FRANCISCO MORENO-FERNÁNDEZ

LANGUAGE EVOLUTION OVER THE LIFESPAN: EXPERIMENTAL APPROACH TO LEXICAL RICHNESS IN SPANISH

INTRODUCTION

Most of the studies on lifelong changes conducted so far have focused on phonetic (Sankoff & Blondeau, 2007, 2013; Reubold & Harrington, 2015) and grammatical (Wagner & Sankoff, 2011) levels, with less attention to pragmatic or discourse issues (Buchstaller, 2015). We therefore consider it of interest to delve more deeply into lexical aspects, which is less usual in panel studies (Gerstenberg, 2015). Likewise, the languages that have received preferential treatment within this specialty are English (Sankoff, 2018b) and French (Blondeau, 2001; Sankoff, 2018a), with partial attention to other languages such as Danish (Jensen, 2017), Portuguese (Da Conceição De Paiva, Duarte, & Guy, 2021) or Swabian German (Beaman & Tomaschek, 2021), for example. However, the accumulated experience in sociolinguistic research in the Spanish-speaking area makes Spanish an ideal language for the study of real-time linguistic change, although little work has been done on it so far. The focus of this text is not so much on the treatment and development of the concept of vocabulary or "lexical richness" as on the exploration of the limits of the study of change over the lifespan using an available panel of Spanish speakers from Madrid (Spain).[1]

Francisco Moreno-Fernández, Alexander von Humboldt Professor at Universität Heidelberg; e-mail: francisco.moreno@uni-heidelberg.de; ORCID: https://orcid.org/0000-0002-3136-4443.

THE STUDY OF VARIATION AND CHANGE IN SPANISH

The Spanish language has a series of language material of great value for the study of variation and change. Among the material, the corpus of recordings of higher education speakers from different cities of the Spanish-speaking territory stands out as pioneering and as a reference for many subsequent works. This collection of material is known as the PILEI project or "Proyecto para el estudio de la norma culta de las principales ciudades hispanohablantes", and succeeded in gathering materials from Barcelona, Bogotá, Buenos Aires, Caracas, La Habana, La Paz, Las Palmas, Lima, Madrid, Mexico, Miami, Montevideo, Panama, San José de Costa Rica, San Juan de Puerto Rico, Santiago de Chile and Seville (Lope Blanch, 1986). The material, gathered through interviews between the 1960s and 1980s, is the oldest available with potential for sociolinguistic study. In fact, much of it the basis for comparison with more recent material, making it very useful for studies of real-time change.

Likewise, Hispanic sociolinguistics has been cultivated, especially in the variationist line, since the 1970s and 1980s, when studies of the Spanish of Panama or of San Juan de Puerto Rico were undertaken. Indeed, Henrietta Cedergren's (1973) doctoral thesis on Panama City Spanish was one of the first works aligned with William Labov's sociolinguistics, and it dealt with the quantitative analysis of phonetic variables such as the aspiration of syllable-final /s/, the elision of various final consonants, or the lenition of the pre-palatal voiceless phoneme. The/Her? corpus gathered interviews from 79 Panamanian informants. The study by Humberto López Morales (1983) on San Juan de Puerto Rico was developed in a similar line and served as a methodological basis for other works, such as the study by José Antonio Samper (1990) on socio-phonetic aspects in the Spanish of Las Palmas de Gran Canaria.

As for corpus building, there have been numerous sociolinguistic experiences of great value for the knowledge of variation and change in Spanish, beyond the PILEI project. Among them, it is worth mentioning the work carried out in Venezuela, specifically in Caracas (Rosenblat & Bentgivoglio, 1979; Guirado, 2014; Galucci, 2020), and in Mexico (Rodríguez Flores & Rodríguez Alfano, 1996; Rodríguez Alfano, 2005). In the latter country, the material gathered has been used for the elaboration of studies on linguistic change, especially that derived from contact between varieties and the incidence of linguistic leaders (Martín Butragueño, 2006). Likewise, among the corpora of spoken language, the material contributed by the PRESEEA project stands out, since the project has served as a driving force for the sociolinguistic study of numerous Spanish-speaking communities in Europe and America since 1996. The objectives of PRESEEA are oriented, in its theoretical and methodological approaches, to the comparative sociolinguistics of numerous Spanish-speaking communities (Moreno Fernández, 2016, 2021a, 2021b).

The PRESEEA project proposes to work on a common basis for all Spanish-speaking communities, which guarantees the comparability of the material. The PRESEEA samples for each community are organized by quotas with uniform distribution, according to the variables of gender, age and educational level. The reference sizes for each sample range from 54 to 108 informants, depending on the social complexity of the urban nucleus. If the number of communities investigated with the same PRESEEA methodology reaches 50, as expected, the total volume of speakers whose discourse will form part of this macrocorpus will exceed 2500 (Moreno Fernández & Cestero, 2020).

Since the PRESEEA project started in 1996 and the collection of material by several research teams was completed before 2000, two decades later, work has begun on the collection of new samples. This is being done for the communities of Mexico, Malaga and Madrid, as was also done in Caracas and Monterrey. In this way, the continuity in the construction of corpora of spoken language is making it possible to gather sociolinguistic material from campaigns carried out at different times and to be in a position to undertake studies of change in real time, both panel and trend studies. As is well known, trend studies analyze the same community at two successive points in time, but without using the same informants on each occasion; panel studies do use the same informants.

In relation to the study of linguistic change in Spanish, it is quite clear that sociolinguistic projects carried out in numerous communities have made it possible to analyze processes of change in apparent time. However, there have also been some approaches from trend studies. An example of this is the work by Antonio Briz (2023) on the Spanish town of Requena, a community in the interior of the peninsula dedicated mainly to vine cultivation and winemaking. The changes observed by Briz (2023) point to three hypotheses concerning the evolution of dialectal speeches: there is a tendency 1) to level with the standard, 2) to delocalize some features, and 3) to maintain some features as a mark or reinforcement of social identities.

Another example of a trend study is that carried out by Hernández-Campoy (2003) on the Spanish of the region of Murcia (Spain) and based on radio material broadcast between the 1970s and 2000. In this study, it was found that the use of standard features of Spanish is gradually and consistently spreading throughout the region of Murcia and among the different social classes and groups, to the detriment of local, southern linguistic features. From a geolinguistic point of view, this slow but steady erosion of local features follows a hierarchical structure of diffusion, from the largest to the smallest urban centers. For the Spanish language, I know of no precedents of panel works oriented to the study of changes over the life span.

LANGUAGE EVOLUTION AND CHANGE: THEORETICAL PREMISES

Our study is aligned with analyses of real-time linguistic change and specifically with studies of change over the lifespan — or over the life course (Hunt, 2005; Lowry, 2022) — from panel material. The epistemological framework from which we work, then, assumes for the most part the proposals made from other life course change research (Labov, 1978, 1994, 2001; Blondeau, 2006; Sankoff & Wagner, 2006; Tagliamonte, 2012; Wagner, 2012; Wagner & Buchstaller, 2017; Sankoff, 2018a, 2018b, 2019; Beaman & Buchstaller, 2021).

From a sociological perspective, we start from the assumption that the "life course" refers to events, transitions, and trajectories in a person's life — in this case, referring to their linguistic practices — that unfold in particular sequences and dynamics across life stages (Lowry, 2022, p. 35). Likewise, this paper accepts the cardinal principles of the life course perspective, namely:

Principle 1. A person's life course is implicated in and shaped by the historical moments he or she lives through.

Principle 2. The effect of a major biographical incident will be shaped by the age and stage of life at which the event occurs.

Principle 3. No life story can be understood independently of other life stories.

Principle 4. Conditions, events, and experiences at any age or life stage cannot be understood independently of conditions, events, and experiences at earlier ages or life stages.

Principle 5. Individuals make choices and perform actions within the limits of their social universe.

On the other hand, it is important to distinguish between life as a biological phenomenon and as a sociocultural phenomenon, there being a strong interrelation between both approaches. Biological conditions include cognitive functions and their neurophysiological correlates (Kemper, 2015). The sociocultural dimension of age is linked to external and social factors, such as education or work. Moreover, it should be considered that the aging process as a social reality is strongly correlated with its perception (Rowe & Kahn, 1997) so, to a large extent, it is a socially constructed process.

From a sociolinguistic perspective, this paper accepts the fundamental conceptual distinctions established for the study of linguistic change across the life span. Thus, a distinction is made between individual change and community change, and the existence of significant intersections between them is recognized. It is also accepted that lifelong change trends may respond to different patterns: continuity or stability, age categorization, regression or even life change (Labov, 1994, p. 83; Sankoff, 2019).

From a linguistic perspective, it should be noted that language features may have different behavior in their development or evolution, depending on their nature. Broadly speaking, the evolutionary process is given the name "linguistic change", but, within such a process, a distinction can be made between modifications of a quantitative nature and of a qualitative nature. Quantitative changes involve an increase or decrease in the frequency of a given condition or in the use or occurrence of a given feature, in which case one could speak of "evolution". Qualitative changes involve the modification, alteration, or substitution of one feature or characteristic for another, as documented in the processes of phonetic elision or assimilation, in cases of grammaticalization or in the adoption of borrowings or neologisms.

In the sociolinguistic analysis of lexical richness, the linguistic factor under study does not imply the modification or mutation of a given feature, nor does it imply the substitution of one element for another. Rather, it is an evolution of a quantitative nature that manifests itself individually and would tend to adopt an asymptotic form (Ávila Muñoz, 2014, p. 175). From this point on, two research questions could be proposed to guide the analysis of change: on the one hand, it is worth asking whether, despite the individual basis in the evolution of lexical richness, communitarian change is possible; on the other hand, we wonder to what extent the social factors of gender, age and educational level of the speakers condition the evolution of lexical richness.

LEXICAL RICHNESS AND ITS EVOLUTION

Our analysis operates on lexical usage data from speakers experiencing "normal aging"; that is, aging unaffected by noticeable cognitive or neurophysiological impairments (Thornton & Light, 2006; Abrams & Farrell, 2011). In psychology, vocabulary size has traditionally been considered a significant variable, which has been linked to the state and development of human intelligence (Wechsler, 1941; Singer et al., 2003, p. 319). However, while numerous papers have been published on the effects of aging on the basis of vocabulary tests, less is known about lexical development in spontaneous speech and oral production over the life course.

About lexical evolution between generations, there are quantitative studies that show how some words cease to be used in favor of innovative forms (Beeching,

2011). Other previous studies show that we should not expect a decrease in vocabulary volume over the life course, but rather that older adults know more words than younger adults (Ramscar et al., 2013). This will be one of the hypotheses we pose for verification based on a panel study of the lexical richness of Madrid speakers. The clearest antecedent to our work is the sociolinguistic study conducted by Annette Gerstenberg (2015), who analyzed a corpus of spoken French in two sets of interviews, conducted in 2005 and 2012, with 28 participants.

Lexical richness within a given discourse or text can be measured in several ways. One of them consists of dividing the number of different types of words (types) appearing in a discourse or a text by the total number of words (tokens) in the discourse or text under analysis. Thus, to measure lexical richness, the proportion of different types within the set of words in a text is calculated. This proportion is called Type-Tokens Ratio (TTR) (Baker, Hardie, & McEnery, 2006) and also "lexical density" (Ávila, 1988; Ávila Muñoz, 2014):

TTR = Type / Tokens

Indeed, the TTR is calculated from only two values: the different words (types) and the total number of words (tokens), without considering the frequency of each of them. The suitability of the TTR has been discussed, due to its sensitivity to the length of the discourses or texts: the TTR is not a constant measure but decreases as the length of the texts increases (Stubbs, 2001; Baker, McEnery, & Hardie, 2006; Rojo, 2021). I am aware of the many limitations of this calculation; however, the analysis we present uses the TTR justifying it with two arguments: the first is that it is a well-known and accepted measure in lexical research; the second is that the discourses analyzed do not show great variation or diversity in terms of length, since they all come from a source with identical or very similar characteristics. If we consider lexical richness as a complex concept defined by the triad density, diversity, sophistication, this paper focuses exclusively on lexical diversity (Malvern et al., 2004).

As regards other possibilities for analyzing and quantifying the lexicon of a discourse, we can mention the calculation of lexical diversity, using Yule's K, the calculation of lexical peculiarity, using the hapax/tokens ratio (Baayen, 2001) or even the calculation of informational entropy, using the classical procedures of Shannon and Weaver (1963) based on the frequency of types. In contrast to what was practiced in Gerstenberg's (2015) study, in our study we will only consider the calculation of the TTR, pending future approaches with a larger volume of data.

The approach to the lexicon presented here pays attention to the individual character of the uses of spoken language in conversation but is also interested in its social dimension. In this way, we intend to discover the individual and generational

tendencies (Salthouse, 2009, p. 513) observed in speakers of the city of Madrid at two different moments in their lives, so that a possible temporal development can be detected, in correlation with the sociolinguistic variables of sex, age and educational level.

As far as these social factors are concerned, studies have been carried out that offer very varied conclusions. Thus, for example, Wechsler (1941) observed that the influence of subjects' educational and cultural opportunities on lexical tests is rather weak, and that vocabulary seems to be more stable in senescence, although it may be affected by cognitive impairments in old age, such as loss of working memory, inhibition deficits or general slowing (Abrams & Farrell, 2011). Similarly, Ramscar et al. (2013) found that possible cognitive impairment contrasts with other patterns of lexical growth and learning across the lifespan, while other analyses indicate that linguistic competencies of lexical-semantic knowledge are relatively stable in healthy aging (O'Hanlon, Wilcox, & Kemper, 2005).

Specifically sociolinguistic studies that have been interested in the lexicon have shown that it can be conditioned or affected by social factors (Sankoff & Lessard, 1975; Tainturier, Tremblay, & Lecours, 1992; Ávila Muñoz, 2014), but in different ways. Sankoff and Lessard (1975), for example, found by analyzing 120 French informants in Montreal that socioeconomic status or residential environment have no direct effect on lexical diversity, whereas educational level does. They also observed a continuous enrichment of productive vocabulary with increasing age, at least up to the age of 50, a fact explained by the greater malleability of the lexicon, as opposed to that of phonology or syntax. This malleability makes possible a prolonged acquisition over time, especially among more educated speakers (Buchstaller, 2015).

METHODOLOGICAL QUESTIONS

The analysis of lexical richness in Spanish presented here is of an experimental nature and has been carried out on the transcriptions of semi-guided conversations held in the Spanish city of Madrid according to the methodological guidelines of the PRESEEA project, which involve the investigation of samples of informants selected according to the variables of gender (male = H / female = M), age (1 = 20–34 years old / 2 = 35–54 years old / 3 = 55 years old or older) and educational level (1 = primary education / 2 = secondary education / 3 = higher education) (Moreno Fernández, 2021a).

The semi-guided conversations analyzed for this study were held in two different time periods: 2001 and 2022. It is, therefore, a panel study, for which the same informants were interviewed in both periods. Because the number of informants

with whom it has been possible to repeat interviews in Madrid is still scarce, the interviews analyzed are relatively few compared to those used in other studies. Here we analyze the interviews held with six Madrid informants at the two indicated times (12 interviews in total).[2] The sociolinguistic profiles of the speakers interviewed are shown in the table. The identifying codes for each informant indicate, in this order, gender; age and educational level; thus, H11 refers to a 20–35-year-old male with primary education.

Table 1
Madrid Panel, With Indication of Speakers' Sociolinguistic Profile, Name and Year of Recording

PANEL DE MADRID
H12 ROBERTO 2001 / H22 ROBERTO 2022
H13 DANIEL 2001 / H23 DANIEL 2022
H22 JOSE 2001 / H32 JOSE 2022
M12 BLANCO 2001 / M22 BLANCO 2022
M13 PAULA 2001 / M23 PAULA 2022
M22 ANGELES 2001 / M32 ANGELES 2022

Note. Gender: M = Male, F = Female. Age (years): 1 = 20–35, 2 = 36–55, 3 = 56+. Education: 1 = primary, 2 = secondary, 3 = higher.

The semi-guided interviews lasted between 45 and 60 minutes and were organized around thematic modules that were reiterated in all of them, although with some flexibility. These interviews were transcribed by the team responsible for the sociolinguistic research in Madrid using as a reference the system of minimum tags and labels of the PRESEEA project (Moreno Fernández, 2021b). This system uses the XML coding language, an adapted and simplified version of SGML, to facilitate data exchange and selective information retrieval. The tagged texts are in TXT format and are organized internally into two parts: a header and the text itself. The header consists of a series of fields that provide data on the file itself, the interview recording, the interview transcription and review, and on the speakers participating in the interview.

For the analysis of lexical richness, the texts received a treatment that allowed them to be handled with different tools. This treatment was based on the TXT format files derived from the PRESEEA project and consisted of the following tasks: elimination of headings; elimination of formal elements typical of the interview

---

[2] All PRESEEA Madrid panel interviews available to date have been used for this analysis.

(e.g., turn-taking indications); elimination of XML tags and labels; elimination of paralinguistic elements, cut words and hesitations typical of the conversation; and elimination of proper names. This treatment of cleaning codes and elements foreign to the object of study results in a TXT text exported to UTF-8 code and which meets the formal conditions to be subjected to lexicostatistical analysis.

The analysis of the 12 conversations and specifically the analysis of lexical richness was carried out with the help of the software called T-Lab (Lancia, 2020). This software consists of a set of linguistic, statistical, and graphical tools designed for the analysis of texts, whatever their type (newspaper articles, transcripts of interviews and speeches, answers to open questions, Twitter messages, business documents, legislative texts, books, etc.). The program performs quantitative analyses that allow us to extract patterns based on relationships between words and significant topics. The treatment that the texts receive are the following: corpus normalization, detection of word sequences and empty words, segmentation into elementary contexts, automatic lemmatization and keyword selection. The analyses performed are the following: 1) co-occurrence analysis, which allows us to explore, measure and map relationships between keywords; 2) thematic analysis, oriented to the search for keywords within the texts and their cluster analysis; and 3) comparative analysis, where differences and similarities between different types of texts are analyzed and mapped.

The analysis offered here is of an exploratory nature, given the size and composition of the corpus, and is focused on the calculation of lexical richness through the TTR. Experimentally, during the research process, other analytical possibilities were tested but finally discarded, such as the calculations that take into account the hapax (forms that appear only once in each text), as well as the ratio of hapax and types or tokens. In the case of the hapax/types calculation, the result usually maintains a more constant ratio in texts of different length or magnitude (Rojo, 2021). However, in this study we have disregarded this calculation because of the large amount of hapax legomena shown in the texts, not because of the appearance of peculiar words with high informative load, but because of the presence of altered or modified forms (including errors), very frequent in spoken conversations and interviews. The quantitative results provided by the lexicometry tool were in turn subjected to various statistical tests, mainly descriptive, for which three computer resources were used: for general statistics, SPSS and XLSTAT; for regression analysis, GOLDVARB Z, an updated version of the methodological tool used for variationist sociolinguistic analysis (Tagliamonte, 2021).[3]

---

[3] Given the small size of the dataset, the applied statistics is very simple so it was not worth using other resources such as the *quanteda* package for R.

LEXICOMETRIC ANALYSIS

Based on the 12 interviews conducted in 2001 and 2022 from six informants in Madrid, a first obligatory step for the analysis of lexical richness is the quantitative description of the texts, including the length or length of the texts, their number of words. In our texts, it is observed that their length is fairly even. This is because the 2001 and 2022 interviews were conducted following the same methodological guidelines. Each of the units or words that make up a text is called a "token", so that the total number of tokens in a text is its total number of words; that is, its total length including all possible multiple occurrences of each word.

Another measure of quantification is that of word types. A type is considered to be each different word, disregarding its possible repetitions. Thus, the number of types within a text will be the number of different words it contains, while the total number of words will be expressed as the number of tokens. Table 2 shows the number of types and words in each interview.

Table 2
Absolute Frequencies of Types and Tokens in the Six Informants Interviewed in Madrid in 2001 and in 2022, With Indication of Speakers' Sociolinguistic Identification Code

Types	H12 Roberto	H13 Daniel	H22 Jose	M12 Trinidad	M22 Angeles	M13 Paula
2001	1572	1742	1733	1686	1747	1469
2022	1484	1616	1549	1904	1805	1934
Tokens						
2001	9635	11306	10114	13958	9895	8867
2022	7366	8093	6046	12061	9237	11215

Figure 1 shows the closeness of the interviews in terms of the number of types and vocabulary used by the informants. This quantitative factor is relevant because the analysis of lexical richness is sensitive to significant differences in the length of the speeches or texts analyzed. The averages of word types and total number of words for all informants are as follows:

– types: 2001: 1658, 2022: 1715;
– tokens: 2001: 10629, 2022: 9063.

Figure 1

Bar Charts of Number of Types and Number of Tokens per Interview in 2001 (Series 1) and in 2022 (Series 2)
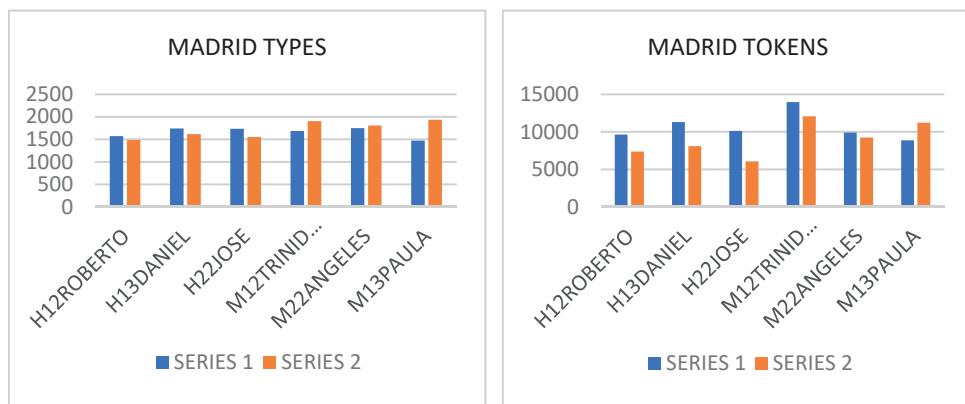
Another interesting measure for the quantitative characterization of the analyzed interviews is the number of words with content; that is, those types of words that have lexical and not grammatical meaning. Excluded from this count, therefore, are articles, pronouns, prepositions, conjunctions and other elements without lexical meaning. In this case, we do not observe relevant differences between 2001 (Series 1) and 2022 (Series 2), except in the case of a woman with higher education: PAULA M13 > M23.

Figure 2

Number of Types With Content in Interviews With Six Madrid Speakers in 2001 (Series 1) and 2022 (Series 2)
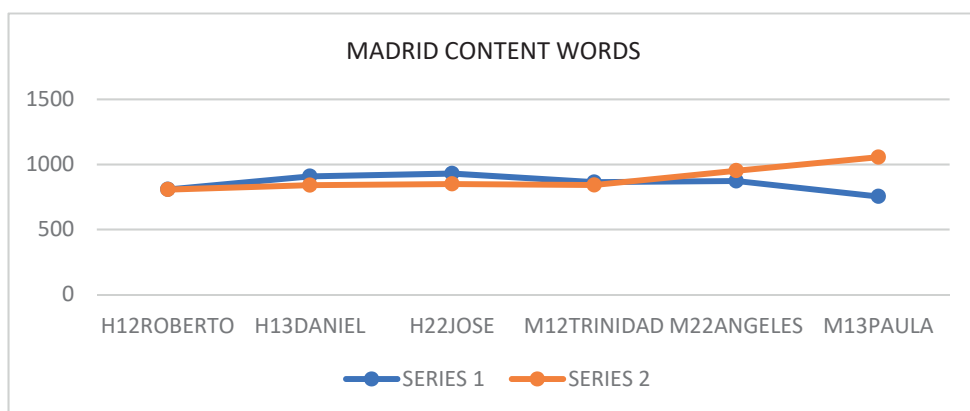
	H12 Roberto	H13 Daniel	H22 Jose	M12 Trinidad	M22 Angeles	M13 Paula
2000	808	909	930	864	873	754
2022	807	841	850	842	952	1056

The calculation of the lexical richness of the Madrid speakers analyzed was carried out by applying the Types/Tokens ratio (TTR). The results of the TTR calculation are shown in Table 3 and its visual representation in Figure 3.
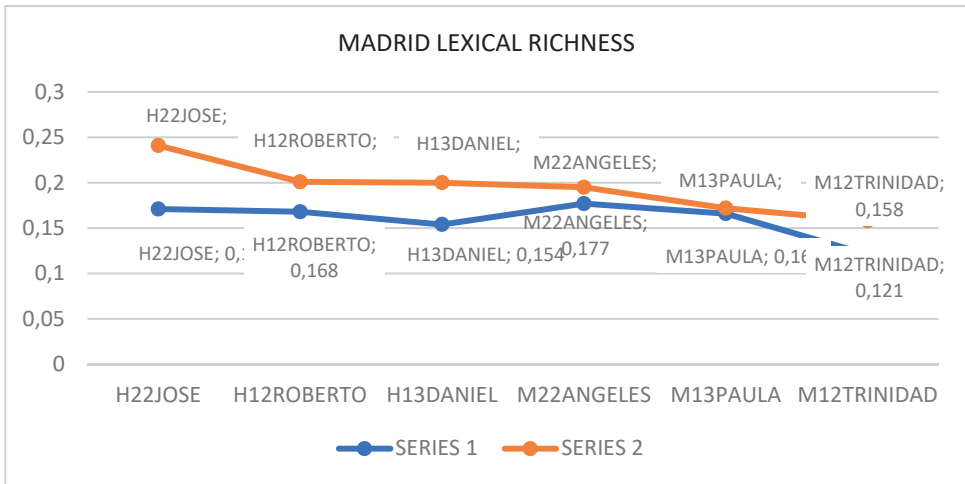
Table 3

TTR Calculation Results by Speaker and Time Period

	H22 Jose	H12 Roberto	H13 Daniel	M22 Angeles	M13 Paula	M12 Trinidad
2001	.171	.168	.154	.177	.166	.121
2022	.242	.201	.200	.195	.172	.158

Figure 3

TTR Calculation Results by Speaker and Time Period

According to these results in all cases, the interviews conducted in 2022 show a higher index of lexical richness than the interviews conducted in 2001. Only the fact that the informant M13, who had shown a greater difference in the frequencies of words with content, offers a smaller distance between the lexical richness of 2001 and 2022 deserves a comment.

To test the strength of the relationship between the two-time series analyzed, we applied two statistical tests. On the one hand, a hypothesis was tested using Friedman's test for analysis of variance; on the other hand, we calculated Cohen's *d*, which measures the effect size or the difference between two variables, using the standard deviation of the samples. In the application of Friedman's test, we started from the null hypothesis that the distributions found in our 6 Madrid speakers in 2001 and 2020 are equal. In this case, the significance yielded by the test is 0.173, which is higher than 0.05, the usual reference for Social Sciences. This means that the null hypothesis cannot be rejected and that it must be accepted that the distributions are equal. As for Cohen's *d*, its values are less than 0.20, so it should be understood that the difference between the samples is small.

Table 4
Cohen's d *in Lexical Richness of Madrid Speakers in Samples Collected in 2001 and 2020*

M22 ANGELES	.012728
H13 DANIEL	.032527
M13 PAULA	.004243
M12 TRINIDAD	.026163
H12 ROBERTO	.023335
H22 JOSE	.050205

In the process of the analysis, other statistical tests have been applied experimentally, whose significance indices have generally exceeded the 95% confidence interval, so their inclusion in this work has been discarded. Undoubtedly, the small size of the samples has been a factor that has hindered us from reaching conclusions of greater significance.

Having verified the growth of vocabulary richness in all the Madrid speakers, the hypothesis that this increase is correlated with the sociolinguistic factors of sex/gender, age and educational level has to be verified. To test this hypothesis, we adapted our data for logistic regression analysis. The adaptation consisted basically in creating a dependent variable related to the increase in lexical richness (increase/no increase) and in omitting the variant "generation 1", since it is not possible to appreciate any change in lexical richness with respect to the previous time reference.

The regression analysis, performed using the GOLDVARB Z tool, shows age as the only variable with explanatory power; that is, with a significance below the 0.05 threshold. For the variables sex/gender and educational level, the significance index exceeded this level.

Table 5

Result of Regression Analysis of the Increase in Lexical Richness

Input 0.381
Sex/gender = H: 0.680, M: 0.320. Log likelihood = –12.291. Significance = 0.519
Age = 2: 0.195, 3: 0.805. Log likelihood = –9.595. Significance = 0.017
Educational level = 3: 0.441, 2: 0.623, 1: 0.434. Log likelihood = –12.304
Significance = 0.820
Fit: χ-square = 7.437, accepted, p = 0.2845

Note. Gender: M = Male, F = Female. Age (years): 2 = 36–55, 3 = 56+. Education: 1 = primary, 2 = secondary, 3 = higher.

Most probably, even though it is a variable phenomenon, the low overall significance of the explanatory variables may be related to the scarcity of the data, as well as to their distribution. Accepting the limitations of statistical significance, it could be affirmed that in the sex/gender and educational level variables there is a tendency towards an increase in lexical richness in men and in speakers with an intermediate education, although in the latter case the differences with respect to the other variants are smaller.

DISCUSSION

This work was based on two fundamental research questions related to lexical richness in spoken dialogic language. The first one referred to the possible existence of an increase in lexical richness throughout life, tested on the basis of spoken language in an interview situation in two periods of time. The second concerned the incidence of sociolinguistic factors in the possible increase of lexical richness.

In relation to the first question, data from 6 informants from the Spanish city of Madrid, in interviews conducted in 2001 and in 2022, indicate that in all cases an increase in lexical richness is observed, measured through the TTR index. This result is in line with that of many other studies on the evolution of the lexicon in healthy speakers over the course of their lives. Also, individual observations allow us to state that such a process can be interpreted on an individual basis, since the observed increases are variable; indeed, previous studies have pointed out that individual behavior may be unstable over time (Sankoff, 2005).

From a sociolinguistic perspective, the significance of the factor "age" is redundant with the general increase in lexical richness, so it would be the factors sex/gender and educational level that would merit specific discussion. In general terms, we can accept the hypothesis that biographical and environmental factors, expressed

through educational level and socio-professional class, could be correlated with lexical richness indicators. This was tested by Sankoff and Lessard (1975) for Montreal French. Such biographical factors, however, have not been found to be significant at a community level among Madrid informants.

According to these observations, lexical richness evolves inversely to the standard language use explained by Allan Bell (2014) (U-shaped shift), according to which younger and older people tend to use non-standard forms, while middle-aged people, tend to a greater use of the standard. In the case of the lexicon, there would be a process of continuous enrichment, in the form of an asymptote, which would only decline due to cognitive deterioration, if it were to occur. Thus, in the later stages of life, a double possible trajectory can be found: 1) the level of lexical richness can be maintained or even increased, as observed; 2) the level of lexical richness declines.

Figure 4
Scheme of Evolution of Lexical Richness, With Continuous Increase and Possible Fall Due to Cognitive Impairment

Middle-aged

Younger   Older
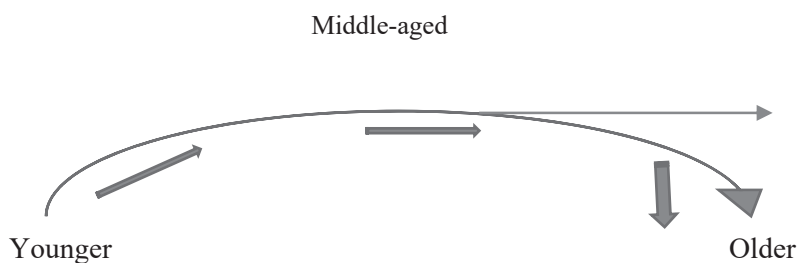
If we relate this analysis to the proposals of theoretical models referring to the processes of change throughout life, it could be concluded that the observed increase in lexical richness could be interpreted as age-grading. This age-graded evolution necessarily implies individual changes throughout life, but it does not imply a change of the language in the community. That is, the increase in lexical richness over the course of a speaker's lifetime does not mean that the whole community increases its lexical richness. If anything, such an increase could be interpreted as a generational evolution towards the dominant standard (Bell, 2014), which does not necessarily mean that the whole community experiences such an evolution.

Thus, the individual increase in lexical richness responds to individual evolutions not linked to community processes of variation and change but correlated with age and the succession of life stages, according to the general dynamics of each community. In a society there may be cyclical factors with the capacity to condition individual trajectories, such as schooling at different ages, the usual forms of family planning

or training and professional progress (Buchstaller, 2015). Among the conditioning factors experienced by individuals, the social salience of certain phenomena with repercussions on lexical uses can also be important, such as the weight of ideologies in certain areas and moments of life or the social and indexicalized perception of prestige or stigma.

CONCLUSION

This paper has presented a panel study of the evolution of lexical richness over the lifespan in which, for the first time, material from the Spanish language is used. It is a partial and experimental analysis, since it works only with 6 informants from the Spanish community of Madrid, with whom the same type of semi-guided interview was conducted in 2001 and 2022. The material is part of the PRESEEA project, which, in communities such as Madrid, Malaga or Mexico City, are being collected in new campaigns, two decades after the previous sample collection. The analysis carried out is specifically interested in the evolution of lexical richness in conversation, as well as in the factors that may condition it.

The methodology used presents some peculiarities that must be taken into account for the interpretation of lexical richness. As mentioned above, one of them is the relative scarcity of material used for its analysis, but the particular nature of the material and texts on which the analysis is carried out should also be mentioned. This is material from interviews in which the presence of conversational elements such as repetitions, hesitations or deictics is significant. Likewise, the presence of hapax in this type of discourse is more intense than in other discourses, due to the fact that errors or specific original forms may occur. All this must be disregarded in the analysis of lexical richness.

In the interviews conducted in Madrid in 2001 and 2022 with speakers of different social profiles (men and women, three age groups and three educational levels), lexical richness increased appreciably in the speech of all the individuals investigated. This evolution of lexical richness over time tends to become more frequent in men than in women and in people of intermediate educational level, although the analysis did not show significant differences for these two social factors.

The evolution of lexical richness could be included among the processes of age-grading, which can manifest themselves in an unstable manner and which respond to an evolutionary development that does not necessarily have to be linked to changes at a community level. In this evolution of lexical richness, a continuous increase can be observed in the absence of cognitive deterioration in the speakers. This increase can be justified by the malleability of the lexicon, as well as by its

relative stability, derived from the capacity of healthy memory to accumulate lexical information as life experiences accumulate. In general, the observations made on the evolution of lexical richness corroborate the principles established by Deborah Lowry in her conceptual delimitation of the life course.

REFERENCES

Abrams, L., & Farrell, M. T. (2011). Language processing in normal aging. In J. Guendouzi, F. Locke, & M. J. Williams (Eds.), *The handbook of psycholinguistic and cognitive processes*: *Perspectives in communication disorders* (pp. 49–73). Psychology Press.

Ávila Muñoz, A. M. (2014). Patrones sociolingüísticos de la riqueza léxica. Estudio basado en una propuesta original para el cálculo del índice de la densidad léxica virtual de los hablantes. *Lingüística Española Actual*, *36*, 171–194.

Ávila, R. (1988). Lengua hablada y estrato social. Un acercamiento lexicoestadístico. *Nueva Revista de Filología Hispánica*, *36*(1), 131–148.

Baayen, H. (2001). *Word frequency distributions*. Kluwer Academic Publishers.

Baker, P., Hardie, A., & McEnery, T. (2006). *A glossary of corpus linguistics.* Edinburgh University Press.

Beaman, K., & Tomaschek, F. (2021). Loss of historical phonetic contrast across the lifespan: Articulatory, lexical, and social effects on sound change in Swabian. In K. Beaman and I. Buchstaller (Eds.), *Language variation and language change across the lifespan: Theoretical and empirical perspectives from panel studies* (pp. 209–234). Routledge.

Beaman, K., & Buchstaller, I. (Eds.). (2021). *Language variation and language change across the lifespan: Theoretical and empirical perspectives from panel studies*. Routledge.

Beeching, K. (2011). The sociolinguistics of lexical variation in Standard French: A diachronic perspective. In T. Pooley, D. Lagorgette, & R. A. Lodge (Eds.), *On linguistic change in French*: *Socio-historical approaches. Studies in honour of R. Anthony Lodge* (37–53). Université de Savoie.

Bell, A. (2014). *The guidebook to sociolinguistics*. Wiley.

Bentivoglio, P., & Malaver, I. (2012). Corpus sociolingüístico de Caracas: PRESEEA Caracas 2004–2010. Hablantes de instrucción superior. *Boletín de Lingüística*, *24*(37–38), 144–180.

Blondeau, H. (2001). Corpora comparability and changes in real time within the paradigm of the personal pronouns in Montreal French. *Journal of Sociolinguistics*, *5*(4), 453–474.

Blondeau, H. (2006). Panel studies and language variation. In K. Brown (Ed.), *Encyclopedia of language and linguistics* (pp. 150–152). Elsevier.

Briz, A. (2023). Tres hipótesis explicativas del cambio dialectal: el caso del habla de la comarca de Requena-Utiel. *Homenaje a José Antonio Samper que le brinda la Academia Canaria de la Lengua* (in print). Academia Canaria de la Lengua.

Buchstaller, I. (2015). Exploring linguistic malleability across the life span: Age-specific patterns in quotative use. *Language in Society*, *44*, 457–496.

Cedergren, H. (1973). The interplay of social and linguistic factors in Panama [Doctoral dissertation]. Cornell University.

Da Conceição De Paiva, M., L., Duarte, M. E., & Guy, G. R. (2021). Alignment of individuals with community trends: Subjects from the Portuguese. In K. Beaman & I. Buchstaller (Eds.), *Language variation and language change across the lifespan: Theoretical and empirical perspectives from panel studies* (pp. 56–76). Routledge.

Gallucci, M. J. (2020). Paola Bentivoglio (1932–2019) y la lingüística en Venezuela. *Lingüística*, *36*(1), 173–184.

Gerstenberg, A. (2015). A sociolinguistic perspective on vocabulary richness in a seven-year comparison of older adults. In A. Gerstenberg & A. Voeste (Eds.), *Language development: The lifespan perspective* (pp. 109–128). John Benjamins.

Guirado, K. (2014). Corpus diacrónico del habla de Caracas 1987/2013. *Boletín de Lingüística*, *24*(41–42), 19–44.

Hernández Campoy, J. M. (2003). Complementary approaches to the diffusion of standard features in a local community. In D. Britain & J. Cheshire (Eds.), *Social dialectology: In honour of Peter Trudgill* (pp. 23–37). John Benjamins.

Hunt, S. (2005). *The life course. A sociological introduction*. Palgrave.

Jensen, T. J. (2017). Generic 'du' in time and context: A study of intra-individual variation and change. In D. Duncker & B. Perregaard (Eds.), *Creativity and continuity: Perspectives on the dynamics of language conventionalization* (pp. 209–234). Copenhagen University Press.

Kemper, S. (2015). Language production in late life. In A. Gerstenberg & A. Voeste (Eds.), *Language development: The lifespan perspective* (pp. 59–75). John Benjamins.

Labov, W. (1978). On the use of the present to explain the past. In P. Baldi & R. Werth (Eds.), *Readings in historical phonology* (pp. 275–312). Pennsylvania State University Press.

Labov, W. (1994). *Principles of linguistic change: Vol. 1. Internal factors*. Blackwell.

Labov, W. (2001). *Principles of linguistic change: Vol. 2. Social factors*. Blackwell.

Lancia, F. (2020). T-LAB Plus 2020 – User's Manual. *T-LAB*. Retrieved December 6, 2022, from https://www.tlab.it/bibliography

Lope Blanch, J. M. (1986). *El estudio del español hablado culto. Historia de un provecto*. Universidad Nacional Autónoma de México.

López Morales, H. (1983). *Estratificación social del español en San Juan de Puerto Rico*. Universidad Nacional Autónoma de México.

Lowry, D. (2022). *Aging and life course. Social and cultural contexts*. Rowman & Littlefield.

Malvern, D., Richards, B., Chipere, N., & Durán, P. (2004). *Lexical diversity and language development: Quantification and assessment*. Palgrave Macmillan.

Martín Butragueño, P. (2012). Variación y cambio lingüístico en el español mexicano. *Español Actual*, *98*, 11–38.

Martín Butragueño, P. (2006). *Líderes lingüísticos: estudios de variación y cambio*. El Colegio de México.

McEnery, T., & Hardie, A. (2012). *Corpus linguistics.* Cambridge University Press.

Moreno Fernández, F., & Cestero Mancera, A. M. (2020). El proyecto PRESEEA: desarrollos analíticos. In A. J. Gallego & F. Roca Urgell (Eds.), *Dialectología digital del español* (pp. 119–138) (Verba Anexo 80). Servizo de Publicacións e Intercambio Científico Campus Vida.

Moreno Fernández, F. (Ed.). (2021a). *Metodología del "Proyecto para el estudio sociolingüístico del español de España y de América. (PRESEEA)"*. Universidad de Alcalá (Documentos PRESEEA de trabajo).

Moreno Fernández, F. (Ed.). (2021b). *Marcas y etiquetas mínimas obligatorias para materiales de PRESEEA*. Universidad de Alcalá (Documentos PRESEEA de trabajo).

Moreno Fernández, F. (2016). En torno a PRESEEA: notas de investigación y sociología de la ciencia. *Boletín de Filología*, *51*(2), pp. 369–376.

O'Hanlon, L., Wilcox, K. A., & Kemper, S. (2005). Age differences in implicit and explicit (associative memory: Exploring elaborative processing effects. *Experimental Aging Research*, *27*(4), 341–359.

Ramscar, M., Hendrix, P., Love, B. C., & Baayen, H. (2013). Learning is not decline. The mental lexicon as a window into cognition across the lifespan. *The Mental Lexicon*, *8*(3), 450–481.

Reubold, U., & Harrington, J. (2015). Disassociating the effects of age from phonetic change: A longitudinal study of formant frequencies. In A. Gerstenberg & A. Voeste (Eds.), *Language development: The lifespan perspective* (pp. 9–38). John Benjamins.

Rodríguez Alfano, L. (Coord.). (2005). *Investigación sociolingüística. El habla de Monterrey. Su trayectoria en una página electrónica*. Universidad Autónoma de Nuevo León.

Rodríguez Flores, D. E., & Rodríguez Alfano, L. (1996). El habla de Monterrey: Historia de un proyecto. In D. E. Rodríguez Flores & L. Rodríguez Alfano (Eds.), *Lenguaje y sociedad. Metodologías y análisis aplicados al habla de Monterrey* (pp. 11–34). Trillas.

Rojo, G. (2021). *Introducción a la lingüística de corpus en español.* Routledge.

Rosenblat, Á., & Bentivoglio, P. (Eds.). (1979). *El habla culta de Caracas. Materiales para su estudio*. Facultad de Humanidades y Educación, Universidad Central de Venezuela.

Rowe, J. W., & Kahn, R. L. (1997). Successful aging. *The Gerontologist*, *37*(4), 433–440.

Salthouse, T. A. (2009). When does age-related cognitive decline begin? *Neurobiology of Aging*, *30*, 507–514.

Samper, J. A. (1990). *Estudio sociolingüístico del español de Las Palmas de Gran Canaria*. La Caja de Canarias.

Sankoff, D., & Lessard, R. (1975). Vocabulary richness: A sociolinguistic analysis. *Science, New Series*, *190*(4215), 689–690.

Sankoff, G., & Blondeau, H. (2013). Instability of the [r] ~ [R] alternation in Montreal French: An exploration of stylistic conditioning of a sound change in progress. In L. Spreafico & A. Vietti (Eds.), *Rhotics: New data and perspectives* (pp. 249–265). Free University of Bozen-Bolzano Press.

Sankoff, G., & Blondeau, H. (2007). Language change across the lifespan: /r/ in Montreal French. *Language*, *83*(3), 560–588.

Sankoff, G., & Wagner, S. (2006). Age grading in retrograde movement: The inflected future in Montreal French. In M. L. Friesner & M. Ravindranath (Eds.), *University of Pennsylvania Working Papers in Linguistics: Papers from NWAV 34* (pp. 203–216). University of Pennsylvania.

Sankoff, G. (2005). Cross-sectional and longitudinal studies in sociolinguistics. In P. Trudgill (Ed.), *Sociolinguistics: An international handbook of the science of language and society* (pp. 1003–1013). De Gruyter.

Sankoff, G. (2018a). Before there were corpora: The evolution of the Montreal French Project as a longitudinal study. In S. Wagner & I. Buchstaller (Eds.), *Panel Studies of Variation and Change* (pp. 21–52). Routledge.

Sankoff, G. (2018b). Language change across the life-span. *Annual Review of Linguistics*, *4*, 297–316.

Sankoff, G. (2019). Language change across the life-span: Three trajectory types. *Language*, *95*(2), 197–229.

Shannon, C. E., & Weaver, W. (1963). *The mathematical theory of communication.* University of Illinois Press.

Singer, T., Verhaeghen, P., Ghisletta, P., Lindenberger, U., & Baltes, P. B. (2003). The fate of cognition in very old age: Six-year longitudinal findings in the Berlin Aging Study (BASE). *Psychology and Aging*, *18*, 318–331.

Stubbs, M. (2001/2006). *Words and phrases. Corpus studies of lexical semantics.* Blackwell Publishing.

Tagliamonte, S. (2021). *Goldvarb Z*. http://individual.utoronto.ca/tagliamonte/goldvarb.html.

Tagliamonte, S. (2012). *Variationist sociolinguistics: Change, observation, interpretation*. Wiley-Blackwell.

Tainturier, M. J., Tremblay, M., & Lecours A. R. (1992). Educational level and the word frequency effect: A lexical decision investigation. *Brain and Language*, *43*(3), 460–474.

Thornton, R., & Light, L. L. (2006). Language comprehension and production in normal aging. In J. E. Birren & K. W. Schaire (Eds.), *Handbook of the psychology of aging* (6th ed., pp. 261–287). Elsevier.

Wagner, S., & Sankoff, G. (2011). Age grading in the Montréal French inflected future. *Language Variation and Change*, *23*, 1–39.

Wagner, S., & Buchstaller, I. (Eds.). (2017). *Panel studies of variation and change*. Routledge.

Wagner, S. (2012). Real-time evidence for age grad(ing) in late adolescence. *Language Variation and Change*, *24*, 179–202.

Wechsler, D. (1941). *The measurement of adult intelligence* (2nd ed.). Williams & Wilkins.

Yule, G. U. (1944). *Statistical study of literary vocabulary*. Cambridge University Press.

LANGUAGE EVOLUTION OVER THE LIFESPAN:
EXPERIMENTAL APPROACH TO LEXICAL RICHNESS IN SPANISH

S u m m a r y

This paper aims to contribute to the knowledge of linguistic change over the lifespan by means of panel studies. The analysis offered presents two elements that have received little treatment within the field. On the one hand, the analysis is interested in the evolution or development of the lexicon in dialogical speech; on the other hand, the analysis is carried out on material from a panel study on the Spanish language. This text analyzes the evolution of lexical richness by means of an exploratory panel study based on semi-guided interviews carried out in Madrid, a Spanish city included in the framework of the international PRESEEA project.

Keywords: language change; panel studies; lifespan; vocabulary richness; Spanish; Madrid; PRESEEA.

EWOLUCJA JĘZYKA NA PRZESTRZENI ŻYCIA.
EKSPERYMENTALNE PODEJŚCIE DO BOGACTWA LEKSYKALNEGO W JĘZYKU HISZPAŃSKIM

S t r e s z c z e n i e

Niniejszy artykuł ma na celu poszerzenie wiedzy na temat zmian językowych zachodzących w ciągu całego życia jego użytkowników przy użyciu badań panelowych. Prezentowana analiza podejmuje dwa aspekty, które dotychczas odgrywały marginalną rolę jako przedmiot badań językoznawczych. Prezentowana analiza stanowi wgląd w zagadnienie ewolucji czy też rozwoju zasobów leksykalnych mówców w interakcji dialogicznej, a przeprowadzono ją na materiale z badań panelowych dotyczących języka hiszpańskiego. Niniejszy artykuł analizuje ewolucję bogactwa leksykalnego mówców za pomocą badania zawierającego częściowo ukierunkowane wywiady przeprowadzone w Madrycie, hiszpańskim mieście objętym międzynarodowym projektem PRESEEA.

Słowa kluczowe: zmiana językowa; badania panelowe; długość życia; bogactwo słownictwa; hiszpański; Madryt; PRESEEA.

Francisco Moreno-Fernández holds a PhD in Hispanic Linguistics from the Universidad Complutense (Madrid) and a degree in Political Science and Sociology. He is Alexander von Humboldt Professor at the University of Heidelberg, Professor of Ibero-American Linguistic, Cultural and Social Studies, and Director of the Center for Ibero-American Studies at the same university. A specialist in dialectology, sociolinguistics, sociology of language and linguistic methodology; among his most recent publications are *A Framework for Cognitive Sociolinguistics* (2017), *La lengua española en su geografía* (5th ed., 2020), *Variedades de la lengua española* (2020), *La lengua y el sueño de la identidad* (2020), *La lengua de los hispanos unidos de América* (2022) and, with Rocío Caravedo, *The Routledge Handbook of Spanish Dialectology* (2022).