. Three new resources for the study of Spanish in contact with Catalan in Majorca include the Corpus Mallorca (https://www.corpusmallorca.es), a corpus of personal letters and court testimonies written by Catalan–Spanish bilinguals covering the 18th and 19th centuries; the COSER-Mallorca corpus (https://www.corpusrural.es), a collection of interviews with rural elderly informants; and the PRESEEA-Palma corpus, which contains sociolinguistic interviews in the urban

https://doi.org/10.1075/ahs.12.co8enr

setting of Palma. These new data sources include speech samples from a diverse pool of participants from unbalanced bilinguals in rural settings to Spanish-dominant urban speakers with just a passive knowledge of Catalan.

Thanks to these resources we are beginning to understand in all its complexity the wide range of idiolectal variation found in Majorcan Spanish, and at the same time, the more we know we are faced with new questions and challenges. A comparison of contemporary data with historical documents produced by bilinguals reveals a number of traits that present a remarkable historical continuity for at least three centuries. The persistence of these features of Majorcan Spanish poses interesting theoretical questions regarding the creation and historical continuity of contact varieties. As it is normally assumed that Catalan speakers in Majorca use Catalan among themselves with Catalan-Spanish interlanguage being a transitory register used only in interactions with Spanish monolinguals, there is no easy or immediate explanation for the transgenerational survival of traits which cannot be reasonably analyzed as arising spontaneously and consistently generation after generation. In order to try to find an explanation for the historical survival of contact features in the Spanish of Majorca, this paper deals with one non-canonical feature that has been documented as early as 1700 in historical documents produced by bilinguals and is still found in the speech of Catalan-dominant rural elderly speakers: the use of the preposition *en* (instead of *a*) to convey direction of movement (*vamos en Palma* 'we go to Palma'). The main objectives that will be addressed in the pages to follow are:

a. To determine the linguistic origin of this element (whether it derives from Catalan transfer, or it is the result of Spanish internal evolution, or both)
b. To describe its distribution and frequency of use
c. To understand how this element has been transmitted across generations

More generally this study aims to explore the linguistic, cognitive, and social processes involved in the transmission and the consolidation –or demise– of the linguistic features that are characteristic of Majorcan Spanish.

2. Data and methodology

The present study is based on the analysis of historical and contemporary data. The historical data has been drawn from the *Corpus Mallorca* (https://www.corpusmallorca.es).[1] For this particular study we have focused on 18th century texts: the

1. This effort is part of the CHARTA network (https://www.charta.es), a consortium of research groups that compile textual corpora made up of historical documents issued in different parts of the Spanish-speaking world.

Cecilia Zaforteza epistolary archive which includes a collection of over two hundred personal letters in Spanish and Catalan dated between 1739–1788 (for a full description see Enrique-Arias 2014b), as well as court testimonies from legal cases from the Royal Court of Palma. The legal documents that I have selected for this study are dated between 1720 and 1799. In order to have a control corpus of comparable monolingual Spanish from the same period I have also searched twelve legal files from the Corpus Diacrónico del Reino de Granada, dated between 1700–1771.

As for the contemporary data, in order to examine the Spanish produced by Catalan-dominant speakers interviews were collected in 15 rural localities in Majorca within the Rural Spanish Corpus (COSER in its Spanish acronym, at https://www.corpusrural.es). All participants were elderly people (between 64 and 95, with average age 86.5), mostly farmers with lifetime residence in their places of origin or at the most sporadic residence outside of the area. Additional criteria for the selection of participants included occupation in the primary sector (agriculture, raising livestock, or fishing), no access to formal education beyond elementary school and little contact with Spanish monolinguals and canonical forms of Spanish. All the participants are Catalan-dominant and acquired Spanish during the little schooling they had, or through mass media in Spanish and interactions with Spanish monolinguals. Since the mid 20th century Majorca has received heavy immigration from Spanish monolingual areas in the mainland and also has developed a massive tourism industry; as a result, work in the service sector is pervasive in the island, and in a few cases the participants had spent some time in Palma working as gardeners or taxi drivers, and in one case a speaker had been married to a Spanish-speaking man. Nevertheless, all of the participants exhibit the traits typically associated with the Spanish of rural Catalan-dominant Majorcans, such as *seseo*, devoicing of –d in final position or dark velarized /l/ in syllabic coda. The interviews, which lasted over one hour, were collected in May of 2017. The format was free conversations on topics dealing with rural life and life in old times in general.

The rural interviews were complemented with sociolinguistic interviews carried out within the PRESEEA project in the urban setting of Palma between 2007 and 2010. Informants were selected using sampling based on quotas of sex, age, and level of education; each interview lasted around 45 minutes. Following the methodology designed for the PRESEEA project, a fixed number of thematic modules were introduced during the course of all the interviews (for the PRESEEA project methodological guidelines see http://preseea.linguas.net/). The 54 participants are equally divided between men and women belonging to three age groups (20–34, 35–54, 55 and over) and three education levels (primary, secondary, college) distinguishing language dominance according to a survey of biographical data and language use. For this study we have analyzed only data produced by the 18 speakers classified as Catalan-dominant.

3. Directional uses of *en*

Catalan-dominant speakers in Majorca use the preposition *en* to express direction of movement in contexts that would be ungrammatical in monolingual varieties of Spanish. This use was first observed by Moll (1961: 472), who provides the following examples:

(1) a. Se fue **en** Barcelona
'He left for Barcelona.'
b. A las nueve llegamos **en** la ciudad
'At nine we arrived in the city.'

The structure in Examples (1a)–(b) is ungrammatical in standard Spanish, which uses *en* for location and *a* for direction as shown in the Examples (2a)–(b).

(2) a. Estoy *__a__ / __en__ Palma
'I am in Palma.'
b. Voy *__en__ / __a__ Palma
'I go to Palma.'

The historical corpus used in this study shows that the use of *en* to express direction of movement was rather frequent among Catalan-Spanish bilinguals in the 1700s (cf. Enrique-Arias 2010: 111–14, 2019). In my thorough search of 23 court cases in legal texts in the historical corpus, I found 54 occurrences of directional *en* dated between 1721 and 1799. The wide number of writers – 11 different hands with texts produced in 9 different localities – attests to the pervasiveness of the phenomenon in the 18th century. Likewise, directional *en* in the court documents appears in clauses with 18 different verbs in contexts conveying direction of movement (the number between parentheses indicates number of cases when the same verb appears more than once): *ir* 'go' (13), *pasar* 'pass' (9), *venir* 'come' (5), *volver* 'return' (5), *arrimarse* 'come closer' (3), *llegar* 'arrive' (3), *acudir* 'go' (2), *conducir* 'drive' (2), *remitir* 'send' (2), *retirarse* 'retreat' (2), *acompañar* 'escort', *asistir* 'attend', *echar* 'throw', *enviar* 'send', *llevar* 'carry', *marcharse* 'leave', *mudarse* 'move', *traer* 'bring'. In the private letters, which constitute a somewhat less formal genre, we also find 10 cases of directional *en* with verbs *pasar* 'pass' (4) *ir* 'go' (3), *llegar*, *arrivar* 'arrive' and *venir* 'come'. Here are some examples from the historical corpus ((3a)–(d) from legal cases and 3e–f from personal letters) with indication of the year and place of production:

(3) a. al llegar en la plaza de la Seo [1724, Palma]
'upon arriving to the Cathedral square'
b. el declarante se fue en casa [1760, Palma]
'the person testifying went home'

c. se marcharon en la villa de Pollensa [1769, Sa Pobla]
'they went to the town of Pollensa'
d. vino en mi casa Gabriel Mateu [1771, Montuiri]
'Gabriel Mateu came to my house'
e. Llegamos en esse puerto [1759, Malta]
'We arrived at that port.'
f. en el viernes pasado pasamos en Palma [1773, Palma]
'Last Friday we went to Palma.'

Directional uses of *en* are not completely unheard of in the history of Spanish (see Hanssen 1945: 297 for some medieval examples and Palacios 2019: 241–242 for 16th and 17th century ones), but they faded progressively and became extremely rare by the 18th century. In order to establish a control corpus I examined 12 legal cases from the Corpus Diacrónico del Español del Reino de Granada produced in the provinces of Granada and Almeria in southern Spain. The documents, dated between 1700–1771, are very similar in content and style to the court documents in the Majorcan corpus. After an exhaustive search I found only one example of directional *en* in a 1731 document dated in Monachil (Granada): *el que declara se bolvió en casa* 'the one testifying returned home'. We can conclude that this use that, as we have seen, is fairly frequent in the Spanish of 18th century Majorca, was rather infrequent in monolingual Spanish of the time.

As for the current existence of these constructions, contemporary data from our oral corpora indicates that directional *en* occur in present-day Majorcan Spanish but it is confined almost exclusively to older rural Catalan-dominant bilinguals. In the 15 COSER interviews 12 participants (that is, a remarkable 80%) exhibited at least one example of directional *en*. In total there are 38 examples in the corpus involving 7 different verbs: *ir* 'go' (26), *venir* 'come' (7), *acompañar* 'escort', *cambiarse* 'move', *llevar* 'take', *marcharse* 'leave', *traer* 'bring'. Here are a few examples with indication of the speaker's age, gender and place of origin:

(4) a. nos cambiamos en Alcudia [female, 83, Alcudia]
'we moved to Alcudia'
b. esto era por no ir en África [male, 83, Campos]
'this was in order to not go to Africa'
c. mi padre se fue en La Habana [male, 83, Campos]
'my father went to Havana'
d. los jóvenes no vais en la peluquería [female, no data, Consell]
'you young people do not go the the hair salon'
e. nos fuimos en Barcelona [female, 82, Llubí]
'we went to Barcelona'
f. los hemos traído en Pollensa [female, 70, Pollensa]
'we have brought them to Pollensa'

At the same time this use has largely disappeared among urban speakers. Even though the 54 sociolinguistic interviews in the PRESEEA-Palma corpus include 18 Catalan-dominant speakers, only two examples were found, both of them from a 71 year old woman with primary education:

(5) a. ellos van en la casa de otros amigos [3M1-Cat 148]
'they go to other friends' houses'

b. mi hermana arregló los papeles para mi padre y lo llevó en una residencia [3M1-Cat 440]
'my sister arranged the papers for my father and took him to a retirement home'

In short, the age of the speakers and the density of the Catalan-dominant population in the community where they live seem to be the main predictors of the persistence of the directional uses of *en*. As was to be expected, the more widespread use of the Catalan language in daily interactions and, crucially, the comparatively lower exposure to canonical Spanish in rural areas favors the continuation of this structure. On the other hand, in the city of Palma exposure to monolingual Spanish is greater, which in turn explains why directional *en* is very rare, even among Catalan-dominant bilinguals. The reason is that in Palma Catalan speakers have frequent face-to-face interactions with both Spanish monolingual speakers and Spanish-dominant bilinguals and thus are more exposed to standard Spanish structures.[2] Although the rural informants in the COSER corpus are somewhat older than the ones in Palma, the two populations are still comparable as the difference regarding date of birth, which is about 8–10 years, is not enough to establish a generational gap between the two groups.[3]

Additional data from grammaticality judgement tasks demonstrates that Majorcans as a whole do not find the directional uses of *en* to be more acceptable compared to Spanish monolingual speakers from other areas (cf. Enrique-Arias 2019: 173–74). In an evaluation task distributed freely by email and in social media

2. For instance, according to the data in Melià and Vanrell (2018: 61) Catalan is the most common language in everyday use for 27.9% of people surveyed in Palma, but this percentage goes up to 45.5% in the rest of the island. In small rural settings, like the ones surveyed in the COSER corpus used in this study, and especially among rural elderly speakers with little formal education, the use of Catalan is even higher, almost exclusive, while there is little exposure to monolingual Spanish.

3. In the Palma corpus the six Catalan-dominant speakers in the older group were aged between 57 and 77 (average age 68.5) at the time when they were interviewed as compared to 86.5 in the rural group. Since the PRESEEA interviews were collected between 2007 and 2010 and the rural ones in 2017, the speakers in the urban group were born just 8–10 years after the rural ones, which means that they belong to the same generational group.

(for a description of the methodology employed see Enrique-Arias 2014a: 280) participants were asked to give a score between 7 ('completely natural') and 1 ('completely unacceptable') to the following sentences with directional uses of *en*:[4]

(6) a. Hace tiempo que la familia se trasladó *en* este barrio
'Some time ago the family moved to this neighborhood'
b. Salgan *en* la calle para ver la cortina desde fuera
'Go out in the street to see the curtain from outside'
c. Me pongo nervioso cuando tengo que ir *en* un sitio nuevo
'I get nervous when I have to go to a new place'
d. Cuando llegamos *en* Barcelona ya no quedaban hoteles
'When we arrived to Barcelona there were no hotel rooms left'

The questionnaire was distributed to Catalan-speaking bilinguals in Majorca and Spanish-speaking monolinguals from various geographical areas (Castile, Madrid, Andalusia). In all locations participants' responses rendered average values well below 4 (labelled 'acceptable but sounds a bit strange'): Majorca (Catalan-dominant) 2.31, Majorca (Spanish-dominant) 2.24, Castile 1.79, Madrid 1.97, Andalusia 2.20. The results seem to indicate that the directional uses of *en* are idiosyncratic (i.e., non community-wide) structures that appear sporadically in only a few bilingual speakers. Although acceptability among Majorcans is slightly higher than what we observe in non-contact areas, the difference is rather small.

The different data sources analyzed in this section exhibit a sharp contrast between the data from the rural corpus, where directional *en* is nearly universal, and the sociolinguistic interviews in Palma, where it is virtually absent, while the questionnaires show that young, educated Majorcans exhibit attitudes towards this structure that are comparable to those of monolingual Spanish speakers elsewhere. All of the above seems to indicate that after three centuries of continuity this trait of Majorcan Spanish is not being transmitted to the younger generations and is becoming a case of residual variation after a process of convergence with the monolingual standard (Enrique-Arias 2019: 175).

4. I am well aware of the methodological problems associated with grammaticality judgements and more so in a survey that has been distributed online with no direct contact with the participants (Silva-Corvalán and Enrique-Arias 2017: 86–88). Also it should be noted that the sample has a clear slant toward women (64%) and individuals with university education (83%). However, the distribution of genders and levels of education among the participants from each geographical area is reasonably homogeneous, which serves our main purpose of comparing attitudes towards directional *en* in bilingual and monolingual areas.

4. Looking for the source of Majorcan Spanish directional *en*

4.1 Majorcan Catalan

As we have seen, the use of directional *en* in the Spanish of Majorca clearly correlates with Catalan language dominance, that is, it is more frequent and intense among unbalanced bilinguals living in rural communities where there is little contact with monolingual Spanish speakers. Based on this, we would expect that this structure is the result of on-line interference from Catalan and thus it parallels prepositional uses in that language; as we are about to see, however, this is not the case, as the Catalan currently spoken in Majorca uses *a*, and not *en*, to express *both* location and direction as in the examples under (7):

(7) a som a Palma / *som en Palma
'I am in Palma'
b. vaig a Palma / *vaig en Palma
'I go to Palma'

Some reference grammars (such as Wheeler *et al.* 1999: 246 or Sancho Cremades 2002: 1733) point out that in Catalan the preposition *en* can be used optionally to express direction in front of articles and demonstratives that begin with a vowel. But, at least for the Catalan of Majorca, it could be argued that this is not the preposition *en* but a phonetically conditioned variant of the preposition *a*. In the dialect area traditionally known as Eastern Catalan (which includes Majorca) non-tonic [a] and [e] neutralize as [ə]. Therefore when the preposition *a* (pronounced [ə]) precedes an article or demonstrative that begins with a vowel speakers insert an epenthetic [n] segment to negotiate the phonotactic difficulties created by having to articulate two adjacent identical vowels in hiatus (see Example (8a) below); this possibility, however, is treated by grammarians as if this was the preposition *en* while the proposals to use alternative spellings to render this allomorphic variant of *a* as *an* have met with little success (Escandell-Vidal 2007: 30). Instead, as this phonetically conditioned variant of the preposition *a* is homophonous with the preposition *en* (they are both pronounced [ən], and in fact they are usually spelled *en*), the idea that there is a generalized confusion of the prepositions *a* and *en* in Majorcan Catalan has taken hold. Nevertheless, if we look at those cases in which there is no epenthetic -[n], such as words beginning with a consonant, it is clear that there is no such confusion (see 8b). Also, it should be noted that place names are excluded from the epenthesis rule (cf. 8c) which, again, reinforces our analysis of the sequence [ən] in directional contexts as a variant of the preposition *a* and

not the preposition *en*. In short, the use of *en* in this structure is a mere spelling convention rather than an actual grammatical rule.[5]

(8) a. *Vaig **an** es poble.*
'I go to the town.'
b. *Vaig ***en** / **a** sa plaça.*
'I go to the square.'
c. *Vaig***en** / **a** Alcudia*
'I go to Alcudia'

Furthermore, the examples that we have collected of directional *en* in the Spanish of Majorca do not occur predominantly in those contexts in which Majorcan Catalan uses *an*, that is, before articles and demonstratives that begin with a vowel. As can be seen in Table 1, in both the historical and modern corpora three fourths of the examples correspond to contexts in which Majorcan Catalan uses only *a* for direction (i.e., before words beginning with a consonant and toponyms), and only one fourth occur in contexts that trigger the phonetically conditioned *an* variant.[6]

Table 1. Directional *en* in the Majorcan corpora sorted by phonetic context after the preposition

	Articles and demonstratives	Consonant	Toponyms
Historical corpus ($N = 64$)	15 (23.4%)	8 (12.5%)	41 (64.1%)
Modern corpus ($N = 40$)	10 (25%)	11 (27.5%)	19 (47.5%)
TOTAL ($N = 104$)	25 (24%)	19 (18.3%)	60 (57.7%)

From a comparison of the related Spanish and Catalan structures it follows that the directional uses of *en* in the Spanish of Majorca are not the result of simple direct transfer from present-day Catalan structures. As Majorcan Catalan uses *a* for both static location and direction of movement, if bilinguals were merely transferring the Catalan structure into their Spanish one we would expect an overuse of the preposition *a*, not *en*, to express spatial relations. Although such examples exist both in the historical corpus and the COSER interviews, they are rather infrequent compared to the much more frequent cases of directional *en*.

5. This analysis is further supported by the fact that the same epenthetic [n] emerges in other uses of the preposition *a*, such as the prepositional indirect object plus the "personal article" *en* that precedes personal names: *vaig donar el llivre an en Pau* (*an en* is pronounced [ənən]) (Escandell-Vidal 2007: 30).

6. The historical corpus includes the 54 cases in the court documents and the 10 occurrences in the private letters (64 total), and the modern corpus comprises the 38 examples in the COSER interviews and the two cases in the PRESEEA corpus (40 total).

4.2 Simplification and hypergeneralization

In the only study that has observed the directional uses of *en* in Majorcan Spanish (aside from my own investigations) Moll (1961: 472) simply assumes that because the uses of *en* and *a* are different in Catalan and Spanish, bilingual speakers have difficulties mastering their distribution, which results in an overuse of *en* in their Spanish utterances. Although Moll does not elaborate on the issue we may surmise that his proposed mechanism operates as follows: Catalan-dominant speakers notice that in Spanish *en* is used in locative sentences such as *estoy en Palma* 'I am in Palma' in contexts in which the equivalent Catalan structure would use *a*; hence they infer *en* is to be used when speaking Spanish in at least some of the contexts where Catalan would use *a*, but they overgeneralize this rule to include direction of movement. Hypergeneralization processes of this type are a common phenomenon in the speech of bilinguals when operating in the less dominant language, particularly in situations of reduced input and infrequent use.

In explaining this phenomenon of Majorcan Spanish we have to first understand that the standard use of prepositions is difficult to achieve for both L1 and L2 learners. As Romaine (1995) points out, prepositions are a difficult grammatical category to acquire and understand for native speakers of any language; this is illustrated by the work of Kvaal et al. (1988), who in a study of the acquisition of ten Spanish morphemes by monolingual Spanish children ages 2;0 to 4;8 found that the preposition *en* was one of the latest to be acquired only second to the irregular preterit indicative. The challenge of acquiring the standard use of prepositions may be even greater for second-language speakers including cases involving languages that, like Spanish and Catalan, are closely related. Solà (1980) gives a detailed description of grammatical items discussed in prescriptivist guides for "good use" of Spanish intended for Catalan speakers. One of the errors that are commonly emphasized and corrected are prepositional uses associated with the speech of Catalan-dominant bilinguals such as non standard interchanges of prepositions *a* and *en*, *por* and *para*, *con* and *en*, among others (cf. Solà 1980: 577–578).

The directional uses of *en* in Majorcan Spanish would therefore be motivated by the inherent difficulty in achieving a more standard monolingual use of prepositions and the fact that there is no direct correspondence to the same preposition in the two languages involved (Catalan and Spanish). This would result in potential incomplete learning in the part of Catalan-dominant bilinguals. This scenario is supported by the fact that, in various contact situations across the Spanish-speaking world, speakers produce directional uses of *en* regardless of the status of the corresponding prepositions in the contact language. For instance, Granda (1988) and others (cf. Lipski 1990: 19; Casado-Fresnillo 1995: 290) report a number of idiosyncratic prepositional uses in the Spanish of Equatorial Guinea; the most common one is the

interchanging of *en* for *a* with verbs of movement, which Lipski (1990: 59–61) illustrates with abundant examples such as *voy en Bata* 'I go in Bata', *iremos en Malabo* 'we will go in Malabo', among many others. Both Lipski and Casado-Fresnillo consider that this is caused by the influence of the vernacular Bantu languages such as Fang which employs the same structure for direction of movement and static location: *make ô Mâng* 'I go to Bata'; *mene ô Mâng* 'I am in Bata.' Likewise, in the Spanish of Paraguay the use of *en* with verbs of directional modality is rather common (Choi 2001; Granda 1991: 47–50, Palacios 2019). Granda explains this phenomenon as the result of the retention of a structure which existed in 16th-century Spanish encouraged by contact with Guarani: in this language the postpositional element *-pe* / *me* may signify either 'to' / 'towards' or 'in' depending on whether the verb is one of motion or state. Directional uses of *en* have also been reported among Spanish-Quechua bilingual speakers in Ecuador: *voy ir en Guayaquil* 'I am going to go to Guayaquil' (Haboud 1998: 255); and Peru: *llegué en una panadería* 'I arrived to a bakery' (Escobar 1990: 75); *en su casa se había ido* 'he had gone to his home' (Minaya Portella 1978: 470). In contrast to Fang and Guarani, which uses the same marker to convey direction of movement and static location, Quechua uses distinct markers for each one of these functions: whereas the suffix *-man* is used to express destination (*yachana-man-mi riguni* 'I am going to class') the suffix *-pi* is used to express location (*yachana-pi-mi kan* 'he is in class'). Finally, MaryEllen García has observed the extension of the preposition *en* with directional verbs in the Spanish of two Texas (United States) cities: El Paso (*me vendría aquí en El Paso* 'I would come here to El Paso', García 1982: 92) and San Antonio (*he ido en México* 'I've gone to Mexico', García 1995: 201). García explains this phenomenon as the linguistic simplification and restructuring that is common in language contact situations. In her view the use of *en* with directional verbs could be the result of language contact but it is not necessarily caused by direct transfer from English. For instance, in the two examples just provided the corresponding preposition in English is not *in* but *to*.[7]

From the previous discussion of language contact varieties of Spanish it follows that bilinguals tend to favor directional uses of *en* not only in those cases in which the contact language makes no clear distinction between direction of movement and static locality (Guarani, Catalan, Bantu) but even when, like in English or Quechua, it has a system similar to the monolingual Spanish one. It seems reasonable to explain this pattern with psycholinguistic factors. In language contact situations bilingual speakers develop certain strategies in order to alleviate the cognitive burden of having to remember and use more than one linguistic system. Two of these strategies are the simplification of grammatical categories and semantic

7. Directional uses of *en* have also been reported in the Spanish of Chile (Chiloé and Valdivia) (cf. Oroz 1966: 399).

distinctions, and the hypergeneralization of linguistic forms (Silva-Corvalán 1994). These processes typically result in the extension or higher frequency of use of a form X in a context Y at the expense of a form Z that decreases in its contexts of use. Silva Corvalán (1994) and others such as Romaine (1995) consider that languages are strikingly impermeable to foreign influence at the syntactic level and thus, prepositions are rarely borrowed, although it is accepted that the language contact situation may motivate the restructuring of their distribution. In Majorcan Spanish hypergeneralization would have worked in the following way: Majorcan Catalan uses the same marker for two functions (*a* for both direction of movement and static location) while standard monolingual Spanish uses different markers for each of those functions (*a* and *en*, respectively). Catalan-dominant bilinguals realize that Spanish speakers use a different marker in some locative contexts, but, due to limited input and infrequent use, are unable to discriminate between them correctly and tend to produce this "foreign" item (*en*, in this case) with both functions whenever speaking their less dominant language. On the other hand, more educated speakers from urban settings are free from the interference, due to their typically higher level of instruction and frequent exposure to Spanish while uneducated rural speakers are less firm in their choice of prepositions.

Another factor to take into account is that prepositional use is an area of the grammar that is particularly vulnerable to change, and more specifically the uses of *a* and *en* have fuzzy limits in the monolingual standard varieties of all Romance languages (cf. Choi 2001: 188–193). For instance, in standard Spanish *en* may express direction with quite a few verbs, such as *entrar* 'enter', *penetrar* 'penetrate', *caer* 'fall', *subir* 'go up', *echar* 'toss' (*entraron en la casa* 'they entered in the house'); likewise *a* may express location (*esperaban a la puerta de su casa* 'they waited at the door of his house'); that is, the contact situation would have contributed to the extension of a structure that already existed in standard Spanish to more linguistic contexts.

But we are still faced with the question of why *en* is favored even in those cases, like Catalan-Spanish contact, in which a tendency to overuse *a* would be expected. A possible reason is the relatively higher frequency of the preposition *en* in Spanish as compared to Catalan in the expression of spatial relations: as Catalan uses *a* for both location and direction and Spanish divides those uses between *a* and *en*, Catalan speakers will perceive the latter preposition as more salient (i.e., alien to their native usage). This difference could have the following effect: Catalan-dominant speakers that feel insecure about what Spanish preposition to select when expressing spatial relations may be inclined to pick the one that is identified as more properly Spanish and thus more likely to be correct and, as a result, overuse *en*.

4.3 Historical data

One of the more common situations in which researchers tend to establish that contact-induced change has taken place is when in a given variety we find a feature that (a) is different from the structural equivalent in the non-contact variety and (b) exhibits a structural parallel with the presumed source. Nevertheless, in our search for the source of directional *en* in Majorcan Spanish this method does not yield positive results: as we have seen in our contrasting analysis of Spanish and Catalan, this trait cannot be explained as direct transfer from Majorcan Catalan which uses *a* and not *en* to express direction of movement. As I have explained elsewhere (cf. Enrique-Arias 2010, 2014a, 2019) the problem here is that in locating the source for a contact-induced change it is not enough to look at the current situation; we also need to look at the history of the languages involved and examine whether the proposed interference features were not present in past stages of the language or even in the pre-contact variety. In fact, a diachronic perspective complements the synchronic factors outlined above, unveiling a crucial factor: unlike present-day Catalan, in the language of the 18th century *en* was frequently used to express direction. During the medieval period both Spanish and Catalan exhibited variation between *a* and *en* in the expression of spatial relations; by the 16th century, however, the two languages had diverged substantially: while in Spanish *a* had become the usual preposition for directional uses, in Catalan *en* was still fairly frequent in such structures and, in fact, it was the predominant preposition in certain directional contexts, such as with place names referring to countries and regions (Cabanes 1989). By the 18th century directional *en* was a rarity in Spanish but quite common in Catalan. An analysis of the Catalan letters in the epistolary corpus shows that *en* and *a* were used interchangeably to express 'direction' (see 9a for directional *en* and 9c for directional *a*) and 'location' (see 9b and 9d for *en* and *a*, respectively). I include here only examples with toponyms to show that these are genuine occurrences of *en* and not phonetically conditioned variants of *a*.

(9) a. ***en*** 'direction to':
aura de ***passar en*** *Vich* [1743, Barcelona]
'he will have to get to Vic'
G. Thomàs, que ***arribà*** *lo dia de pasqua* ***en*** *Madrit* [1756, Barcelona]
'G. Thomàs, who arrived the day of Easter to Madrid'

b. ***en*** 'location':
jo ***me trobave en*** *Llucmajor* [1748, Vilafranca de Bonany]
'I found myself in Llucmajor'
Su Ilustrisima ***estera en*** *Muro* [1779, Muro]
'his honor will be in Muro'

c. ***a*** 'direction to'
y **torna a** Vinagrella en salut [1780, Llubí]
'and returned to Vinagrella in good health'
y **áribaran** ap perfeta salut **á** Madrid [1770, Barcelona]
'and they will arrive in perfect health to Madrid'

d. *a* 'location'
bolgue que lo **acabas de pagar á** Madrid [1770, Barcelona]
'he wanted me to finish paying it in Madrid'
y no mi troba a causa de **eser** jo **á** la Granja [1770, Madrid]
'and he didn't find me because I was in La Granja'

This means that first, directional uses of *en* were quite normal in Catalan at the time when it came into contact with Spanish in Majorca; and second, *en* and *a* did not have a separate distribution to express location or direction. Although in the standard peninsular Spanish of the time the current distribution of *a* for direction and *en* for location was well established, the possibility of using directional *en* in Majorcan Catalan would have encouraged a parallel Spanish construction with *en* in the speech of Catalan-dominant speakers, particularly at a time when, as I will explain in the next section, exposure to canonical forms of Spanish must have been rather limited. In sum, this trait of 18th century Catalan must have been transferred into Spanish at that time and was probably preserved in rural communities due to their rather limited contact with native speakers of Spanish. The larger implication for the field of contact linguistics is that an interference phenomenon characterizing a present-day variety may not find a structural parallel in any of the varieties currently in contact, but rather be due to options active in those varieties during earlier stages in their evolution. Therefore, historical research may be crucial in reconstructing current features of contact varieties.

5. On the historical continuity of directional *en*

From the previous data and discussion we may conclude that directional *en* in Majorcan Spanish could be explained by a combination of two factors: (1) the tendency towards the simplification of rules and hypergeneralization of variants that takes place in language contact situations and (2) the fact that in the 18th century Majorcan Catalan did have a structure with *en* to express direction which could be transferred to the Spanish of Catalan-dominant bilinguals.[8] As we have

8. Martínez i Taberner (2000) features an interesting example that illustrates both the existence of directional *en* in Majorcan Catalan and the tendency to hypergeneralize it when producing equivalent Spanish structures. Following the 1768 Royal Decree of Aranjuez, the minutes from

seen this prepositional use cannot be attributable to direct online transfer from present-day Catalan, as this language does not use *en* to express direction of movement; rather it seems that at some point, when Catalan had directional uses of *en* in the 18th century, the structure was transferred to Spanish. What is noteworthy about this non-canonical feature is how widespread it is in the Spanish of rural Catalan-dominant bilinguals, as it co-occurred among 80% of the participants who provided data for the present study. This scenario suggests limited input from monolingual Spanish speakers for an extended period, which implies circumstances in which Catalan-dominant bilinguals communicated with one another in Spanish or at least received considerable input from Catalan-Spanish interlanguage.[9]

An investigation of the sociolinguistic situation in Majorca during the 18th and 19th centuries reveals a number of situations in which intracommunity use of Spanish among Catalan-dominant speakers must have taken place. One factor to take into account is the slow and imperfect introduction of education in Spanish language in Majorca. Following the Royal Decree of Aranjuez in 1768, Spanish became the official language of instruction in public education. But as Kailuweit (1993) explains, the efforts to introduce Spanish in the schools were largely unsuccessful. He reviews abundant documentation that provides evidence that well into the 1800s teachers were ill-prepared to teach in Spanish as they were Catalan-dominant speakers with little command of the official language. This means that Majorcans were learning Spanish mainly or solely from Catalan-speaking Majorcans. Moreover, Majorcan clergy with limited competence in Spanish were in charge of religious instruction across the island. Another important factor is that, due to the low status of Catalan vis à vis Spanish, Majorcans would shift to Spanish in the presence of Spanish monolinguals; this means that even in group interactions where there was a majority of Majorcans, they would switch to Spanish for the benefit of even just one Spanish speaker. The result is that in such interactions

---

municipal meetings had to be done in Spanish. According to Martinez i Taberner (2000), municipal authorities continued holding their meetings in Catalan, took notes in Catalan, and later translated them to Spanish to fulfill the legal requirement. In one 1779 example from the town of Calvià transcribed by the author we have the following text and its translation:

> *no pugue ser impedit si vol* ***anar en Ciutat*** *o* ***a altre part*** *fora vila.*
> *no sera inpedido si quiere* ***irse en Palma*** *o* ***en otra parte*** *fuera del presente districto*
> 'he can't be prevented from going to Palma or to another place outside the town'

The Catalan version exhibits alternation of *en* and *a* (*en Ciutat … a altre part*) and there is no phonetically motivated *en* before *altre*. In the Spanish version, possibly more carefully drafted as it is intended to be the official one, we have *en* generalized to all contexts (*en Palma … en otra parte*).

9. See Lipski (2016) for a similar explanation for the historical continuity of some traits of Andean Spanish.

Majorcans would end up using Spanish with each other. Additionally, we should keep in mind that, since the early 1700s, many Majorcans abandoned Catalan and shifted to Spanish for writing purposes, both for official and private uses, which again provides another venue for the intra-community use of Majorcan Spanish traits. Finally, as Moll (1961: 473) reports, there has been the practice on the part of Majorcan parents of speaking Spanish to their children to promote their social and economic progress, which again results in the transmission of Majorcan Spanish traits. All of these sociolinguistic factors combined would explain why there has been a situation of community recycling of Catalan-Spanish interlanguage that could facilitate the transgenerational transmission of directional *en* in Majorca.

After three centuries of continuity this trait of Majorcan Spanish seems to be receding in more recent times: it is not being passed on to younger generations and it is becoming confined to elderly rural speakers. This development is due to the changing linguistic ecology of the main two languages used in Majorca in the last half century or so. Until two or three generations ago the rural communities of the kind featured in this study had very limited access to face-to-face interactions with Spanish monolinguals. Schooling in rural areas was incomplete, as children abandoned their studies to work in the fields. The incomplete formal education produced the effect of limiting the acquisition of canonical varieties of Spanish. In recent generations there has been a major demographic change: the continuous tourism boom that has been taking place in Majorca since the 1960s has attracted a massive number of Spanish-speaking immigrants from the mainland. This change together with the establishment of universal schooling and the spread of Spanish mass media has put younger generations into closer contact with canonical varieties of Spanish. The contemporary environment is completely different as regards exposure to conventional Spanish, as well as opportunities to speak Spanish as compared to previous generations. One immediate consequence is the discontinuation among younger urban speakers of several traits that have characterized Majorcan Spanish for the last few centuries, such as *seseo* or the directional uses of *en*.

6. Summary and conclusions

In this study I have used a combination of historical and contemporary data of Spanish produced by Catalan-dominant bilinguals in Majorca to explore the origin and historical evolution of the directional uses of *en* in Majorcan Spanish. The presence of similar phenomena in other Spanish contact varieties suggests that this trait of Majorcan Spanish is related to the inherent difficulty of achieving standard use of prepositions for bilingual speakers, and the fact that the limits in the distribution of *en* and *a* are not clearly defined in Spanish and more so in Catalan. The

historical data reveals that the current directional uses of *en* in Majorcan Spanish are the continuation of a structure that was already firmly established in the Spanish used in Majorca in the 18th century. In turn, this feature of Majorcan Spanish would have been encouraged by the existence of a parallel construction in 18th century Catalan. The systematic nature of these non-canonical prepositional uses and their transgenerational survival for over three centuries indicate that this feature of Majorcan Spanish cannot be dismissed as spontaneous learners' errors. The historical transmission points to a scenario of intra-community recycling of a linguistic innovation that emerged as the result of incomplete grammatical competence in Spanish and was then perpetuated by limited contact with monolingual Spanish speakers. The changing linguistic ecology of Majorca, however, with wider access to canonical varieties of Spanish in the last half century, is facilitating the demise of this and other traits among younger urban generations and its replacement with conventional Spanish structures. The combination of contemporary and historical sources presented in this study offers new insights to the study of a largely unknown language contact environment.

Acknowledgments

I am grateful to Inés Fernández-Ordóñez for her assistance in the compilation of the COSER-Mallorca interviews used in this study; to Miguel Calderón Campos for facilitating access to the 18th century texts in the CORDEREGRA corpus and to Laura Camargo and Beatriz Méndez Guerrero for their assistance in compiling and transcribing the PRESEEA-Palma corpus of sociolinguistic interviews. Also thanks are due to three anonymous reviewers for their helpful comments to a previous version of this paper.

Funding

The research reported here was funded by two grants from the Ministry of Economy and Competitiveness of Spain (AEI/FEDER, UE) references FFI2014-59135-R and FFI2017-83899-P.

References

Cabanes, Vicent. 1989. Ús de les preposicions a/en davant els noms propis de lloc: Segles XIII-XIV-XV-XVI. *Caplletra* 6. 141–170.

Casado-Fresnillo, Celia. 1995. Resultados del contacto del español con el árabe y con las lenguas autóctonas de Guinea Ecuatorial. In Carmen Silva-Corvalán (ed.), *Studies in language contact and bilingualism*, 279–292. Washington: Georgetown University Press.

Choi, Jenny K. 2001. The genesis of voy en el mercado: The preposition *en* with directional meaning in Paraguayan Spanish. *Word* 52(2). 181–196. https://doi.org/10.1080/00437956.2001.11432511

Enrique-Arias. Andrés. 2010. On language contact as an inhibitor of language change: The Spanish of Catalan bilinguals in Majorca. In Anne Breitbarth, Christopher Lucas, Sheila Watts & David Willis (eds.), *Continuity and change in grammar*, 97–118. Amsterdam: John Benjamins. https://doi.org/10.1075/la.159.05enr

Enrique-Arias, Andrés. 2014a. Efectos del contacto de lenguas en el castellano de Mallorca: una perspectiva histórica. In Andrés Enrique-Arias, Manuel J. Gutiérrez, Alazne Landa & Francisco Ocampo (eds.), *Perspectives in the study of Spanish language variation*, 271–297. Santiago de Compostela: Universidad. Anexos de Verba, 72.

Enrique-Arias, Andrés. 2014b. Lengua y escritura en la Mallorca del XVIII: el archivo epistolar de Cecilia Zaforteza. In Antonio Castillo Gómez & Verónica Sierra Blas (eds.), *Cinco siglos de cartas. Historia y prácticas epistolares en las épocas moderna y contemporánea*, 313–328. Huelva: Universidad.

Enrique-Arias, Andrés. 2019. Testing contact-induced change in the Spanish of Mallorca. Insights from a historical perspective. In Gabriel Rei-Doval & Fernando Tejedo-Herrero (eds.), *Lusophone, Galician and Hispanic linguistics: Bridging frames and traditions*, 168–183. London: Routledge. https://doi.org/10.4324/9781315403946-10

Escandell-Vidal, Victoria. 2007. Topics from Ibiza: Differential Object Marking and Clitic-Dislocation. In Georg A. Kaiser & Manuel Leonetti (eds.), *Proceedings of the workshop "Definiteness, Specificity and Animacy in Ibero-Romance Languages," 23–43. Arbeitspapiere Fachbereich Sprachwissenschaft*, 122.

Escobar, Anna María. 1990. *Los bilingües y el castellano en el Perú*. Lima: Instituto de Estudios Peruanos.

García, Mary Ellen. 1982. Syntactic variation in verb phrases of motion in U.S.-Mexican Spanish. In Jon Amastae & Lucía Elías-Olivares (eds.), *Spanish in the United States: Sociolinguistic aspects*, 82–92. Silver Spring: Institute of Modern Languages.

García, MaryEllen. 1995. En los sábados, en la mañana, en veces: A look at *en* in the Spanish of San Antonio. In Carmen Silva-Corvalán (ed.), *Spanish in four continents: Studies in language contact and bilingualism*, 196–213. Washington, DC: Georgetown University Press.

Granda, Germán de. 1988. Origen y configuración de un rasgo sintáctico en el español de Guinea Ecuatorial y en el portugués de Angola. *Anuario de lingüística hispánica* 4. 81–98.

Granda, Germán de. 1991. De nuevo sobre la causación múltiple en el español de América (a propósito de dos rasgos morfosintácticos del español paraguayo). In Germán de Granda (ed.), *El español en tres mundos. Retenciones y contactos lingüísticos en América y África*, 41–54. Valladolid: Universidad de Valladolid and Caja Salamanca y Soria.

Haboud, Marleen. 1998. *Quichua y castellano en los Andes ecuatorianos: los efectos de un contacto prolongado*. Quito: Ediciones Abya-Yala.

Hanssen, Federico. 1945. *Gramática histórica de la lengua castellana*. Buenos Aires: El Ateneo.

Kailuweit, Rolf. 1993. Prohibició i continuïtat del català com a llengua escrita i escolar – revisió d'alguns documents de Mallorca. In Antoni Ferrando & Albert Hauf (eds.), *Miscel·lània Joan Fuster. Estudis de llengua i literatura*, 141–161, vol. VII. Barcelona: Publicacions de l'Abadia de Montserrat.

Kvaal, Joy T., Nancy Shipstead-Cox, Susan G. Nevitt, Barbara W. Hodson & Patricia B. Launer. 1988. The acquisition of 10 Spanish morphemes by Spanish speaking children. *Language, Speech, and Hearing Services in Schools* 19. 384–394. https://doi.org/10.1044/0161-1461.1904.384

Lipski, John. 1990. *El español de Malabo: procesos fonéticos-fonológicos e implicaciones dialectológicas*. Madrid/Malabo: Instituto de cooperación para el desarrollo and Centro cultural hispano-guineano.

Lipski, John. 2016. On the tenacity of Andean Spanish: Intra-community recycling. In Sandro Sessarego & Fernando Tejedo-Herrero (eds.), *Spanish language and sociolinguistic analysis*, 109–134. Philadelphia: John Benjamins. https://doi.org/10.1075/ihll.8.04lip

Martínez i Taberner, Catalina. 2000. *La llengua catalana a Mallorca al segle XVIII i primer terç del XIX*. Barcelona/Palma: Publicacions de la Abadia de Montserrat and Universitat de les Illes Balears.

Melià, Joan & Maria del Mar Vanrell. 2018. *Enquesta d'usos lingüístics a les Illes Balears 2014. Anàlisi*. Palma: Govern de les Illes Balears/Generalitat de Catalunya/Universitat de les Illes Balears.

Minaya Portella, Liliana. 1978. Descripción sintáctica del habla del niño ayacuchano. *Lingüística y educación: Actas del IV Congreso Internacional de la ALFAL*, 464–477. Lima: Universidad Nacional Mayor de San Marcos.

Moll, Francisco de Borja. 1961. El castellano en Mallorca. *Homenaje ofrecido a Dámaso Alonso por sus amigos y discípulos con ocasión de su 60° aniversario*, 469–474. Vol. 2. Madrid: Gredos.

Oroz, Rodolfo. 1966. *La lengua castellana en Chile*. Santiago: Universidad de Chile.

Palacios, Azucena. 2019. La reorganización de las preposiciones locativas *a*, *en* y *por* en el español en contacto con guaraní. *Círculo de Lingüística Aplicada a la Comunicación* 78. 233–254. https://doi.org/10.5209/clac.64380

Romaine, Suzanne. 1995. *Bilingualism*. Oxford: Blackwell.

Sancho Cremades, Pelegrí. 2002. La preposició i el sintagma preposicional. In Joan Solà, Maria Rosa Lloret, Joan Mascaró & Manuel Pérez Saldanya (eds.), *Gramàtica del català contemporani*, 1689–1796. Barcelona: Empúries.

Silva-Corvalán, Carmen. 1994. *Language contact and change: Spanish in Los Angeles*. Oxford: Clarendon.

Silva-Corvalán, Carmen & Andrés Enrique-Arias. 2017. *Sociolingüística y pragmática del español*. Washington: Georgetown University Press.

Solà, Joan. 1980: Tractats de catalanismes. In Emilio Alarcos Llorach et al. (eds.), *Miscel·lània Aramon i Serra: estudis de llengua i literatura catalanes oferts a Ramon Aramon i Serra en el seu setantè aniversari*, vol. 2, 559–582. Barcelona: Curial.

Wheeler, Max, Alan Yates & Nicolau Dols. 1999. *Catalan: A Comprehensive grammar*. London: Routledge. https://doi.org/10.4324/9780203300275

Anthroponymic perseverance of Spanish vestigial <x>

Maryann Parada
California State University, Bakersfield

Until the early 19th century, the letter <x> was a common representation of the voiceless velar fricative /x/ (*Quixote*, *brúxula*). Despite the Royal Spanish Academy's (RAE) 1815 elimination of <x> for /x/, it has survived in a number of Mexican indigenous toponyms and their derivatives as well as in a handful of Spanish anthroponymic variants. Drawing on diatopic, diachronic demographic data, this paper traces the retention of vestigial <x> in six anthroponymic variants: the given name *Ximena* and the surnames *Ximénez/Ximenes, Mexía/s*, and *Roxas*. These forms can be attributed to the often exceptional orthography of proper nouns from a normative perspective. In the case of *Ximena*, a powerful resurgence of the feature is linked to robust indexicalities of <x> in Mexican society.

Keywords: Spanish, names, vestigial, spelling, identity, standardization

1. Introduction

In the border town of Juarez, Mexico, in the *Plaza de la Mexicanidad* 'Plaza of Mexicanness', stands a recently erected, nearly 200-foot-tall red monument in the shape of an <X>. Almost straddling the border, it contains an eye-shaped center viewing area and is known as *La Equis de Juárez* 'The Ex of Juarez' or simply *The X*. The 2013 'X-Fest' inaugurated the monument by Mexican sculpture artist Sebastián (Enrique Carbajal Gonzales) and featured musical artists ranging from the El Paso Symphony Orchestra to Juarez native Juan Gabriel (Candelaria 2013). The monument is described as "a symbol full of meanings, a mark to welcome people to Mexico and to demonstrate the ancestral Mexican culture" (Martinez 2015, para. 4). According to Sebastián, "the two intersecting arms of the 'X' symbolize the *mestizaje*, or the merging of two cultures, in Mexico – the indigenous people and the Spanish" while also representing an Aztec religious symbol: the

https://doi.org/10.1075/ahs.12.c09par

Nahui-Ollin (Martinez 2015, para. 5). The artist also intended the sculpture as "a tribute to former Mexican president Benito Juárez, who changed the spelling of the country's name in the 1800s, from *Méjico* to *México*. […] Being the first president with Aztec blood, Benito Juárez is credited with bringing together the races" (Martinez 2015, para. 5).

The graphemic saturation of <x> in Mexico's indigenous toponymy and anthroponymy, unrivaled elsewhere in the Spanish-speaking world, both reveals and reinforces its symbolic importance for that nation. The Juarez monument illustrates that the letter's connotations exceed the purely linguistic to encompass historical, political, ethno-cultural, and spiritual aspects of Mexico's past and present. But is the strength of the associations such that they hold in <x> variant non-indigenous names as well? This study examines the diachronic popularity of the Spanish baby name *Ximena*, giving special consideration to its significance in the Mexican context. In addition, to provide a broader picture and contextual backdrop of the anthroponymic presence of vestigial <x> across the Spanish-speaking world, the study also provides a frequency and density survey of a selection of <x> variant Spanish surnames.

Simply linguistically speaking, the <x> in Mexico and neighboring regions is complex. It has at least four pronunciations observed across different personal and place names. First, some indigenous names have maintained the native [ʃ] value of <x>, for instance *Xoco* ['ʃoko], a commercial district in Mexico City, *mixiote* [mi'ʃjote], a traditional pit-barbecued meat dish in central Mexico, and *Xela* ['ʃela], the second largest city in Guatemala. Second, in word initial position [s] is frequent, often observed in indigenous items with an internal [tʃ] (*Xochimilco, Xóchitl*, *Xochitepec*) but also in Greek-origin names (*Xiomara, Xenia*). The third and most prevalent pronunciation is that of [ks], sometimes alternatively realized as [gs] as a result of syllable coda stop neutralization, for intervocalic <x> (*Necaxa* [ne'kaksa], a main river in the Mexican state of Puebla) or implosive <x> (*Tuxpan* ['tukspan],[1] a municipality and city in the Mexican state of Veracruz). In Spain, by contrast, implosive <x> is frequently realized as [s] (*exponer*, [espon'er]) (RAE 2005), reflecting the final stage of syllable final lenition: elision. The fourth, vestigial pronunciation as [x][2] (*México*, *Oaxaca*) will be discussed in the next paragraph and in evolutionary detail in Section 1.3. This realization of <x> is most intimately tied to specific high frequency Mexican indigenous toponyms and their derivatives, but, interestingly, for a handful of Spanish toponyms with a final <x>

1. In Nahuatl it is pronounced ['tuʃpan].

2. [x] is used because it is the most frequent realization of /x/ in Mexico. The author acknowledges [h] and [χ] (among others) as standard pronunciations in other regions.

typically pronounced [ks] or [s], an underlying /x/ emerges in their corresponding demonyms: *Sax* (sajeña/o) in Valencia, *Borox* (borojeña/o) in Castile-La Mancha, *Guadix* (guedijeña/o) in Andalusia, but *Tolox* (toloseña/o) and *Torrox* (torroseña/o), also in Andalusia (RAE 2005).

Beyond its nationalistic and ethnic indexicalities in the context of Mexican indigenous toponomy, /x/ represented as <x> also appears in a handful of Hispanic anthroponymic variants that, in theory, extend across the Spanish speaking world. This use of <x> simply represents a vestige of Old Spanish. However, little is known of the present geographical distribution and incidence of these items, nor of their indexical relationship to Mexican indigeneity. This exploratory two-part study examines the persistence of vestigial <x> in two name types: first, in the feminine given name variant *Ximena* and, second, in five surname variants: *Ximénez, Ximenes, Mexías, Mexía*, and *Roxas*. Though there are certainly other /x/ as <x> variants in use (especially of given names – *Xavier* in variation with *Javier* in México comes to mind), these six appear to be by far the most prevalent, as they are frequently referenced across a variety of sources and in popular commentary. Though the study's main focus in on the given name *Ximena*, the surnames are also analyzed for context as to the broader incidence and distribution of the orthographic variable as well as for insight into the potential correlations between <x> variant given names and <x> variant surnames. The distinction between the two name types must not be underestimated, however, given the vastly different processes by which they are bestowed and propagated.

Anthroponyms are key in examining the continuance of vestigial <x> for several reasons. It is only through names and name derivatives that the feature persists in official usage. Surnames today, at least in Western cultures, are characterized by a nearly unassailable patronymic perpetuation. (Though it is rare and the legal process varies internationally, names can be legally changed) (Alford 1988). Thus the current geographical distribution and incidence of the <x> variant surnames in question are likely to remain fairly stable for generations to come. By contrast, given names involve a process of selection underpinned by the forces of parental values and societal tastes and associations, which reflect the feature's overt significance in a given population. The remarkable trending of *Ximena* in recent years, most drastically in Mexico where the symbolic status of <x> seems at play, extends the feature beyond passive traditions. It suggests that non-indigenous names may be imbued with the emblematic force of <x> traditionally and typically conveyed through indigenous toponyms, anthroponyms, and ethnonyms.

2. The evolution of Spanish <x>

Until the early 19th century, the voiceless velar fricative /x/ was commonly represented orthographically by the letter <x> in addition to <j> (before any vowel) and <g> (before <e> and <i>). This is observed in the former spelling of names like *Quixote* (retained in English versions of the famed text) and of common words like *brúxula* (brújula) 'compass', *exemplo* (ejemplo) 'example' and *dexar* (dejar) 'to leave/let (something/someone)'. Such items were mostly patrimonial Latinisms, but there were also Arabisms (e.g. *xarabe* (jarabe) 'syrup'). Contrastively, later Latinate literary borrowings containing <x> preserved the Latin-like value [ks], e.g. *exquisito* 'exquisite.'

The confluence of <x>, <j>, and <g> in their expression of /x/ was a result of earlier sound changes. In medieval Spanish, <x> was used to represent the voiceless pre-palatal fricative (one of six sibilant phonemes), a sound akin to the modern English <sh> or French <ch> (/ʃ/).[3] Thus, the <x> in words like *lexos* (lejos) ('far') or *xabón* ('jabón') was pronounced [ʃ]. In colonial Latin America, indigenous toponyms containing a sound similar to medieval Spanish [ʃ] were transcribed by the Spanish using <x>: *México*, *Texas*, *Xalapa*, *Oaxaca*, *Xalisco*, *Xilotepec*, etc.

In medieval Spanish there was also a voiced pre-palatal fricative (/ʒ/) expressed orthographically by <j> or <g> (only before <e> and <i>). Accordingly, words like *fijo* (hijo) 'son' and *muger* (mujer) 'woman' were pronounced ['fiʒo] and [mu'ʒeɾ], respectively (RAE 2010). In the 16th century, as part of a trend in which voiced sibilants devoiced and merged with their voiceless counterparts, /ʒ/ and /ʃ/ converged (Pharies 2015). According to Penny (2002, p. 99), sibilant devoicing appeared first in the northern Castilian speaking territory. Around the middle of the 16th century many northerners migrated to Madrid, the new capital, occasioning a southward spreading of the feature.

By the middle of the 17th century, another important sound change had taken place: the velarization of /ʃ/. This was normalized in all varieties of Spanish (except Judeo-Spanish, which existed outside the reach of such standardization forces), producing /x/. Despite their homophony as voiceless velar fricatives, the graphemes <x>, <j>, and <g> went on to coexist in the writing system until the RAE's publication of the eighth edition of the *Ortografía* in 1815, which marked the official directive to substitute <x> with <j> in all words that had undergone the consonant shift (Rosenblat 1960: 139), reserving <x> for [ks]. Accordingly, a word like *embaxador* 'ambassador' was standardized as *embajador* and *crucifixo* 'crucifix' as *crucifijo* (RAE 2010).

---

3. In most of Spain's official or regional languages (Asturian, Leonese, Galician and Catalan) and Portuguese, [ʃ] is a possible pronunciation for <x>. Portuguese has four possible pronunciations for <x>: [ks, s, z, ʃ].

The change applied to names as well, with most eventually losing the <x> in favor of <j>: e.g., *Javier*, *Guadalajara*, *Jalisco*, *Jamaica*, *Juárez*, and *Jaramillo*. But because "*los nombres propios son más conservadores que las voces corrientes*" ('proper nouns are more conservative than common nouns') when it comes to normative reform (Rosenblat 1960: 140), /x/ as <x> has survived in many Mexican indigenous toponyms and their derivatives (e.g., *México*; *oaxaqueño*; *Bexar County*, *Texas*), as well as in highly infrequent (in the case of surnames) variants of certain Spanish anthroponyms (e.g., *Ximena*, *Ximénez*/*Ximenes*, *Mexía*/*s*, *Roxas*). However, until the Mexican Revolution, when <x> is believed to have acquired special status as a national symbol, there appears to have been a great deal of alternation between the <x> and <j> variants of *México* and other place names (Moreno de Alba 2003; Rosenblat 1960). In fact, in Spain, until relatively recently, the toponyms *México*, *Oaxaca*, and *Texas* and their derivatives were normatively spelled with a <j>. Such variants are still considered valid (and more orthographically consistent), but since 1999 the recommendation of the RAE is to defer to the use of <x> in these and other indigenous toponyms, in accordance with preferences in Mexico and across Latin America (RAE 1999). Villa's (2015: 228) analysis of three ideological debates around Spanish orthography during the middle of the 19th century supports the notions of "spelling as an identity marker and political tool" which result in periods of alternation between different orthographies.

2.1 <x> in vestigial variants of personal names

As stated, this study probes the retention of vestigial <x> through comparative frequency analyses of the given name *Ximena* and five Spanish surnames: *Ximénez, Ximenes, Mexía, Mexías, Roxas*. Although these names also occur as toponyms and trade names, the focus of this paper is on their incidence as personal names. They are oft-cited examples of vestigial <x> yet little is known about their prevalence and perceptions in the Hispanic world. Again, that the items reflecting the "archaic" (vestigial) spelling are names is not surprising, given that in phonological and orthographic evolution names often behave in non-standard, de-regulated ways (Carney 1994, Rosenblat 1960, Ryan 2016, Sebba 2007). This is because names are more bound up with identity, making change less appealing and less necessary (i.e., a visible etymological connection to related words is of little functional importance).

The first and main objective was to comparatively trace the diachronic popularity of the given name *Ximena*, through an analysis of baby naming trends, across the three largest Spanish speaking populations of Mexico, the United States (U.S.), and Spain. The analysis examines the popularity of *Ximena* relative to its variants *Jimena* and *Gimena* over the last eight decades. In contrast with surnames, which are typically inherited as a matter of course, given names reflect parental values and

tastes, which are informed by broader socio-cultural trends. Thus, the data have the potential to reveal both patterns of retention and resurgence of the feature, inviting consideration of the socio-cultural underpinnings of such trends. For broader context on the onomastic incidence of the feature, the second objective was to provide a simple synchronic pulse on the diatopic frequency and distribution of the five <x> variant surnames. More specifically, this analysis assesses the present-day, international distribution and prominence of each of the surnames in relation to their variant counterparts. These data have the potential to shed light on the relationship between the incidence of <x> variant surnames in a given society and that of <x> variant first names. In addition, these first-stage analyses will provide a foundation for future cross-sectional quantitative work on the parental selection of <x> variant baby names, as well as for qualitative approaches to exploring the situated use, attitudes, and folk understandings of such names.

Before proceeding to the next section, a brief overview of each name/name pair is in order.

2.1.1 *Ximena*

Ximena is the Spanish feminine equivalent of the Hebrew biblical name *Simon*, meaning hearkener, listener (Behind the Name). Variants are *Jimena*, *Gimena*. Masculine forms, which are out of use but derived from *Ximénez* (*Jiménez, Giménez*), are *Ximeno, Jimeno*, and *Simón*. *Ximena Díaz* was the wife of the Spanish national hero known as *El Cid*. Several modern Latin American celebrities bear this name, including *Ximena Duque* and *Ximena Sariñana*. Mexican actress and 2010 Miss Universe *Jimena 'Ximena' Navarrete* unofficially swapped the <j> of her name for <x> following her crowning. *Ximena Sinfuego* was a recurring character on the U.S. ABC series *The Fosters* (2013-2018). Toponymic/commercial instances of the masculine variant include *Ximeno Avenue* in Long Beach, California, U.S.A. and *Ximeno Records* in Los Angeles, California.

2.1.2 *Ximénez/Ximenes*

Ximénez and *Ximenes* are patronymic surnames whose variants are *Jiménez, Giménez, Jimenes*, *Gimenes*. The 12th century Navarrese-born Archbishop *Rodrigo Ximénez de Rada* is a well-known figure in the history of Spain. *Miguel Ximénez* and his son *Juan Ximénez* were Spanish Renaissance painters. *Miguel Ximénez* is a Uruguayan footballer who currently plays for Club Cienciano of Peru. In Spain today, the name may bear special ethnic connotations. Gamella et al. (2014: 159) found that *Ximénez* was the second most frequent Gitano surname in the 1783–1785 Spanish censuses, affirming, "the surnames found in these censuses are the same that Gitanos still carry today." According to the study, when Spanish patronymics like *Ximénez* came to identify Gitanos, "they were infused with new meanings,

as they became tokens of ethnic identity and cultural distinction" (Gamella et al. 2014: 160). Thus, particularly in Andalusia where Spain's Gitano population largely resides, *Ximénez* may be popularly associated with Gitano lineage. Toponymic/commercial instances of the name include *Pedro Ximénez* grapes in Spain and *Colegio Mayor Ximénez de Cisneros*, a residential school in Madrid.

2.1.3 *Mexía/s*

Mexía and *Mexías* are Sephardic patronymic surnames from the Hebrew *mashiaj* 'annointed': 'messiah' (Tibón 2001). Variants include *Mejía/s, Macías, Mesía. Álvaro Mexía* was a 17th-century Spanish explorer and cartographer stationed near San Agustín, Florida. *Ynes Mexía* was a Mexican American botanist who is believed to have collected around 150,000 specimens before her death in 1938. Curiously, she lived much of her life in Mexia (now standardly pronounced [məˈheɪə]), Texas, named after *José Antonio Mexía*, a general for the Republic of Texas Army during the Texas Revolution. According to the city website, Mexia's motto is "A great place, no matter how you pronounce it" (City of Mexia), presumably referencing outsiders' mispronunciation as [ˈmɛksiə]. Among its many public and commercial entities bearing the name are the *Mexia Unified School District* and *The Mexia News*. There is also a nearby *Lake Mexia*.

2.1.4 *Roxas*

Roxas is a habitational Spanish surname derived from *Roxo* (rojo) 'red' (Tibón 2001). The most common variant is *Rojas*. It is prevalent in the Philippines, a former Spanish colony, where it is borne by politicians (e.g. *Mar Roxas*), including a former president (*Manuel Roxas*), actors (e.g. *Van Roxas*), municipalities, a city, street names, businesses, and an airport.[4]

2.2 Socio-cultural significance of <x> in México and broader valorizations

Decades of work on the relationship between language variation and society have "shown very clearly that wherever choices are possible, or made possible, they have the potential to take on social meaning – and usually do so" (Sebba 2012: 1). Scripts and orthographies, although understudied in the sociolinguistic and sociology of language traditions, which have overwhelmingly focused on spoken language, also exhibit socially meaningful variation and performative functions (Sebba 2007; Jaffe, Androutsopoulos, Sebba & Johnson 2012). Regarding the marked nature of /x/ as <x> in Mexican indigenous names, Venezuelan philologist Angel Rosenblat

4. In contrast, *Rojas* figures as the 30,817th most frequent name in the Philippines, with only 437 bearers.

(1960: 140), reflected on how symbolically curious it was that an orthographic "archaism" had come to index transformation and progress in the Mexican context:

> Me dicen que en México se ha hecho de la x bandera de izquierdismo. Y que, en cambio, la j es signo de espíritu conservador o hispanizante. El hecho se presta para una filosofía de los símbolos, porque es curioso que el espíritu renovador, del que se podía esperar una atrevida modernización ortográfica, se aferre a una grafía arcaica. Y que sean los conservadores o tradicionalistas los partidarios de la j moderna.
> 'I am told that in Mexico, *x* has become a leftist symbol. And that, conversely, *j* is a sign of conservative or Hispanizing spirit. This fact lends itself to a philosophy of symbols, since it is curious that a spirit of renewal, which one might expect to go hand in hand with bold orthographic modernization, is bound up with an archaic letter, and that conservatives or traditionalists are the defenders of the modern *j*.'

Indeed, the onomastic retention of vestigial <x> once evoked significant debate among writers, grammarians, philologists, and politicians alike. Critics cited the orthographic inconsistency it presented as well as the unwarranted special status it granted to Mexico (summarized in Moreno de Alba 2003). However, the victory of the <x> variant for many Mexican toponyms (most notably *México*), despite critique from across the Spanish speaking world, including oppositional RAE stances until 1999, suggests a type of reappropriation and indexical reversal at the orthographic level. Today in Mexico the letter continues to be creatively and emblematically employed in the performance of collective and individual identities, such as in the Juarez monument previously discussed, recent campaigns to replace D.F. with CDMX in signage, national tourism marketing, and in the rise of <x> in baby naming.

The attachment to <x> in Mexico is better understood by considering how identity and iconicity are conveyed through spelling. Sebba (2007: 161), in his comprehensive treatment of orthography as social practice and postcolonial orthographies, stresses "contested orthographies should be viewed as sites of contested identities" and suggests that marginalized or oppositional linguistic features wind up as prime options for identity stances of resistance. Thus, vestigial <x> has only been able to iconify Mexican identity through its known opposition to more widely used and standardized alternatives (mainly <j>), as well as through its associations with indigenous heritage. In this sense, it was not arbitrarily selected. Though /x/ as <x> is a vestige of Old Spanish, its purposeful and persistent presence of /x/ as <x> in Mexico's toponomy and anthroponomy has signaled more of a cultural and socio-political resistance than linguistic. Affinity toward a particular non-endemic letter is not unusual. The Basque display a symbolic resistance to mainstream Spanish culture through their attachment to <k> over Spanish <c> (Sebba 2007: 50) as do factions within the pro-indigenous Mexican-American movement (following orthographic tendencies in modern Nahuatl) (mexika.org; Tlalpoyawa 2000). Because the indexicalities of linguistic features, including graphemes, are subject

to continual flux and renegotiation (Agha 2003; Johnstone et al. 2006; Jaffe et al. 2012), origins are often unimportant or obscured and specific trajectories of use rather unpredictable (Jaworski 2019).

As the most basic linguistic forms, letters are easily semiotic and lend themselves to folk interpretations and applications. We might consider the declared symbolisms of the newly erected 'X' monument in Juarez described earlier as well as those outlined in a recent Mexican government-sponsored competition, titled '*El valor de su* X,' which used the story of the <x> in *México* to invite Mexican artists to tap into the determination and fighting spirit of their Mexican heritage in order to design an <x> for the logotype of the official www.mexico.mx website and its related social media accounts ('Convocatoria: México, el Valor de su X' 2016). 'Xicana' feminist writer and activist Ana Castillo (1995: 12) has argued for <x> in the spelling of the ethnonym to "reclaim indigenism" by paying "homage to the Nahuatl language of the Mexica and their use of <x>". Socially motivated orthographic resistance and affinities tend to be more visible in names, given the peripherality of names relative to language norms and standardization processes, as well as their discursive salience (Puzey & Kostanski 2016).

According to Jaworski (2019: 116), <x>, being "persistent[ly] linked [to] specific types of products, services, and activities," is especially salient and "prone to manipulations" in the global semioscape (Thurlow & Aiello 2007). It has been characterized as "the sign of signs" and the "absolute embodiment of symmetry" (Frutiger 1998; Gale 2015: 71–72). According to Jaworski (2019), despite (or perhaps because of) the rarity of <x> in Latin-based alphabets, it has experienced wide and salient usage in branding, signage, and advertising copy. The letter is a common symbol for the unknown, including the mysterious, the anonymous, the generic, and the experimental, whether negative or positive, from popular culture to science (e.g. the X factor, x-ray, X marks the spot, x rated, The X-Files, X-Men, Malcolm X, Latinx) (Sacks 2003). Following the intensifying linkage between the taboo and <x> in previous decades, the letter began to undergo mainstreaming and legitimization in the 1990s, principally in its use for the marketing of cutting-edge products, especially related to software (Sacks 2003). Pelkey (2017) describes the X-mark as a culturally significant chiasmus figure due to its relation to the spread-eagle body position, "capable of signifying a multitude of frequently oppositional communicative acts: severe warning, triumphant elation, humiliating defeat, exertion in a fitness routine, pain and exhaustion under torture, among others" (Jaworski 2019: 119). In recent years, these multiple meanings and functions have been heavily exploited in stylized brands and logos containing <x>. This trend toward positive associations and creative uses of the letter has likely further facilitated its important presence in Mexican baby naming and as a focal symbol in country and city name branding efforts in Mexico in recent decades.

3. Methods

To research the given name *Ximena* in relation to its variant counterparts, sources providing general diachronic trends were first consulted. The Google Books *N*-gram corpus (Michel et al. 2011) is considered a useful starting point in sociolinguistic research to broadly identify trends (Meyerhoff et al. 2015). Specifically, the *N*-gram Spanish 2009 corpus was searched, which includes popular and academic Spanish language works[5] published between 1800 and 2008. In addition, in a worldwide demographic count of the name *Simona*, Mollerup (2015) provided open source frequency data that include its Hispanic variants *Ximena* and *Jimena*.

Next, official population data were consulted for Mexico, the U.S., and Spain, the top three countries with respect to number of Spanish speakers.[6] Data inclusive of the top 1000 names for female and male births for each year between 1950 and 2016 were requested and granted by Mexico's Registro Nacional de Población e Identificación Personal (RENAPO), managed by the Secretaría de Gobernación (SEGOB). Comparable U.S. data were obtained from the Social Security Administration (SSA). Each calendar year, the SSA publishes national data on the relative frequency of given names in the population of U.S. births (where the individual has a Social Security number), and state-specific as well as U.S. territory lists are also available. To protect privacy, the lists are restricted to names with at least five occurrences in the population. The lists do not combine different spellings of similar names. Thus, the variants *Ximena*, *Jimena*, and *Gimena* each appear alongside their individual annual frequency count and rank. These were analyzed over eight decades, from the 1930s through the first half of the 2010s. Their ranking trajectories in specific Southwestern and other Latinx-heavy states between 2010 and 2016 were also evaluated. Rankings of the name variants for Puerto Rican births were also accessed and analyzed, though independently from the U.S. data; Births in U.S. territories are not included in the SSA national data but are provided separately and in a more limited fashion (only the annual top 100 names since 1998). Similar to the U.S., Spain's *Instituto Nacional de Estadística* (INE) provides baby name data by year and province through as far back as 1930. Thus, the three name variants were again measured by decade, starting with 1930.

To investigate the five surnames, distributional and frequency data on surnames were obtained from various online data sources, including genealogy portal

5. According to Zhang (2015), the changing composition of the corpus over time can be problematic. For example the pre-20th century corpus is sermon heavy, while scientific literature is increasingly overabundant.

6. The U.S. is home to nearly 60 million Hispanics, comprising approximately 18% of the population (Noe-Bustamante & Flores 2019).

Forebears.io, United States Census Bureau, Spain's INE, and Mexico's *Instituto Nacional Electoral*. While all the previous sources were consulted, the Forebears database provided the most comprehensive international data, which is what is principally reported in this study. The site launched in 2012 with the aim of providing worldwide frequencies and a wealth of other information for thousands of surnames. To do so, they compile a variety of sources, including genealogical records, and data from newspapers, phone books, and electoral roles. Thus, the frequencies cited from this database should be considered estimates only.

4. Results

Figure 1 displays the trajectories of *Ximena* and variants based on the Spanish 2009 Google corpus of Spanish language books. The diachronic comparison provides a point of reference for the frequency trends of the variants across the Hispanic world, measured by percent of total words per year. The variant *Jimena* emerges as the most consistently used over time, and *Gimena* the least. *Ximena* has experienced steady growth in recent decades, surpassing *Jimena* in the late 1980s. Its spike in the early 19th century seems tied to a heavy density of references to Ximena of *El cantar de mio Cid*, based on a cursory look through the period's corresponding texts.

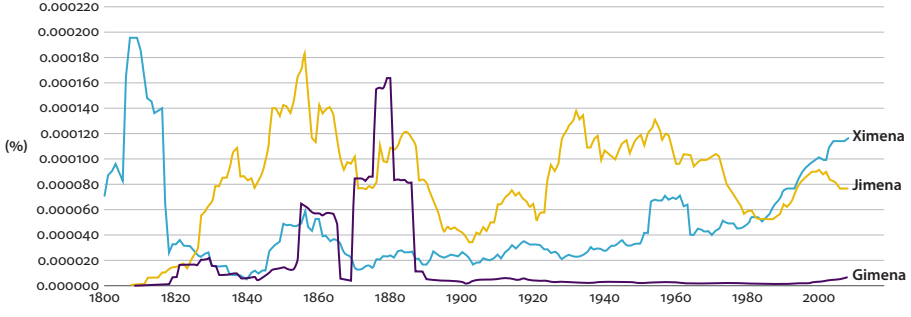


Figure 1. Frequencies of *Ximena, Jimena*, and *Gimena* in the Spanish 2009 Google book corpus from 1800 to 2008. The graph was made with the Google Books Ngram Viewer (Michel et al. 2011), with a smoothing of 5

The map in Figure 2 likewise points to the impressive contemporary popularity of the name. It shows the countries where the variants *Ximena* or *Jimena* are documented as high ranking (i.e., among the 100 most frequent names, either in the population or among newborns) in recent years (c. 2014). *Ximena* clearly predominates in the Americas while Spain favors *Jimena*.

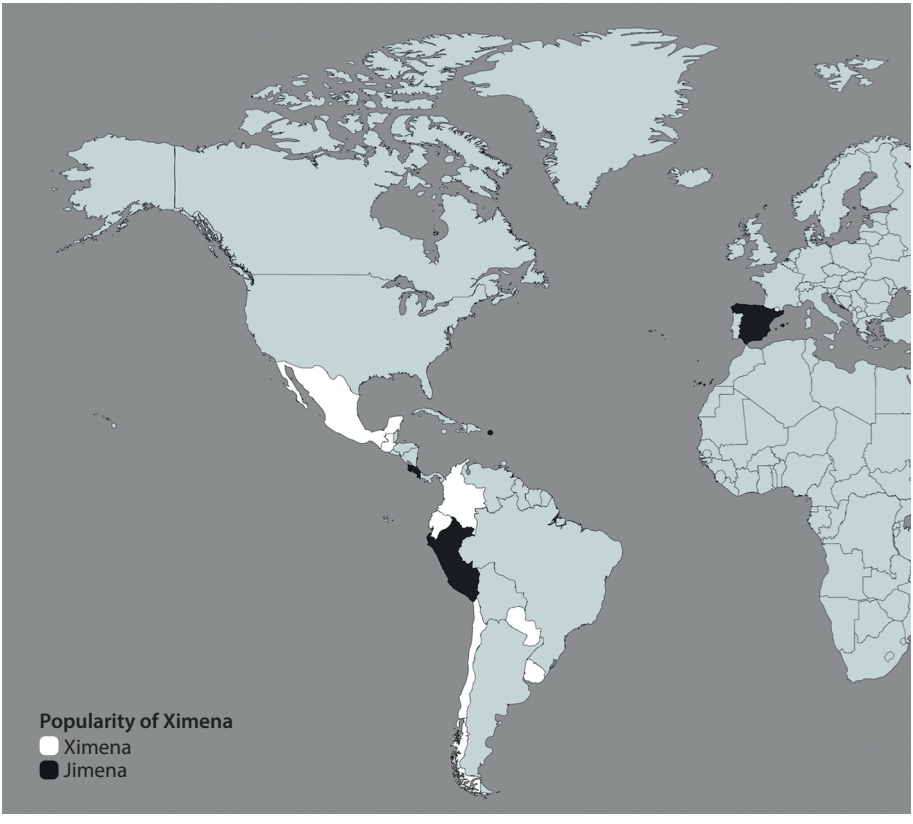


Figure 2. Popularity of *Ximena*. Source: Mollerup (2015)

We will now turn to country-specific baby naming data. In Mexico, as represented in Figure 3, a variant hierarchy similar to that of the Google Books corpus is observed. In terms of ranking, since 2007 *Ximena* has hovered at the very top of the charts. Only in 2009 and 2010 did it descend to second ranked, after *María Fernanda*. It has been among the 10 most popular baby names since 2001, among the top 100 since 1985, and among the top 1000 since 1987. Since 2000, a growing variety and frequency of composites with *Ximena* as a first or second name (e.g., *Ximena Guadalupe*; *María Ximena*) among the top 1000 further magnifies the exceptional prominence of the variant in contemporary Mexico. *Jimena* entered the top 1000 in 1981, slightly earlier than *Ximena*, but did not enter the top 100 until 2001, from which point it has remained among the top 20. The variants' unremarkable ranking differences mask their significant frequency division. In 2016, *Ximena* was given to a full 4% of all baby girls born in Mexico while *Jimena* was bestowed on .06%.

Figure 3 displays the top 1000 ranking trajectory of the variants since 1980. The composite forms were tallied and added to the simple form counts for each year

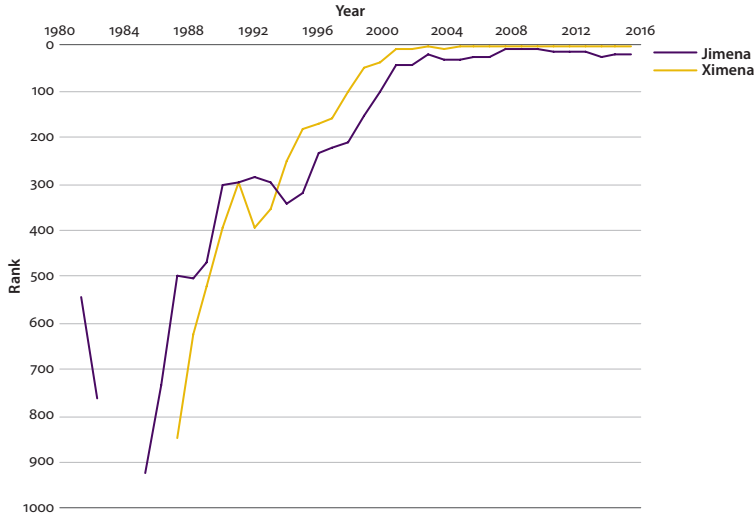


Figure 3. Top 1000 variant rankings for Mexican births, 1980 to 2016. Source: RENAPO

from which revised rankings were determined. However, because of the already extreme popularity of both variants, the addition of the various composites to the simple form counts altered the rankings very little. Since the year 2000, when the name began picking up speed, composite forms have accounted for 32% of the total *Ximena* tokens but only 13% of the total *Jimena* tokens. An analysis of the composites is provided in Table 1. While the *Jimena* composite types totaled only three, there were 45 for *Ximena*. By far, *Ximena Guadalupe* has been the most popular form, representing 7.4% of all Ximena tokens since the year 2000. At 9%, *Jimena Guadalupe* is likewise the most frequent composite for that variant.

Table 1. Summary of *Ximena*/*Jimena* composites for Mexican births, 2000–2016

Variant	As first name Types: Tokens	Highest frequency forms	As second name Types: Tokens	Highest frequency forms
Jimena	1: 4,593	Jimena Guadalupe (9%)	2: 2,108	María Jimena (3%) Karla Jimena (.08%)
Ximena	26: 43,381	Ximena Guadalupe (7.4%) Ximena Alejandra (1.8%) Ximena Mon(t)serrat(h) (1.4%) Ximena Sara(h)í (1.3%) Ximena Valentina (.08%)	19: 20,264	Karla Ximena (1.9%) María Ximena (1.9%) Ana Ximena (1.2%) Alexa Ximena (.09%) Danna Ximena (.08%)

Table 2 provides a frequencies summary by decade for the two variants (inclusive of composites) in which we observe the exponential growth of *Ximena* at the start of the 21st century relative to *Jimena* (obscured in the ranking data). Data for the first half of the current decade is reported showing the extreme between-variant frequency divergence. As of 2015, the *Jimena* counts were a mere fifth of the corresponding *Ximena* counts.

Table 2. Raw frequencies of *Ximena* and *Jimena* in México since 1930. Note: Frequencies are only available for names figuring in the top 1000 in any given decade (which did not include the variant *Gimena*)

Decade	Ximena	Jimena
1930	–	–
1940	–	–
1950	–	–
1960	–	–
1970	–	–
1980	488	1,575
1990	6,156	5,527
2000	80,334	28,670
2010–2015	98,870	19,096

Turning to comparable trends in the U.S., Table 3 displays the raw frequencies of the two variants (plus those of *Gimena*) for births in the U.S. by decade since 1930. Like Mexico, in the 1990s a trend favoring *Ximena* emerges, producing an identical hierarchy. *Ximena* debuted on the top 1000 baby girl name charts in 2001. Inclusion in that list is significant because it accounts for a majority of names given to baby girls. For example, in 2016 the top 1000 names accounted for approximately 73 percent of all U.S. baby girl births. The frequency of *Ximena* in the U.S. just in the first half of the 2010s is remarkable and points to a similar but slightly delayed popularity divergence from *Jimena* as seen in Mexico.

Table 3. Raw frequencies of *Ximena*, *Jimena*, and *Gimena* in the U.S. since 1930

Decade	Ximena	Jimena	Gimena
1930	0	0	0
1940	0	0	0
1950	0	0	0
1960	5	5	0
1970	56	5	0
1980	89	127	0
1990	385	371	16
2000	6,792	5,694	216
2010–2015	11,027	4,244	113

Figure 4 shows how in a span of 15 years *Ximena* has advanced in ranking by more than 600 positions, occupying the 118th slot in 2016 with 2,673 babies given this name. (Combining all spelling variants, the name would have ranked 80th in 2016.) The exponential growth of *Ximena* since 1980 parallels immigration patterns from neighboring Mexico and thus mirrors trends there.

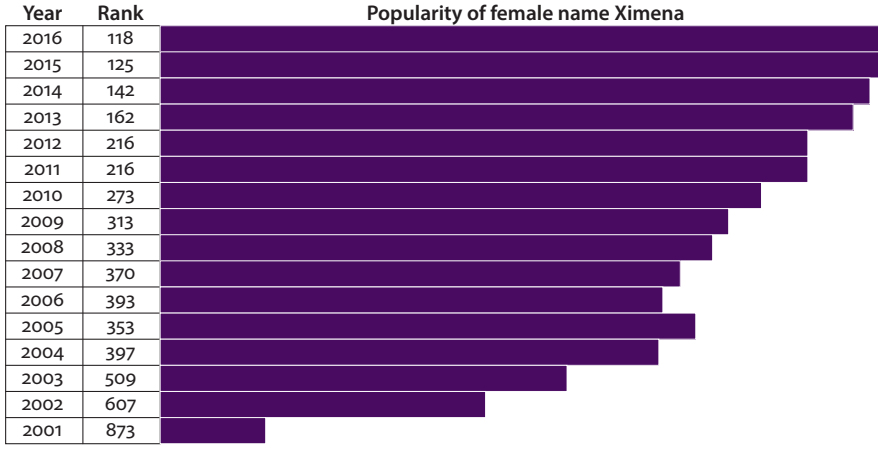


Figure 4. Popularity ranking of *Ximena* since 2001 in the U.S. Source: Social Security Administration. Note: Longer bars indicate higher ranking

For a regional U.S. perspective, Figure 5 displays the change in rank of *Ximena* since 2010 in four Southwestern states. The variant has reached the top 50 in all four states, most vigorously in New Mexico where it ranked 11th in 2016. (New Mexico, while not the Southwestern state with the largest Hispanic population, has the largest proportion of Hispanic residents at 47%.) In nearby Colorado and Nevada, it ranked 76th and 77th. The trend is observed outside of the Southwest as well, in areas with significant Mexican-American communities. For instance, in

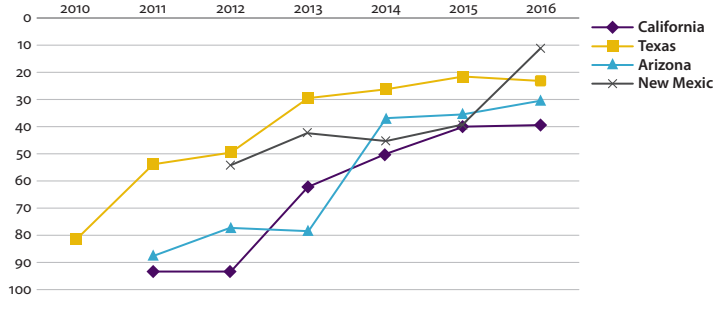


Figure 5. Rise in rank of *Ximena* in four Southwestern states, 2010–2016. Source: Social Security Administration

Illinois it ranked 85th and in Oregon, 89th. As of 2016, it had not appeared in the top 100 in Florida or New York, states with sizeable Latinx, yet not predominantly Mexican origin, populations. Puerto Rico, like Florida, New York, and other northeastern states, has not seen the same level of captivation with the name. Analyses of Puerto Rican baby naming from 1998 to 2016 revealed that it was not until 2015 that *Ximena* appeared on the top 100 charts, occupying the 96th slot, but it fell out of the top 100 in 2016.

Turning to Spain, Table 4 provides the frequencies of the three variants among all residents according to their year of birth. Like the previous Mexico and U.S. data, the years span from 1930 to 2015. According to the INE, practically all individuals named *Ximena* born before 2000 were born abroad. For example, 91% of the *Ximenas* born in the 1960s currently living in Spain were not Spanish-born. Likewise, 96% of those born in the 1980s were born abroad. However, for those born in the 2000s, only 16% were foreign born. Currently, out of 786 total residents named *Ximena*, 581 (74%) are born in Spain. Since 2000, *Ximena* is most often given to babies born in the province of Madrid, followed by Galicia. This pattern is likely due to the popularity of the name among the large Latin American (particularly South American) immigrant demographic in Madrid, given their mass arrival beginning around this time, as well as to *Ximena* as a Galician name in Galicia characterized by a distinct pronunciation. Importantly, as displayed in Table 4, the variant hierarchy differs from that of Mexico and the U.S. in the strong favoring of *Jimena*.

The general popularity of *Jimena* is reflected in its considerable growth over the last two decades, with INE ranking it as the 31st most popular baby name in 2016. The disparate average ages of the variants' bearers suggest the same: 21.6 years for *Ximena*, 19.9 for *Gimena*, 7.9 for *Jimena*. Though noteworthy, the recent popularity of *Jimena* in Spain is still but a shadow of the super status of *Ximena* in Mexico in recent decades.

Table 4. Frequencies of *Ximena* and variants in Spain since 1930. Source: INE

Decade	*Ximena*	*Jimena*	*Gimena*
1930	0	10	6
1940	0	14	0
1950	7	17	0
1960	23	30	0
1970	147	239	24
1980	145	368	32
1990	88	518	20
2000	213	5,492	70
2010–2015	157	7,013	76

We will now turn to the surname data. Displayed in Table 5 are four variants of the surname *Mejía* and their corresponding international profiles in comparison. *Mexía* is an infrequent form of the widespread last name, most prevalent in Mexico but also found in Portugal and Greece (*Mexias*).

Table 5. Worldwide frequencies of four variants of the surname *Mejía*. Source: https://Forebears.io. Note: These names in non-Spanish speaking countries typically do not bear an orthographic accent

	Estimated frequency	International rank	Most prevalent	Highest density
Mejía	1,008,194	495	Mexico	Honduras
Mejías	95,779	5,707	Venezuela	Venezuela
Mexía	5,630	78,315	Mexico	Portugal
Mexías	241	771,381	Greece	Greece

Looking more closely at the <x> variants, *Mexía* and *Mexías*, Tables 6 and 7 summarize their frequencies and rankings by country. The Forebears data presented in Table 6 shows that Mexico leads with the most occurrences of *Mexía*. There are an estimated 262 individuals with the surname *Mexía* in the U.S., likely correlated with the name's incidence in Mexico. The 2010 U.S. census reports *Mexía* with a slightly higher absolute frequency of 326 individuals, 85.3% of whom claim Hispanic heritage. Likewise, Spain's 2016 INE *Apellidos y nombres frequentes* 'Frequent Surnames and Names' report shows larger frequencies than Forebears for the name: 59 individuals have it as a first surname and 40 as a second surname versus the 70 listed by Forebears. Those numbers suggest that in Spain it is more often patrilineal, especially in the Asturias province (Figure 6), where the pronunciation of <x> is likely in line with Asturian (see Endnote 2) and thus not tied to the phenomenon of vestigial <x>.

Table 6. Frequencies of the surname *Mexía* by country. Source: https://Forebears.io

Country	Estimated frequency	Relative frequency (No. / million)	Rank in nation
Mexico	3,978	30	2,240
Portugal	1,210	115.5	799
United States	262	0.8	61,585
Brazil	79	0.4	92,589
Spain	70	1.5	33,508
Canada	9	0.3	212,166
Angola	3	0.2	2,935
Colombia	3	0.1	14,126
Peru	2	0.1	24,269
Greece	2	0.2	31,767

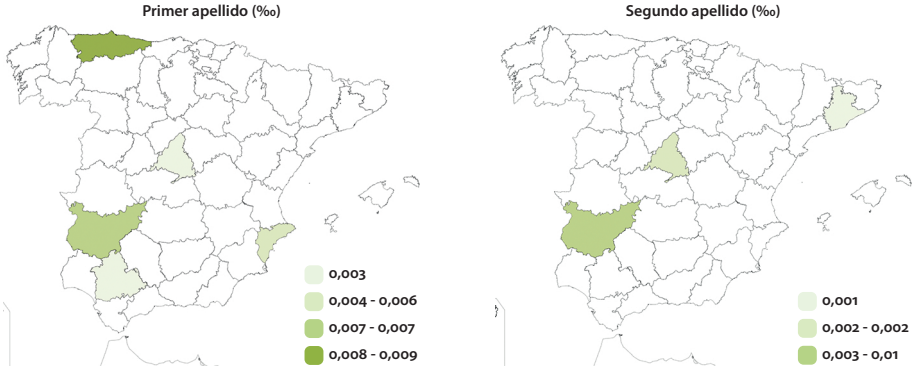


Figure 6. Percent of residents with *Mexía* as a first or second surname in Spain's provinces. Source: Instituto Nacional de Estadística (INE)

The variant *Mexías* is an extremely infrequent surname worldwide and, as shown in Table 7, is in fact unregistered in Spanish speaking countries. This is confirmed by a search of the U.S. 2010 surname census data, which turned up zero trace of the variant (U.S. Census Bureau 2016), as did a similar search of Spain's 2016 INE data. Its absence is striking given the prevalence of its <j> counterpart, *Mejías*.

Table 7. Frequencies of the surname *Mexías* by country. Source: https://Forebears.io

Country	Estimated frequency	Relative frequency (No. / million)	Rank in nation
Greece	123	11.1	15,938
Brazil	95	0.5	83,365
England	16	0.3	155,219
Australia	6	0.3	179,761
Portugal	1	0.1	21,562

Turning to the surname *Ximénez*, Table 8 confirms it is the least frequent spelling of its type worldwide with an estimated 4,378 bearers. When only considering Spanish speaking nations, it is the fourth most frequent, after *Jiménez*, *Giménez*, and *Jimenes*. It is most prevalent in Mexico, but most dense in Uruguay. When it comes to the <s> ending variants, although *Jimenes* is widely documented in Hispanic societies, *Ximenes* is largely concentrated in Brazil.

Looking more closely at the <x> variants, Table 9 summarizes the frequencies and rankings of *Ximénez* by country. It is mostly found in Spanish-speaking nations, but also documented in Brazil (ranked second after Mexico), France, and Portugal. After Spain, the U.S. again makes the list. In contrast, Table 10 confirms the narrow presence of *Ximenes* in Hispanic nations. In fact, its documentation in the U.S., Spain, and Argentina can likely be attributed to Portuguese speaking immigration.

Conversely, its presence in the Philippines (322 cases) probably represents an archaic Spanish variant.

Table 8. Worldwide frequencies of six variants of the surname *Jiménez*. Source: https://Forebears.io

	Estimated frequency	International rank	Most prevalent	Highest density
Jiménez	2,216,322	216	Mexico	Costa Rica
Giménez	300,036	1,818	Argentina	Paraguay
Jimenes	18,228	27,751	Mexico	Dominican Republic
Ximénez	4,378	96,510	Mexico	Uruguay
Gimenes	42,461	12,509	Brazil	Brazil
Ximenes	57,028	9,468	Brazil	East Timor

Table 9. Frequencies of the surname *Ximénez* by country. Source: https://Forebears.io

Country	Estimated frequency	Relative frequency (No. / million)	Rank in nation
Mexico	1,597	13.3	4,146
Brazil	1,065	5.2	12,691
Spain	442	9.5	8,750
United States	437	1.4	58,651
Uruguay	367	111.7	1,207
Venezuela	125	4.1	10,984
Ecuador	92	5.8	6,636
Argentina	90	2.1	50,444
France	49	0.7	155,398
Portugal	37	3.5	9,524

Table 10. Frequencies of the surname *Ximenes* by country. Source: https://Forebears.io

Country	Estimated frequency	Relative frequency (No. / million)	Rank in nation
Brazil	35,853	176.6	561
East Timor	18,251	15,151.5	5
Indonesia	873	3.5	13,942
Portugal	499	47.6	1,469
Philippines	322	3.2	34,535
United States	265	.8	87,845
France	257	3.9	38,971
Italy	246	4	35,595
Spain	98	2.1	26,993
Argentina	75	1.8	57,588

Tables 11 and 12 provide comparable data for *Roxas*. Although the head variant, *Rojas*, is as highly ranked or higher ranked than the head variants *Jiménez* and *Mejía*, it differs in that its <x> counterpart, *Roxas*, is entirely absent in Spain and Latin America. Given this, the documentation of *Roxas* in the U.S. (Table 12) is presumably due to immigration from the Philippines, where it is the 57th most frequent surname[4]. The same explanation likely applies for the rest of the countries on the list, considering migration patterns and the name's extreme frequency in the Philippines.

Table 11. Worldwide frequencies of surnames *Rojas, Roxas*. Source: https://Forebears.io

	Estimated frequency	International rank	Most prevalent	Highest density
Rojas	1,992,282	240	Mexico	Costa Rica
Roxas	95,193	5,751	Philippines	Philippines

Table 12. Frequencies of the surname *Roxas* by country. Source: https://Forebears.io

Country	Estimated frequency	Relative frequency (No. / million)	Rank in nation
Philippines	88,290	881.8	57
United States	1,942	6.1	16,982
Saudi Arabia	1,539	51.3	2,309
United Arab Emirates	826	87.4	1,289
Qatar	331	152.3	695
Canada	263	7.4	14,299
Singapore	254	47	1,626
Taiwan	232	9.9	1,368
Oman	186	46.6	2,626
Malaysia	152	5.0	11,447

5. Discussion

The socio-cultural cachet currently sustaining *Ximena* at the top of the charts in Mexico and in bordering U.S. states merits a critical examination. <x> has long enjoyed symbolic status in Mexico, but in recent decades seems to have experienced heightened awareness, as the trajectory of *Ximena* suggests, among other sociolinguistic and semiotic phenomena. In Mexico, as shown, it has been a top 100 name for many decades, but it has only occupied the highest rankings in recent years. The popularity spike of *Ximena* in Mexico and its diasporas is likely a byproduct of the revival of names of indigenous origin in Mexico since the end of the 20th century,

many of which are or can be spelled with <x> (e.g. Xochitl, Xitlali). The rise of the non-standard variant can likely also be attributed to a growing worldwide preference for innovative, unique names. This is especially the case in baby girl naming. Because names for daughters are more often chosen for style and beauty than tradition, they tend to be more plentiful, less conventional, more orthographically variable, and more vulnerable to shifting trends (e.g., Barry & Harper 2014; Lieberson 2000; Orenstein 2003; Parada 2016). In addition, one cannot ignore the powerful influences of popular culture in the promotion of particular names. Mexican celebrities like Ximena Sariñana and Ximena Navarrete undoubtedly have served to further imbue the name and its features with fashionable connotations, impacting the naming choices of parents. The timing of the trending would make sense in light of Evans' (2007) "tsunami curve" theory on the delayed impact of popular culture on naming. These socio-onomastic realities help explain the immense and rapid rise of the marked form *Ximena*, but also suggest that its extreme popularity will wane sooner rather than later.

In the U.S., viewers are presented with prime-time Latina characters portraying the name, a reality that not only reflects the name's perceived presence in Mexican-American communities, but also carries implications for non-Spanish speakers' encounters and increased familiarity with the orthographic archaism contained therein. In fact, a popular U.S. baby naming site *Oh Baby Names* recently featured the name, highlighting its vintage yet 'exotic' charm: "This is a beautiful name and the revived spelling with an 'X' makes Ximena at once distinctive and antiquated. Exotically Spanish and one of the very few X names in existence for females makes Ximena a unique choice!"

Importantly, while metalinguistic commentary on <x> in personal names abounds in baby naming circles (where Greek and other origin <x> names like *Xenia*, *Xiomara*, and *Xandra* are also garnering attention), Hispanist critics of <x> for /x/ seem to have traditionally shown less concern for the feature's retention in personal names of Spanish origin (such as those discussed in this paper) than for names of indigenous origin. The latter have been both more politicized and visible, most notably *México* and its associated demonym *mexicana/o* which were at the core of the historical debates referenced earlier (summarized in Moreno de Alba 2003). Along these lines, on the exceptional status of *México* relative to orthographic norms, Miguel de Unamuno queried: "*¿Ha de ser México más que Guadalaxara en esto? Sobre todo, igualdad ante la ley.*" ('Should Mexico be considered greater than Guadalaxara in this matter? Above all, equality under the law.') (Rosenblat 1960; Moreno de Alba 2003: 86). Unamuno's further unfavorable assessments of the practice, such as '*disparate ortográfico*' ('orthographic absurdity') and '*fetichismo de la letra*' ('letter obsession'), seem to have been singularly directed at Mexican toponymic items, with <x> variant Spanish anthroponyms entirely excluded from

such debates of orthographic 'law breaking.' Thus, unfavorable attitudes emerged not where there was convergence in use (<x> variant Spanish anthroponyms) but where norms and name origin diverged (indigenous toponymic standardization).

Perhaps this asymmetry explains why, despite unequivocal demographic evidence, the *Ximena*–Mexico link appears to remain undernoted in the popular consciousness of the Hispanic world. Spain's mainstream aversion to voluntary <x> for /x/, evident in the country's striking preference for *Jimena*, may communicate something, but the aversion is likely mostly related to regional Iberian associations with the <x> variants of Spanish names. More work, quantitative and qualitative, is clearly needed to probe how different onomastic vessels of vestigial <x> (indigenous and non-indigenous, given names and surnames, etc.) are perceived across Spanish speaking societies and cross-sectionally within said societies.

Names are language elements that are intimately tied to personal and group identities and, as earlier described, are thus often immune to top-down standardization impositions. Ownership concerns are also of special importance. Ownership factors (i.e., tradition and identity) are what have allowed <x> variant names to survive and even strongly resurge (in the case of *Ximena*). While decisions requiring and implying language ownership can be fraught in the context of pluricentric languages, names tend to be exceptional. However, the social indexicalities of <x> in the Mexican context did not emerge happenstance. They were born of a purposeful struggle for national self-determination and identity, and <x> was only able to serve as a viable alternative to <j> because of its already known opposition. The letter's prominence in the Mexican toponymic landscape had also naturally imbued it with iconicity, setting the stage for its symbolism of resistance relative to Spain and to later critics.

With respect to the surnames documented, Mexico stands out in the Spanish-speaking world as hosting the largest number of so-named individuals, particularly for *Ximénez* and *Mexía*. Could this bear any relation to the popularity of *Ximena* in that nation? To the greater societal awareness and reception of the feature? The surname *Roxas* also raises some interesting questions. Its heavy documentation in the Philippines yet absence elsewhere in the Hispanic world motivates a need for research into the historical factors for its modern confinement to that region. One factor in the retention of the <x> variant could be the influence of Mexican values and practices. The Mexico-based Viceroyalty of New Spain and the Philippines enjoyed a long period of close association, given that the islands were ruled under the former until Mexican independence, after which the source of government shifted to Spain until 1898 (Philippine independence). This shift carried linguistic consequences. Today, those later Peninsular influences are reflected in the pronunciation and spelling of some Spanish origin words (Lesho & Sippolo 2018). This underscores the roles of isolation and contact in orthographic standardization

patterns. In like fashion, the absence of the *Mexías* variant in Hispanic nations is perplexing, given the frequencies of both *Mejías* (its <s> ending counterpart) and *Mexía* (its shared <x> counterpart). Archival work is needed to investigate the historical presence of *Roxas*, if any, and why it has fared as it has. Likewise for the other surnames, archival work may be able to shed light on the factors that both favored and compromised the retention of non-standard <x>.

6. Conclusion

A growing agitation for greater individual and collective authenticity and self-determination through naming can be observed worldwide. For example, many countries have been slowly relaxing their traditionally stringent baby naming rules, affording more personal expression. Analogously, in political geography we are witnessing the restoration of pre-colonial place names and pronunciations (e.g., Mumbai; aboriginal toponymic restoration campaigns in Australia, etc.). Recently in the U.N., nations such as Cabo Verde (Cape Verde) have requested official versions of their names that match (or approximate) native forms and/or pronunciations (Johnson 2014). However, sociolinguists have long observed the stylistic limitations of speaker/speech community agency. Speakers and speech communities seek personalized, agentive expression, but are bound to a finite set of options already recognized (or recognizable) in a given society (be they letters, phonemes, morphemes, words, or grammatical constructions)[7].

The present study has shown that vestigial <x> is one such option heavily deployed in recent decades in Mexican baby naming, with its already robust local connotations and ability to convey fashionable and otherwise meaningful qualities. The nationalistic, ethnic, defiant, vintage, and other possible associations of <x> traditionally expressed through indigenous names seem to be expanding to include non-indigenous <x> names, most notably *Ximena*. While baby-naming trends, as well as toponym-heavy public signage and tourism campaigns, are revelatory as to the social significance (including folk understandings) of <x> variant given names and place names, the <x> variant surnames documented in this study may simply serve to further strengthen the feature's presence in the Mexican public consciousness. Thus, the societal presence in Mexico, though obscure, of vestigial <x> surnames could be one of the many conditioning factors in the rise of *Ximena* and other <x> variant first names.

7. An online commenter's quoting of The Most Interesting Man humoristically captures the agentive use of <x> as /x/: 'I don't always spell *México* and *Texas*, but when I do, I prefer *Dos Equis*' (Flimzy 2011).

The roles of isolation and contact were certainly at play in the historical emergence and present status of <x> as a socio-politically significant marker in Mexico. Mexico's isolation and identity building forged through its independence and later revolution permitted said development. Further, increased contact with, awareness of, and celebration of indigenous groups and cultures over time in Mexico's mainstream have strengthened the letter's associations with the prized notions of nativity and cultural authenticity. In addition, as earlier discussed, Mexico's growing contact with compelling representations of <x> in the global semioscape has likely further favored, albeit implicitly, the letter's privileged status in that nation.

Future work on this socio-onomastic topic should look to the use of diverse quantitative and qualitative methods, such as archival research, large-scale surveys, in-depth interviews, and analyses of popular commentary, for further insights into the distribution of <x> names, and related attitudes and folk perceptions.

References

Agha, Asif. 2003. The social life of a cultural value. *Language and Communication* 23. 231–273. https://doi.org/10.1016/S0271-5309(03)00012-0

Alford, Richard. 1988. *Naming and identity: A cross-cultural study of personal naming practices*. New Haven, CT: HRAF Press

Barry III, Herbert & Aylene S. Harper. 2014. Unisex names for babies born in Pennsylvania 1990–2010. *Names* 62. 13–22. https://doi.org/10.1179/0027773813Z.00000000060

Behind the Name. Ximena. Retrieved from https://www.behindthename.com/name/ximena

Candelaria, Adriana. 2013. The arms of the X. *Aquí con Adriana*, https://adrianacandelaria.wordpress.com/2013/06/18/the-arms-of-the-x/.

Carney, Edward. 1994. *A survey of English spelling*. London: Routledge.

Castillo, Ana. 1995. *Massacre of the dreamers: Essays on Xicanisima*. Albuquerque: University of New Mexico Press.

Convocatoria: México, el valor de su X. 2016, February 8. https://www.mexico.mx/es/articles/la-poderosa-x-de-mexico.

Evans, Cleveland Kent. 2007. The tsunami curve and popular culture influences on given names. *Names* 55. 335–342. https://doi.org/10.1179/nam.2007.55.4.335

Flimzy. 2011, November 16. Re: Why is the "X" in México and Texas pronounced as the letter "J"? [Blog comment], https://spanish.stackexchange.com/questions/245.

Forebears. Surnames and forenames: Meaning and statistics. https://forebears.io.

Frutiger, Adrian. 1998 [1978/1979]. *Signs and symbols: Their design and meaning*. Translated by Andrew Bluhm. London: Ebury Press.

Gale, Cathy. 2015. A practice-based evaluation of ambiguity in graphic design, embodied in the multiplicities of X. Brighton: University of Brighton dissertation.

Gamella, Juan F., Antonio Gómez Alfaro & Juan Pérez Pérez. 2014. Nominal assimilation: The ethnic and national identities of the Gitanos or Calé of Spain as shown by their surnames in the 1783–1785 census. *Names* 62. 147–164. https://doi.org/10.1179/0027773814Z.00000000083

Google Books Ngram Viewer. http://ngrams.googlelabs.com.

Instituto Nacional de España. Nombres y apellidos. https://www.ine.es/widgets/nombApell/index.shtml.
Instituto Nacional Electoral. Apellidos más frecuentes. https://www.ine.mx/transparencia/datos-abiertos/#/archivo/apellidos-mas-frecuentes.
Jaffe, Alexandra, Jannis Androutsopoulos, Mark Sebba & Sally Johnson (eds.). 2012. *Orthography as social action: Scripts, spelling, identity and power*. Berlin/New York: Walter de Gruyter. https://doi.org/10.1515/9781614511038
Jaworksi, Adam. 2019. X. *Linguistic Landscape* 5. 115–141. https://doi.org/10.1075/ll.18029.jaw
Johnson, Akilah. 2014, February 10. Cape Verde asks world to call it Cabo Verde: Wants its identity accepted. *Boston Globe*, http://bostonglobe.com.
Johnstone, Barbara, Jennifer Andrus & Andrew E. Danielson. 2006. Mobility, indexicality, and the enregisterment of "Pittsburghese." *Journal of English Linguistics* 34. 77–104. https://doi.org/10.1177/0075424206290692
Lesho, Marivec & Eeva Sippolo. 2018. Historical changes in place names in Metro Manila and Cavite, Philippines. Paper presented at the American Name Society Annual Meeting. Salt Lake City, UT.
Lieberson, Stanley. 2000. *A matter of taste: How names, fashions, and culture change*. New Haven, CT: Yale University Press.
Martinez, Marta. 2015, May. La equis. *Fronteras / Borders*, http://www.martavmartinez.com/My-Journal-Southwest_files/8d10ce4523ad3c39e29db6a7f3286d6d-9.html.
Meyerhoff, Miriam, Erik Schleef & Laurel Mackenzie. 2015. *Doing sociolinguistics: A practical guide to data collection and analysis*. London: Routledge. https://doi.org/10.4324/9781315723167
Michel, Jean-Baptiste, Yuan Kui Shen, Aviva Presser Aiden, Adrian Veres, Matthew K. Gray, The Google Books Team, Joseph P. Pickett, Dale Hoiberg, Dan Clancy, Peter Norvig, Jon Orwant, Steven Pinker, Martin A. Nowak & Erez Lieberman Aiden. 2011. Quantitative analysis of culture using millions of digitized books. *Science* 331. 176–182. https://doi.org/10.1126/science.1199644
Mollerup (User). 2015. *Popularity of name Simona [Digital image]*. https://commons.wikimedia.org/wiki/File:Popularity_of_name_Simona.svg.
Moreno de Alba, José G. 2003. *La lengua española en México*. México DF: Fondo de Cultura Económica.
Noe-Bustamante, Luis & Antonio Flores. 2019. Facts on Latinos in the U.S. Pew Resarch Center. https://www.pewresearch.org/hispanic/fact-sheet/latinos-in-the-u-s-fact-sheet/.
Oh Baby Names. *Ximena*. http://www.ohbabynames.com/meaning/name/ximena/497#.WnxRW7bMxE4.
Orenstein, Peggy. 2003. Where have all the Lisas gone? *New York Times*. http://www.nytimes.com/2003/07/06/magazine/06BABY.html?pagewanted=all.
Parada, Maryann. 2016. Ethnolinguistic and gender aspects of Latino naming in Chicago: Exploring regional variation. *Names* 64. 19–35. https://doi.org/10.1080/00277738.2016.1118858
Pelkey, Jamin. 2017. *The semiotics of X: Chiasmus, cognition, and extreme body memory*. London: Bloomsbury Academic.
Penny, Ralph. 2002. *A history of the Spanish language*. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511992827
Pharies, David A. 2015. *Breve historia de la lengua española: 2nd edn*. Chicago: University of Chicago Press.

Puzey, Guy & Laura Kostanski (eds.). 2016. *Names and naming: People, places, perceptions and power*. Buffalo, New York: Multilingual Matters. https://doi.org/10.21832/9781783094929
RAE. 1999. *Ortografía de la lengua española*. Madrid: Espasa-Calpe.
RAE. 2005. *Diccionario panhispánico de dudas*. Madrid: Espasa-Calpe.
RAE. 2010. *Ortografía de la lengua española*. Madrid: Espasa-Calpe.
Registro Nacional de Población e Identificación Personal (Secretaria de Gobernación). 2017. Annual top 1000 Mexican baby girl and boy names from 1900 to 2016. https://www.gob.mx/segob/renapo.
Rosenblat, Angel. 1960. *Buenas y malas palabras en el castellano de Venezuela*. Caracas/Madrid: Edime.
Ryan, Des. 2016. Linguists' descriptions of the English writing system. In Vivian Cook & Des Ryan (eds.), *The Routledge handbook of the English writing system*, 61–84. London: Routledge.
Sacks, David. 2003. *The alphabet*. London: Hutchinson.
Sebba, Mark. 2007. *Spelling and society: The culture and politics of orthography around the world*. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511486739
Sebba, Mark. 2012. Orthography as social action: Scripts, spelling, identity and power. In Alexandra Jaffe, Jannis Androutsopoulos, Mark Sebba & Sally Johnson (eds.), *Orthography as social action: Scripts, spelling, identity and power*, 1–19. Berlin/New York: Walter de Gruyter. https://doi.org/10.1515/9781614511038.1
Social Security Administration. Popular baby names. http://www.ssa.gov/oact/babynames/.
Thurlow, Crispin & Giorgia Aiello. 2007. National pride, global capital: A social semiotic analysis of transnational visual branding in the airline industry. *Visual Communication* 6. 305–344. https://doi.org/10.1177/1470357207081002
Tibón, Gutierre. 2001. *Diccionario etimológico comparado de los apellidos españoles, hispanoamericanos y filipinos*. México DF: Fondo de Cultura Económica.
U.S. Census Bureau. 2016. Genealogy: Frequently occurring surnames from the 2010 Census. https://www.census.gov/topics/population/genealogy/data/2010_surnames.html.
Villa, Laura. 2015. Official orthographies, spelling debates and nation-building projects after the fall of the Spanish Empire. *Written Language & Literacy* 18. 228–247.
Zhang, Sarah. 2015, October 12. The pitfalls of using Google Ngram to study language. *Wired*, https://www.wired.com.

Index of subjects

Index of varieties, languages, and language families

This interdisciplinary volume explores the unique role of the sociohistorical factors of isolation and contact in motivating change in the varieties of Spanish worldwide. Recognizing the inherent intersectionality of social and historical factors, the book's eight chapters investigate phenomena ranging from forms of address and personal(ized) infinitives to clitics and sibilant systems, extending from Majorca to Mexico, from Panamanian Congo speech to Afro-Andean vernaculars. The volume is particularly recommended for scholars interested in historical linguistics, sociolinguistics, history, sociology, and anthropology in the Spanish-speaking world. Additionally, it will serve as an indispensable guide to students, both at the undergraduate and graduate level, investigating sociohistorical advances in Spanish.

ISBN 978 90 272 0864 4

John Benjamins Publishing Company