Language Variation and Change (2025), **37**, 31–57
doi:10.1017/S0954394525000055

Pathways of actualization across regional varieties and the real-time dynamics of syntactic change

Matti Marttinen Larsson[1,2,3]

[1]Department of Languages and Literatures, University of Gothenburg, Gothenburg, Sweden; [2]Department of Linguistics, Stockholm University, Stockholm, Sweden and [3]Institut für Romanistik, Humboldt-Universität zu Berlin, Berlin, Germany
Email: matti.marttinen.larsson@gu.se

Abstract

This cross-dialectal diachronic study focuses on actualization—the gradual mapping out of structural innovations across linguistic environments—and asks: to what extent can the route by which actualization unfolds be explained by appealing to common cognitive constraints, and to what extent are pathways of actualization shaped by community-specific patterns in language use? I analyze real-time syntactic change in Spanish oblique relative clauses (ORCs), which are undergoing an enhancive change towards increasing use of antecedent-agreeing definite articles (e.g., *la casa en [la] que nací* 'the house in which I was born'). Over 8,000 occurrences of ORCs were compiled from three Latin American varieties. A Bayesian mixed-effects logistic regression analysis shows that various constraints representing the COGNITIVE ACCESSIBILITY OF THE ANTECEDENT influence actualization similarly across varieties. The discussion addresses the cognitive realism of historical corpus data and formulates testable predictions that present profitable lines of inquiry in future cognitively oriented cross-dialectal diachronic research.

Keywords: usage-based linguistics; cognitive linguistics; corpus linguistics; diachronic data; dialectal variation

Introduction

A central goal of theories of language variation and change is accounting for how language change comes about and spreads across linguistic contexts (cf. De Smet, 2012b; Timberlake, 1977; Weinreich, Labov, & Herzog, 1968). There is overwhelming consensus that language change unfolds gradually across linguistic environments (De Smet, 2012a; Timberlake, 1977). The gradual process whereby the consequences of a prior reanalysis are overtly manifested in the grammar of a language is referred to as *actualization* (Andersen, 2001; Harris & Campbell, 1995; Hopper & Traugott, 2003 [1993]; Timberlake, 1977). Actualization thus involves the gradual mapping out of a new underlying structure across linguistic contexts (Timberlake, 1977:141), and can be perceived as a "special manifestation" (De Smet, 2016:84) of Weinreich et al.'s (1968:183) *constraints problem*: what are the possible conditions for change that determine the progression of change across diverse linguistic environments?

An enduring concern of variationist studies has long been uncovering this route by which grammatical innovations incrementally spread in the linguistic system (e.g., Aaron, 2010; Dion, 2023; Naro, 1981; Poplack, Lealess, & Dion, 2013; Poplack & Malvar, 2007; Torres Cacoullos & Walker, 2009), which I will henceforth refer to as the course of actualization (De Smet, 2012b; see also De Smet, 2009). A pressing issue for a theory of actualization—beyond ascertaining the route by which grammar change unfolds—lies in establishing why changes take place in one particular order instead of another (De Smet, 2012b; Dietrich, 2024). Over the past decade, there has been growing recognition that the process of actualization is largely driven by similarity-based generalizations (De Smet, 2012b): change is actualized firstly in contexts that in some (syntactic, formal, probabilistic, etc.) aspect resemble already established uses and where it can infiltrate the grammar inconspicuously. From there, innovations spread across linguistic environments through successive stepwise similarity-based generalizations (De Smet, 2012b; Naro, 1981; see also Andersen, 2001; Timberlake, 1977). This means that the more similarity in form or distribution there is between two contexts or structures, or the stronger the constructional network ties between two structures, the more probable it is that an innovation will be extended.

Importantly though, such generalizations and extensions are also sensitive to frequency. Abstract patterns that are frequent are more likely to extend their domains of application and to serve as analogical models for generalization (Bybee, 2010:7; De Smet, 2012a:57); in addition, high type frequency sparks productivity and the extension of a construction to involve new items (Bybee, 2010:67). Taken together, these observations mean that extensions largely hinge on frequency, since "as an expression becomes more frequent in one grammatical context, its mental retrievability improves, which in turn makes it more easily available in different yet closely related (analogous) grammatical contexts" (De Smet, 2016:83). For this reason, frequent constructions have a "gravitational pull" (De Smet, 2012a:72) and can easily recruit new members to their environment. Granted that the supporting constructions guiding extensions are specific for each item and each language, De Smet (2012b:608-609) predicted that the course of actualization is item-specific as well as language-specific. In other words, variable phenomena's courses of actualization are highly contingent because the possible generalizations differ between structures and languages and their networks of supporting constructions (De Smet & Fischer, 2017:243).

However, a point that remains less clear pertains to cross-varietal stability during actualization of change: if the same instance of change is attested *across different varieties of the same language*, to what extent can we expect the course of actualization to be non-contingent on regional varieties? This question is intriguing, because the cognitive organization of one's language along with the linguistic and cognitive routines of a language user can differ from one individual to another and, on an aggregate level, between communities and social groups (Barking, Backus, & Mos, 2022; Bresnan & Ford, 2010:204; Röthlisberger, Grafmiller, & Szmrecsanyi, 2017; Szmrecsanyi, Grafmiller, Heller, & Röthlisberger, 2016; Verhagen, Mos, Backus, & Schilperoord, 2018; see also Schneider, 2023). That is, patterns of language use and

https://doi.org/10.1017/S0954394525000055 Published online by Cambridge University Press

processing are largely socio-cognitively determined (cf. Schmid, 2020). This means that actualization pathways, being a reflex of cognitive mechanisms and processes, could potentially vary from one community to another (cf. Bresnan & Ford, 2010:205).

From this, various questions with considerable theoretical bearing surface: when grammatical innovations spread across different environments in the linguistic system, do the cognitive underpinnings of this process—namely, actualization—bring about parallel pathways of change in different populations of speakers? What are "universal" cognitive constraints and to what extent is the course of actualization shaped by community-specific patterns in language use?

The present study aims to contribute to our understanding of these key issues. Tackling these questions requires a comparative cross-dialectal approach that examines how change unfolds across linguistic contexts in different varieties. Work along this line of research has already been conducted, most prominently within the area of variationism and, specifically, making use of the comparative variationist method (Poplack & Tagliamonte, 2001). The comparative sociolinguistic approach compares the synchronic degree of generalization of a particular (innovative) linguistic variant across linguistic contexts and communities to infer the progression of language change in different dialects. Even if varieties might differ in rates of change, comparative variationist studies generally presume that when several varieties exhibit the same instance of variation between structures, these dialects are heading towards the same target on a common cline of grammaticalization, with some varieties being more advanced than others (Tagliamonte, 2013:186; Tagliamonte, Durham, & Smith, 2014:80). However, from a diachronic and fundamentally cognitive perspective on the actualization of change, the comparative variationist enterprise presents some theoretical complexities that merit attention. The first challenge involves the socio-cognitive side of language change. If we assume that actualization is largely based on similarity-based generalizations and extensions between networks of supporting constructions (De Smet, 2012b), we would also need to consider that the point of departure underlying such extensions are analogical models; these models, in turn, are sensitive to frequency (cf. Bybee, 2006), in the sense that frequent and conventionalized structures and usage patterns tend to serve as models for generalization (Bybee, 2010:63) because they have a "gravitational pull" that attracts other structures (De Smet, 2012a:72). However, what is important to keep in mind is that the input structures that are most frequent in one speech community may differ from the most frequent structures in another community, which means that the substructure steering the course of actualization may, in fact, be variable across varieties of a language (cf. Grafmiller, Szmrecsanyi, Röthlisberger, & Heller, 2018:2; Verhagen et al., 2018). Thus, in adopting a synchronic comparative perspective, the dynamics of change and its dependency on language use in a specific community are largely left unaddressed (as are the idiosyncrasies of the individual and the respective speech communities in terms of the composition of their probabilistic grammars [Barking et al., 2022; Szmrecsanyi et al., 2016; Verhagen et al., 2018]).

Secondly, comparative sociolinguistic studies are generally limited to a particular synchronic snapshot of different varieties of language. Needless to say, given the synchrony of the data, the incremental stages of change that have taken place in the respective varieties can only be *inferred* rather than ascertained. While this is only natural considering the nature of the data, the general (implicit) idea underlying

https://doi.org/10.1017/S0954394525000055 Published online by Cambridge University Press

the comparative (variationist) approach seems to be that grammaticalization is unidirectional, because when this approach perceives that grammaticalization appears to be parallel between varieties of a language, synchronic states of variation in different varieties are thought to reflect different stages of diachronic change on the same cline of grammaticalization.

Taken together, what all of this shows is that to understand the extent to which the course of actualization is, or is not, contingent on a particular speech community or variety, or what the underlying motivations for convergence or divergence could potentially be, what is needed is an approach that simultaneously tests the effect of independent (socio-cognitive and structural) factors, linguistic varieties, and their evolution in real time—all in one single model.

However, in addition to data sparsity issues (after all, diachronic corpus data does not generally come in abundance), such an approach can also turn out to be complicated for statistical reasons: constructing a statistical model (e.g., a mixed-effects logistic regression model) that includes three-way interactions between independent effects, regional varieties, and real time tends to generate overcomplicated models with (quasi-)separation, convergence errors, "large *p* small *n*" issues, and unreliable estimates, among other undesirable consequences.

The present study proposes to tackle these issues using Bayesian mixed-effects regression modelling, a technique that can more easily handle (quasi-)separation, low or zero variance, and model identifiability issues where frequentist methods (e.g., generalized linear mixed-effects models using, for instance, the *lme4* package in *R*) would otherwise run into convergence problems (Kimball, Shantz, Eager, & Roy, 2019; see also Grafmiller, 2023:8; Levshina, 2022b). By outlining a multivariate, diachronic, real-time approach to the comparative analysis of the actualization of language change, the present study addresses the issue of cross-varietal actualization, modelling three-way interactions between factors relating to cognition and usage, regional varieties, and real time (for a similar approach, see Grafmiller, 2023; Wolk, Bresnan, Rosenbach, & Szmrecsanyi, 2013). Adopting a usage-based perspective on language change (Aaron, 2010; Bybee, 2006, 2007), it is hypothesized that, to the extent that domain-general cognitive mechanisms and processes along with usage factors steer the course of actualization, they should operate similarly across varieties of a language (cf. Grafmiller et al., 2018:3). This, in turn, would mean that actualization should largely align cross-varietally in terms of the contexts that are affected by change and the order in which extensions take place.

To test this prediction, this study analyzes an instance of ongoing syntactic change in Spanish, namely the conventionalization of definite articles in oblique relative clauses (ORCs) (*la casa en que nací* versus *la casa en la que nací*, 'the house in which I was born'). The use of a definite article in the relative has become increasingly conventionalized across varieties of Spanish over the last centuries. In this study, more than 8,000 occurrences of Spanish ORCs from Argentinean, Peruvian, and Colombian Spanish are analyzed using Bayesian mixed-effects logistic regression. A set of factors all relating to cognitive and usage-constrained effects are coded and their effects are measured as a function of regional variety and real time. The analysis confirms the hypothesis by showing that, across the analyzed varieties, there are virtually no significant differences in the effects of the analyzed factors, suggesting that the considered cognitive and usage

https://doi.org/10.1017/S0954394525000055 Published online by Cambridge University Press

factors constrain and guide actualization in a highly similar manner. These findings provide robust empirical support for a usage-based view on actualization with a strong cognitive commitment, and furthermore advance our understanding and statistical modelling of the cross-varietal actuation of change.

The remainder of the paper is structured as follows. The next section outlines the linguistic variable under scrutiny. Subsequently, the hypothesized cognitive and usage-conditioned constraints and their operationalization are discussed. This section is followed by a methodological section, describing the methodological approach of the paper and explaining the Bayesian cross-varietal diachronic variationist approach. Following this, the results are presented, while the last section advances a general discussion along with some concluding remarks.

The variable use of the definite article in Spanish ORCs: background

The linguistic variable that this paper deals with is the variable use of the definite article in Spanish oblique relative clauses (henceforth ORCs). These alternatives are illustrated in the examples below, with the definite article absent in (1) but present in (2). The noun phrase (NP) antecedent is in bold and the ORC is underlined.

(1) *La Gaceta de Buenos Aires publicaba después* ***una carta*** *de Cullen a Rosas* ***en que*** *habían indicios claros […].*

‘The Gaceta de Buenos Aires subsequently published a letter from Cullen to Rosas in which there were clear indications…’

(Domingo Faustino Sarmiento, Facundo. Civilización y barbarie, 1845–1874, Argentina, CORDE)

(2) *Mockus lanzó la idea de que se instaurara* ***una cátedra práctica*** *en los colegios del Distrito* ***en la que*** *los profesores se sentaran a tomar trago con los alumnos y luego analizaran los efectos que les produjo el alcohol en su comportamiento.*

‘Mockus proposed that a practical course should be imparted in the schools of the districts in which the professors would sit down to consume alcohol with the pupils, and after that they would analyze the effects that the alcohol had on their demeanor.’

(El tiempo, 1997-04-07, “Propuestas que no le cuajaron,” Colombia, CREA)

Since the 18th century, the definite article in ORCs (2) has been conventionalizing in different varieties of Spanish (Blas Arroyo, 2021; Blas Arroyo & Vellón Lahoz, 2017, 2018; Girón Alconchel, 2004, 2006; Guzmán Riverón, 2012). In the 18th century, the change was still highly incipient (cf. Nevalainen & Raumolin-Brunberg, 2017:54-55) in Latin American varieties of Spanish. Tellingly, Guzmán Riverón’s (2012:201) data indicates that the innovative variant only constituted about 7% of the author’s data from the first half of the 18th century. In contrast, in European Spanish, it was more advanced (Blas Arroyo, 2021:499; Blas Arroyo & Vellón Lahoz, 2017:495, 2018:16).

In the 21st century, the [PREPOSITION + DEFINITE ARTICLE + *que*] variant is highly conventionalized (cf. Schmid, 2020:87-88) in European Spanish (Vellón Lahoz & Moya Isach, 2017:471) while its conventionalization is progressing at a much slower rate in Latin American varieties (Santana Marrero, 2004).

In what follows, the constraints that are expected to condition the variation between the two variants are detailed.

Cognitive and usage-conditioned constraints on variant selection and change: on the role and operationalization of accessibility

This study focuses on the influence of effects that are, in the scope of a cognitive theory of language, directly linked to cognitive and usage-determined effects on variable coding options. Concretely, I will evaluate a recent proposal advanced by Levshina (2022a) who formulated the *Principle of Negative Correlation between Accessibility and Costs*, according to which there is a tendency "to use shorter forms to express more predictable, expected, typical etc. meanings, and longer forms to express less predictable, expected, typical, etc. meanings" (Levshina, 2022a:24; see also Jaeger, 2010). Starting from this principle, Levshina predicted that "language users should spend less effort and time on highly accessible information, and more effort and time on less accessible information" (Levshina, 2022a:22). The implication of this principle would be that the more accessible the antecedent of an ORC is, the more likely it is to favor the shorter variant of the ORC (without the definite article). Conversely, less accessible antecedents trigger the longer variant of the ORC (with the definite article). This is because the usefulness of the definite article increases when the tie between the antecedent NP and the ORC is weaker. Through the use of anaphora (e.g., *en* ***la*** *que*), the speaker aids the addressee in retrieving the antecedent.

In light of this, it seems that the definite article in the ORCs could potentially have emerged as an accessibility marker. Its role is (initially) to signal to the interlocutor how easily the antecedent can be retrieved, and it is a measure of the processing cost involved in retrieving the antecedent (Ariel, 1990:16). In what follows, constraints that condition the accessibility of a referent will be reviewed and linked to the linguistic variable under study.

One key accessibility constraint is the DISTANCE between an antecedent and a referring expression (Arnold, 2010). Larger distances and lower degrees of unity between an antecedent and a referring expression lead to lower degrees of accessibility of the antecedent (Ariel, 1990:28-29; Arnold, 2010; Clark & Sengul, 1979; Hawkins, 1999). As for the variable article use in Spanish ORCs, an effect of the distance between the antecedent and the ORC has been observed both diachronically and synchronically (Blas Arroyo, 2021; Blas Arroyo & Vellón Lahoz, 2017, 2018; Guzmán Riverón, 2012:182, 198; López García, 1994:440-442; Vellón Lahoz & Moya Isach, 2017:477, 479; among others). Here, DISTANCE was operationalized as the number of words (cf. Tagliamonte & Baayen, 2012). Since anaphoric reference reactivates its referent in the working memory (McKoon & Ratcliff, 1980), it seems plausible to assume that overtly manifested syntactic properties such as gender or number agreement in other structures than the antecedent lexeme itself (e.g., through resumptive pronouns or antecedent-agreeing dislocated adjectives) should plausibly lead to an activation of the

https://doi.org/10.1017/S0954394525000055 Published online by Cambridge University Press

head noun. Therefore, the distance was calculated as the number of words between the antecedent *or* the latest anaphoric expression to the antecedent.

Another factor considered in this study is the RESTRICTIVENESS of the relative clause. According to López García (1994:440-442), the definite article in the ORC anchors the relative pronoun to the antecedent NP, which is particularly useful in nonrestrictive relatives, being more detached than restrictive ones (see also Blas Arroyo, 2021:495). It is therefore hypothesized that nonrestrictive relatives are more inaccessible vis-à-vis restrictive ones and should, therefore, favor the definite article variant.

DEFINITENESS also influences accessibility: definite antecedents are generally more accessible than indefinite antecedents (Ariel, 1996:22). Therefore, accessibility marking is more expected when the antecedent NP is indefinite, since retrieval is cognitively more costly when the antecedent is underspecified. This aligns with earlier findings on the linguistic variable of interest here, since the innovative article variant is mostly used with indefinite antecedents (Blas Arroyo, 2021:500; Blas Arroyo & Vellón Lahoz, 2017, 2018:41; Girón Alconchel, 2006:1530; Vellón Lahoz & Moya Isach, 2017:476). The definiteness of the antecedent (definite, indefinite, or zero-marked) was established based on its last mention, which most frequently was the antecedent lexeme itself (e.g., *una casa* 'a house') but which could also include, for instance, a definite demonstrative pronoun (e.g., *aquella* 'the one/that [one]') referencing the previous mention of an antecedent (definite or indefinite).

Similarly, the GRAMMATICAL NUMBER is known to influence cross-linguistic patterns of grammatical coding: more frequent functions (i.e., singular) tend to have zero or shorter markers than less frequent ones (e.g., plural [Du Bois, 1985:363; Greenberg, 1966:32]). Singular NPs are inherently more accessible than plural because singular tends to refer to more concretely delimited entities, whereas plural is often underspecified and indefinite (see also Jaeger & Wasow, 2005). So far, however, this effect has not been found to influence the variable article use in Spanish ORCs significantly (Blas Arroyo & Vellón Lahoz, 2017, 2018; Vellón Lahoz & Moya Isach, 2017).

The concreteness of the noun also affects accessibility of an antecedent. The so-called *concreteness effect* refers to the faster and easier cognitive processing of concrete nouns, whereas abstract nouns are cognitively costlier to process in different types of tasks (see Jessen, Heun, Erb, Granath, Klose, Papassotiropoulos, & Grodd, 2000). Antecedent concreteness has nonetheless appeared to be non-significant in quantitative studies on the linguistic variable under study here (Blas Arroyo & Vellón Lahoz, 2017, 2018; however, see Girón Alconchel, 2006). Ideally, the degree of concreteness of a particular word is determined on the basis of some external measurement, such as concreteness scores (e.g., Guasch, Ferré, & Fraga, 2016). However, since Guasch et al.'s (2016) concreteness scores do not match the lexical items analyzed here, an alternative approach to approximate the concreteness effect is opted for. Psycholinguistic research suggests that WORD LENGTH is correlated with the abstractness/concreteness of the lexical item (Lewis & Frank, 2016; Reilly, Hung, & Westbury, 2017; Reilly, Westbury, Kean, & Peelle, 2012; see also Lievers, Bolognesi, & Winter, 2021). It is therefore hypothesized that longer words could potentially be perceived as more abstract and, therefore, render higher probabilities of the definite article variant, and vice versa.

Yet another effect known to influence accessibility is the SYNTACTIC FUNCTION of a discourse entity and its thematic prominence. Topical and subject referents are

https://doi.org/10.1017/S0954394525000055 Published online by Cambridge University Press

generally more accessible and, hence, more likely to be expressed using underspecified referential expressions (see Arnold, 2010:190-192). As concerns the ORCs under study here, the potential influence of topicality or subjecthood has not been systematically analyzed (though see Blas Arroyo & Vellón Lahoz, 2017, 2018; Girón Alconchel, 2006:1527). Considering the large quantity of data coded in the present study, an automatic coding process was used to code this variable. Given that subject and topic referents can bring about similar effects in terms of accessibility (Arnold, 2010), and considering that an automatic process was opted for, the analysis does not distinguish between syntactic prominence and topicality, and instead focuses solely on syntactic function. Using the Spanish transformer pipeline of *spaCy* (Explosion, 2023), the data was automatically parsed for dependency relations. In this study, subjecthood and non-subjecthood were contrasted. The dependency relations of the taxonomy that were considered to most directly reflect subjecthood were NSUBJ and ROOT, which were contrasted with the other dependency relations. It should be noted, however, that the corpora only provide a limited Keyword in Context (KWIC), which potentially reduces the accuracy of the classification. To evaluate the classification performance, a random sample of 400 occurrences were hand-coded, indicating that the *spaCy* pipeline achieved an accuracy of 89%, with an F1 score of 0.73 and a Matthews Correlation Coefficient of 0.68. These evaluation metrics suggest that while the classification is generally satisfactory, there remains some room for improvement. The results should be interpreted with these limitations in mind.

Lastly, accessibility is also associated with frequency. Higher-frequency structures are more predictable (and accessible) than lower-frequency counterparts because of their frequent use (Ariel, 1990:22; Bybee, 2007:243; Haspelmath, 2021:624). The use of accessibility markers should therefore also be correlated with frequency and, hence, used most often when the antecedent NP consists of hapax legomena or other lower-frequency nouns, whereas higher-frequency antecedents are inherently more accessible due to their overall cognitive salience (see also Marttinen Larsson, 2024). Neurolinguistic research aligns well with this prediction, demonstrating that during antecedent retrieval, higher-frequency words are more easily accessed and assigned an antecedent role, whereas lower-frequency words are harder to retrieve (Heine, Tamm, Hofmann, Hutzler, & Jacobs, 2006). This study approaches the effect of the antecedent's frequency by measuring ATTRACTION, that is, the constructional predictability of a specific lexeme in a construction, which can also be described as the information content of an antecedent lexeme given a construction (here, an ORC), calculated as the number of occurrences of a given lexeme divided by the total frequency of ORCs in the relevant macroregional dataset (Levshina, 2018:7). Lower values of ATTRACTION reflect antecedents that are frequency-wise unpredictable (e.g., hapax legomena and other lower-frequency antecedent lexemes, such as *una tempestad* 'a storm'), whereas higher values indicate that the antecedent is recurrently documented (e.g., *el año* 'the year,' *la carta* 'the letter,' etc.) as ORC antecedents in the corpus. The antecedent lexeme was annotated in its basic form. Proper and place names were coded as [Name] and time references (dates, exact times, etc.) as [Time_reference].

Having outlined the hypothesized effects, we can predict (in line with Levshina, 2022a:24) that the longer variant [PREPOSITION + DEFINITE ARTICLE + *que*] should firstly be used in contexts involving inaccessible antecedents (i.e., distant, indefinite,

https://doi.org/10.1017/S0954394525000055 Published online by Cambridge University Press

plural, abstract, non-subject/non-topic, and low-frequency antecedents). Over time, actualization should progress as a function of how (in)accessible antecedents are, with actualization moving from contexts involving inaccessibility towards those of increasing accessibility. More importantly for the purposes of the present study, if the actualization of the change is indeed a reflex of fundamentally cognitive processes, we expect diachronic macro-level tendencies to align cross-varietally. The next section describes the methodology and statistical approach.

Methodology and statistical modelling

The consulted corpora are the *Corpus diacrónico del español* (CORDE, data up to 1974) and the *Corpus de Referencia del Español Actual* (CREA, data between 1975 and 2000). While ORCs may be introduced by a variety of prepositions, this study limits itself to *en* ('in' [cf. Blas Arroyo, 2021; Blas Arroyo & Vellón Lahoz, 2017, 2018; Vellón Lahoz & Moya Isach, 2017]).

Data is only compiled from Latin American varieties (in contrast to Blas Arroyo, 2021; Blas Arroyo & Vellón Lahoz, 2017, 2018; Girón Alconchel, 2004, 2006; Vellón Lahoz & Moya Isach, 2017, among others, who all focus on European Spanish). The analyzed time period was limited to mid-19th century up to 2000. This temporal scope is warranted for two reasons: firstly, because it constitutes the era following Latin America's independence from Spain (approximately mid-19th century and onwards), a critical period for processes of variation and change in Spanish (Caravedo, 2019:26); secondly, because the progression of the innovative article variant is still highly incipient during the 18th century (Guzmán Riverón, 2012).

The corpus searches included *en que* and its variants with definite articles (*el*/*la*/*los*/*las*). The data was subsequently binned into three rough time periods (∼1850–1899, 1900–1949, 1950–2000) to gain an overview of the diachronic distribution of the data, revealing more well-populated distributions in the most modern parts of the dataset, and much scarcer ones during earlier time periods. Therefore, drawing a random sample of the entire dataset would yield a diachronically skewed sample. A diachronically stratified random sample of the data was drawn to circumvent this skewedness (see further below). Note that the data was only for sampling purposes; in the analysis itself, TIME is included as a numerical predictor.

The three regional varieties that were most well-represented across each time period were selected for analysis: Argentinian, Peruvian, and Colombian Spanish. All data from the first two time periods were extracted for subsequent annotation. For the most recent time period, a random sample of $n = 1700$ occurrences was extracted for each regional variety. Data cleaning involved excluding the lexicalized expression *la medida* ***en que*** 'the extent to which,' which does not exhibit variation (following Santana Marrero, 2004:68). The antecedent did, at times, fall outside of the left context of the KWIC. In these cases, larger excerpts were manually extracted and added to the dataset. Despite these precautions, there were, nonetheless, a few occurrences that were N/A coded and excluded due to insufficient contextual information.

Bayesian mixed-effects logistic regression was conducted using the *brm* function in the *brms* package (version 2.21.0, Bürkner, 2021) in *R* (version 4.4.1,

https://doi.org/10.1017/S0954394525000055 Published online by Cambridge University Press

R Core Team, 2022). Additional descriptive statistics of the dataset are in the Appendix. The script and data used for the analysis can be accessed through OSF: https://doi.org/10.17605/OSF.IO/FVZKU. The descriptive statistics can be found in Tables A1–A15 in the Appendix. For general information on Bayesian statistics, I refer the reader to Kruschke (2015). Some basic remarks are in order, though. Bayesian models do not yield *p* values, but instead provide posterior distributions from which parameter estimates and credible intervals (CrI) are computed. A CrI indicates that there is a 95% probability that the true parameter value lies within this interval. To test the significance of an effect, an equivalence test is carried out (Region of Practical Equivalence, ROPE). ROPE represents the range of values considered practically equivalent to the null hypothesis. If the highest density interval lies outside the ROPE, it can be taken as evidence for a significant effect (Makowski, Ben-Shachar, & Lüdecke, 2019). Moreover, Bayesian models include priors which reflect prior beliefs about the parameters. Here, weakly informative priors with a Cauchy distribution were included, centered at 0 with a scale parameter of 10 for the intercept and 2.5 for all other model parameters. These priors provide regularization while avoiding strong prior beliefs from being imposed on the posterior distribution (following Gelman, Jakulin, Grazia Pittau, & Su, 2008).

The dependent variable contained two levels: zero (*en que*) or definite article variant (*en el que, en la que, en los que, en las que*).

The independent variables included: ATTRACTION (log), DISTANCE (log), WORD LENGTH (log), DEFINITENESS, GRAMMATICAL NUMBER, and YEAR (centered). The reference level of each categorical factor is set to the most frequent level. Random effects included ANTECEDENT LEXEME and AUTHOR (based on the corpus metadata).[1] The final datasets retained for subsequent analysis consisted of $n = 3553$ for Argentina; $n = 2611$ for Peru; and $n = 2382$ for Colombia.

As mentioned in the introductory section, the cross-dialectal diachronic analysis consists of constructing a regression model consisting of three-way interactions between the respective predictors, regional variety, and real time. If no statistically meaningful differences surface between historical regional varieties and the predictors, this would suggest the existence of cross-varietally comparable trajectories of actualization. It is important to note that, while empirically robust, this approach does not conclusively prove the absence of differences. Rather, if no meaningful differences emerge between the regional varieties in their trajectories of change, this would suggest that there is sufficient evidence to support stability in actualization trajectories across dialects.

In accordance with the hierarchy principle, all three-way interaction terms also included lower-level interactions and main effects. Following Tizón-Couto and Lorenz (2021), this paper adopts a deductive modeling approach, meaning that, with the analysis being grounded in a previously specified theory, all the tested variables are included in the final model rather than stripping the model of insignificant interactions or main effects. In doing so, the analysis is well-suited to directly test the formulated hypotheses independently of whether the analyzed interactions exhibit meaningful differences between regional varieties and/or over time.

https://doi.org/10.1017/S0954394525000055 Published online by Cambridge University Press

The Bayesian mixed-effects logistic regression model was specified to run 6000 iterations (3000 warmup, 2400 total post-warmup draws) on four chains. The target acceptance rate was set to 0.99 to reduce the risk of divergent transitions and increase the validity of the posterior samples (see Bürkner, 2021:10). Several model diagnostics were utilized: Rhat values for all parameters in the final model are 1.0, meaning that all the chains converge and mix well. Bulk-ESS and tail-ESS are all well above 1000, indicating that the estimates of posterior quantiles are reliable. Posterior predictive checks show that the observed data aligns very well with the predicted data, thus reaffirming the validity of the model's predictions. Lastly, Leave-One-Out cross-validation with moment matching (Paananen, Bürkner, Vehtari, & Gabry, 2024) indicated that the model's effective number of parameters (p_loo = 580.9, SE = 14.8) was considerably lower than the model's total number of parameters (2265), revealing no indications of overfitting. Moreover, all Pareto K estimates were lower than 0.7, signaling that these estimates are reliable. In the next section, the results are presented.

Results

Figure 1 shows density plots on the conventionalization of the innovative variant over time, indicating that the innovative [*en* + DEFINITE ARTICLE + *que*] variant gains substantially in frequency throughout the analyzed time span. In the Colombian data, this variant is almost absent until approximately the 1960s, when a rapid phase of change is initiated. In contrast, the Argentinean and Peruvian data shows a much earlier use of the construction, and a seemingly steadier change over the analyzed timespan. These frequency differences notwithstanding, the question that will be the focus in the remainder of the analysis is whether these communities differ or coincide in their pathways of actualization, that is, how change is mapped out across linguistic environments.

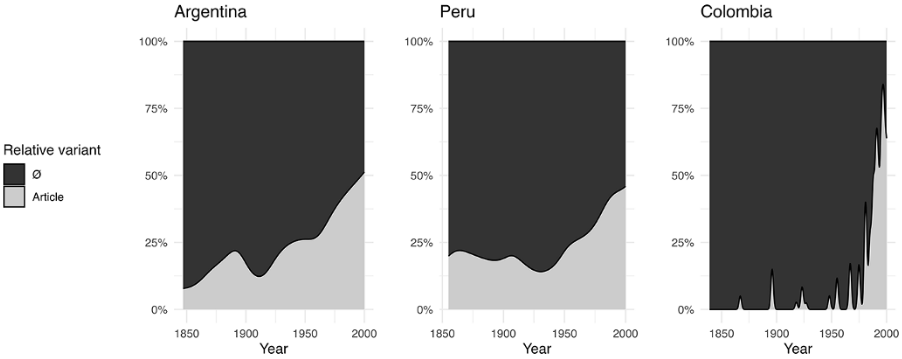

Figure 1. Density plots of the conventionalization of the [*en* + DEFINITE ARTICLE + *que*] variant across the three macroregional varieties. Bandwidth is adjusted to 1.

The final regression model is shown in Table 1. The meaningful (significant) effects are marked in bold.

https://doi.org/10.1017/S0954394525000055 Published online by Cambridge University Press

Table 1. Results from Bayesian mixed-effects logistic regression model

Predictor	Estimate	Est. Error	Lower 95% CI	Upper 95% CI	Inside ROPE
Intercept	−6.06	0.60	−7.24	−4.86	0.00%
I. Main effects					
Variety: Peru	0.22	0.67	−1.15	1.54	21.71%
Variety: Colombia	−0.58	1.32	−3.34	1.80	11.36%
Year (centered)	0.01	0.01	−0.00	0.03	100.00%
Attraction (log)	−0.43	0.07	−0.56	−0.30	**0.00%**
Distance (log)	0.65	0.10	0.46	0.84	**0.00%**
Antecedent number: Plural	0.50	0.15	0.20	0.80	**0.00%**
Restrictiveness: Nonrestrictive	1.53	0.17	1.20	1.87	**0.00%**
Definiteness: Indefinite	1.82	0.16	1.52	2.13	**0.00%**
Definiteness: Zero	1.16	0.19	0.81	1.53	**0.00%**
Antecedent word length (log)	0.39	0.23	−0.07	0.85	17.24%
Syntactic prominence: Subject/Root	0.01	0.16	−0.30	0.32	78.03%
II. Two-way interactions					
Variety: Peru × Year (centered)	0.02	0.01	−0.01	0.04	100.00%
Variety: Colombia × Year (centered)	0.05	0.03	0.00	0.10	100.00%
Variety: Peru × Attraction (log)	0.09	0.07	−0.04	0.22	92.41%
Variety: Colombia × Attraction (log)	0.06	0.18	−0.32	0.40	67.85%
Variety: Peru × Distance (log)	0.01	0.18	−0.36	0.37	72.63%
Variety: Colombia × Distance (log)	−0.03	0.26	−0.55	0.48	53.11%
Variety: Peru × Antecedent number: Plural	0.22	0.23	−0.23	0.67	39.87%
Variety: Colombia × Antecedent number: Plural	0.30	0.51	−0.68	1.31	24.82%
Variety: Peru × Restrictiveness: Nonrestrictive	0.03	0.26	−0.46	0.54	52.50%
Variety: Colombia × Restrictiveness: Nonrestrictive	0.84	0.52	−0.16	1.89	8.20%

(*Continued*)

Table 1. (*Continued.*)

Predictor	Estimate	Est. Error	Lower 95% CI	Upper 95% CI	Inside ROPE
Variety: Peru × Definiteness: Indefinite	0.03	0.23	−0.42	0.47	61.14%
Variety: Colombia × Definiteness: Indefinite	−1.39	0.54	−2.51	−0.38	**0.00%**
Variety: Peru × Definiteness: Zero	0.31	0.26	−0.22	0.82	30.39%
Variety: Colombia × Definiteness: Zero	−0.62	0.60	−1.79	0.55	14.74%
Variety: Peru × Antecedent word length (log)	−0.22	0.29	−0.80	0.33	37.98%
Variety: Colombia × Antecedent word length (log)	−0.45	0.67	−1.76	0.90	18.33%
Variety: Peru × Syntactic prominence: Subject/Root	0.30	0.22	−0.11	0.75	29.17%
Variety: Colombia × Syntactic prominence: Subject/Root	−0.73	0.63	−2.07	0.42	13.11%
Year (centered) × Attraction (log)	0.00	0.00	−0.00	0.00	100.00%
Year (centered) × Distance (log)	−0.00	0.00	−0.01	0.00	100.00%
Year (centered) × Antecedent number: Plural	0.00	0.00	−0.00	0.01	100.00%
Year (centered) × Restrictiveness Nonrestrictive	−0.01	0.00	−0.01	0.00	100.00%
Year (centered) × Definiteness: Indefinite	−0.00	0.00	−0.01	0.00	100.00%
Year (centered) × Definiteness: Zero	−0.00	0.00	−0.01	0.00	100.00%
Year (centered) × Antecedent word length (log)	0.01	0.00	0.00	0.02	100.00%
Year (centered) × Syntactic prominence: Subject/Root	0.00	0.00	−0.00	0.01	100.00%
III. Three-way interactions					
Variety: Peru × Year (centered) × Attraction (log)	−0.00	0.00	−0.00	0.00	100.00%

(*Continued*)

https://doi.org/10.1017/S0954394525000055 Published online by Cambridge University Press

Table 1. (*Continued.*)

Predictor	Estimate	Est. Error	Lower 95% CI	Upper 95% CI	Inside ROPE
Variety: Colombia × Year (centered) × Attraction (log)	−0.00	0.00	−0.01	0.01	100.00%
Variety: Peru × Year (centered) × Distance (log)	−0.00	0.00	−0.01	0.01	100.00%
Variety: Colombia × Year (centered) × Distance (log)	0.00	0.01	−0.01	0.01	100.00%
Variety: Peru × Year (centered) × Antecedent number: Plural	−0.01	0.01	−0.02	−0.00	100.00%
Variety: Colombia × Year (centered) × Antecedent number: Plural	−0.01	0.01	−0.03	0.01	100.00%
Variety: Peru × Year (centered) × Restrictiveness: Nonrestrictive	0.02	0.01	0.00	0.03	100.00%
Variety: Colombia × Year (centered) × Restrictiveness: Nonrestrictive	−0.01	0.01	−0.03	0.01	100.00%
Variety: Peru × Year (centered) × Definiteness: Indefinite	0.00	0.00	−0.01	0.01	100.00%
Variety: Colombia × Year (centered) × Definiteness: Indefinite	0.03	0.01	0.01	0.05	100.00%
Variety: Peru × Year (centered) × Definiteness: Zero	−0.00	0.01	−0.01	0.01	100.00%
Variety: Colombia × Year (centered) × Definiteness: Zero	0.02	0.01	−0.00	0.04	100.00%
Variety: Peru × Year (centered) × Antecedent word length (log)	−0.02	0.01	−0.03	−0.00	100.00%
Variety: Colombia × Year (centered) × Antecedent word length (log)	−0.02	0.01	−0.04	0.01	100.00%
Variety: Peru × Year (centered) × Syntactic prominence: Subject/Root	−0.00	0.00	−0.01	0.01	100.00%
Variety: Colombia × Year (centered) × Syntactic prominence: Subject/Root	0.01	0.01	−0.01	0.04	100.00%

As Table 1 shows, in the vast majority of cases the included interactions did not show statistically credible effects, meaning that, as concerns the process of change and the included varieties at hand, actualization does principally unfold in a comparable manner across the analyzed varieties The only exception is the interaction between VARIETY (Colombia) × DEFINITENESS (Indefinite). I will return to the interpretation of this finding below.

Focusing exclusively on the main effects, numerous factors constrain the analyzed variation, namely: ATTRACTION, DISTANCE, ANTECEDENT NUMBER, RESTRICTIVENESS, and DEFINITENESS (see the first part of Table 1). As indicated by the ROPE (Table 1), ANTECEDENT WORD LENGTH and SYNTACTIC PROMINENCE were not significant. While the reason underlying the statistical non-significance of ANTECEDENT WORD LENGTH and SYNTACTIC PROMINENCE warrants further scrutiny (but see the background section on the inherent limitations of the operationalization of these variables), due to space limitations the rest of the analysis will focus on the factors that do have a meaningful effect on the analyzed variation. What is worth noting, however, is that the same factors were significant and nonsignificant across historical macroregional varieties. This confirms that actualization largely hinges on cognitive and usage-conditioned constraints, and—importantly—that their conditioning is shared across different regional varieties.

To facilitate the interpretation of the interactions, the effects from the regression model are depicted using the *R* package *interactions* (version 1.2.0, Long, 2019).

Attraction

In Figure 2, the influence of ATTRACTION is depicted. Lower values of ATTRACTION (i.e., the lower tercile median) indicate that the antecedent is (relatively) infrequent; conversely, higher values (i.e., the upper tercile median) reflect antecedents that are (relatively) frequent. In accessibility parlance, this means that lower values are symptomatic of antecedents that are less accessible than antecedents that are frequency-wise highly accessible.

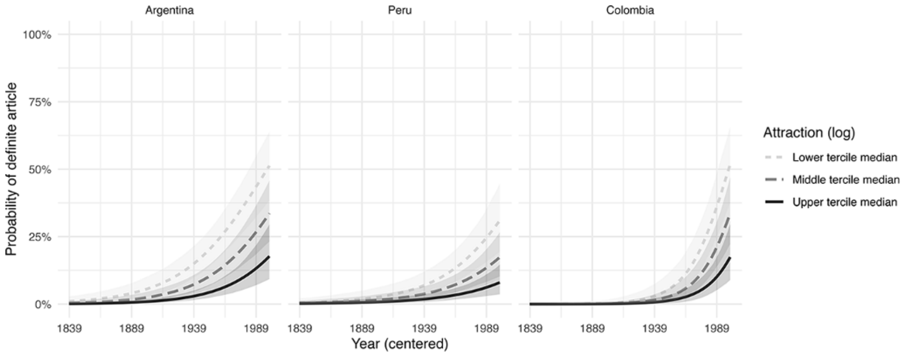

Figure 2. Predicted probabilities of [*en* + DEFINITE ARTICLE+ *que*] under the constraint of ATTRACTION. Values of ATTRACTION are depicted per tercile medians (lower, middle, upper).

As Figure 2 shows, the analyzed varieties are strikingly similar in their trajectories of actualization. The probability of the [*en* + DEFINITE ARTICLE + *que*] variant is significantly influenced by ATTRACTION in all the varieties: hapax legomena

https://doi.org/10.1017/S0954394525000055 Published online by Cambridge University Press

antecedents and other lower-frequency antecedents trigger use of the definite article ORC, whereas higher-frequency antecedents tend not to trigger use of the definite article. This observation corroborates the hypothesis that higher-frequency antecedents benefit from an overall expectedness which endows them with a privileged ease of maintenance and access in working memory. Conversely, less predictable antecedents are more frequently activated through a definite article. There is also likely an entrenchment effect at play here: more frequent collostructional patterns (that is, [high frequency antecedent + relative clause] as in *el año en (el) que* 'the year in which') hold out against the incorporation of the definite article to a larger extent, because frequently co-occurring structures are entrenched in the memory of the language user (Langacker, 1987). Their entrenchment renders modifying them more unlikely than less entrenched patterns (e.g., [low frequency antecedent + relative clause] as in *la tempestad en (la) que* 'the storm in which'; see Levshina, 2018).

Antecedent number

In consonance with the formulated hypothesis, the longer [*en* + DEFINITE ARTICLE + *que*] variant is favored by plural antecedents (Figure 3). This is arguably because the definite article serves to disambiguate and reactivate inaccessible antecedents, and plural entities are cognitively more complex to construct than a single reference (Jaeger & Wasow, 2005). Moreover, Figure 3 shows that the trajectories of actualization converge between the three varieties.

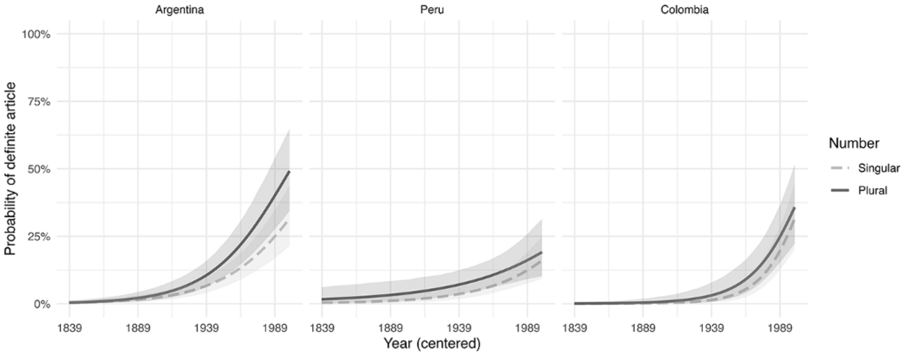

Figure 3. Predicted probabilities of [*en* + DEFINITE ARTICLE + *que*] under the constraint of ANTECEDENT NUMBER.

Distance

Figure 4 shows the effect of the distance between the antecedent and the ORC. The lower tercile median shows smaller distances, whereas higher values indicate larger distances. This predictor is highly influential in constraining the use of the [*en* + DEFINITE ARTICLE + *que*] variant: the more distanced the antecedent, the more probable the use of the definite article. This finding echoes the results of earlier studies on the linguistic

https://doi.org/10.1017/S0954394525000055 Published online by Cambridge University Press

variable and on relative clauses in general (Ariel, 1990:28-29; Blas Arroyo, 2021:500; Blas Arroyo & Vellón Lahoz, 2017, 2018; Clark & Sengul, 1979; Guzmán Riverón, 2012:182; Hawkins, 1999; Vellón Lahoz & Moya Isach, 2017:477). Considering that distance pressures are derived online in the discourse, this finding suggests that part of the variation is entirely motivated by processing effects. Again, the effect is cross-varietally stable.

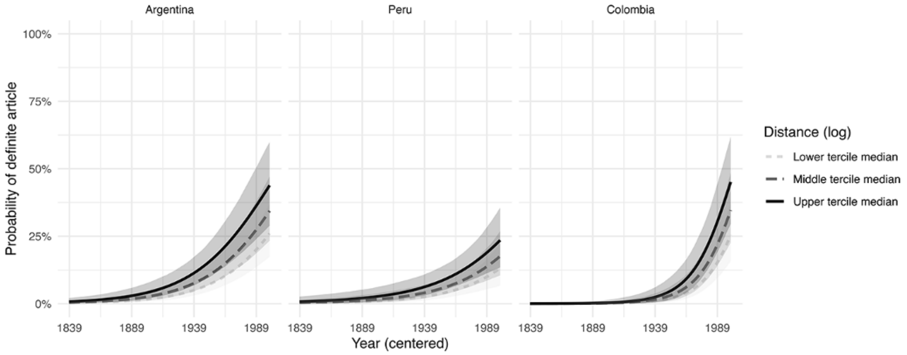

Figure 4. Predicted probabilities of [*en* + DEFINITE ARTICLE + *que*] under the constraint of DISTANCE.

Restrictiveness

Figure 5 depicts the result obtained for the predictor RESTRICTIVENESS. In line with earlier studies (see background section), nonrestrictive relative clauses particularly favor the [*en* + DEFINITE ARTICLE + *que*] variant, whereas the variant spreads to involve restrictive ORCs only at a much later stage. Again, actualization progresses along entirely parallel pathways across the three analyzed varieties.

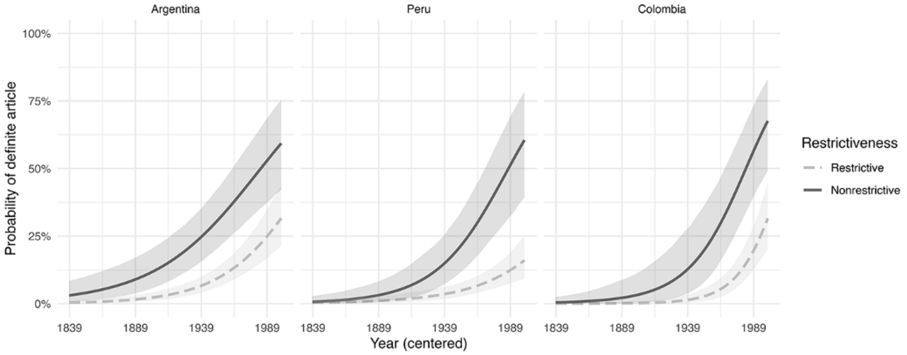

Figure 5. Predicted probabilities of [*en* + DEFINITE ARTICLE + *que*] under the constraint of RESTRICTIVENESS.

https://doi.org/10.1017/S0954394525000055 Published online by Cambridge University Press

Definiteness

Figure 6 shows that the influence of DEFINITENESS. This is the only predictor which was significant in an interaction with VARIETY (see Table 1), revealing that Colombian Spanish is less likely to use the [*en* + DEFINITE ARTICLE + *que*] with indefinite antecedents compared to Argentinian Spanish (cf. log-odds: −1.39, Table 1). The mechanisms underlying this effect are not immediately apparent from the current dataset and merit dedicated investigation in future studies. Apart from this difference, which also appears to be levelled out over time (cf. Figure 6), the trajectories of change are parallel, suggesting a largely uniform course of actualization.

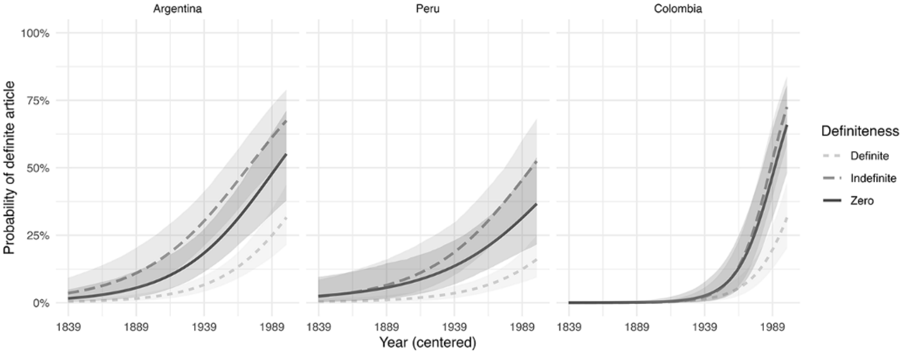

Figure 6. Predicted probabilities of [*en* + DEFINITE ARTICLE + *que*] under the constraint of DEFINITENESS.

In general, the use of the [*en* + DEFINITE ARTICLE + *que*] variant is particularly favored when the antecedent NP is indefinite. In the Colombian data, the difference between indefinite and zero-marked antecedents is minimal. What is shared across the analyzed datasets is that the use of the variant is probabilistically most improbable with definite antecedent NPs. This result agrees with what earlier studies on this linguistic variable have found (Blas Arroyo, 2021:500; Blas Arroyo & Vellón Lahoz, 2017, 2018:41; Girón Alconchel, 2006:1530; Vellón Lahoz & Moya Isach, 2017:476; see also Ariel, 1996:22).

As can be seen across the three varieties, the use of the [*en* + DEFINITE ARTICLE + *que*] variant is usually higher with indefinites. Only more recently has the innovative variant spread to involve definite antecedents.

Discussion and concluding remarks

This study analyzed the real-time mapping out of syntactic change across linguistic environments from a cross-dialectal perspective, seeking to determine to what extent pathways of actualization converge or diverge across varieties of a language. Actualization—defined as the gradual mapping out of new grammatical behavior across linguistic environments—was approximated through the diachronic analysis of real-time effects of fundamentally cognitive processes (cf. Wolk et al., 2013). The fundamental question that the study sought to shed light on was this: When multiple

https://doi.org/10.1017/S0954394525000055 Published online by Cambridge University Press

varieties of a language undergo the same syntactic change, do they follow parallel trajectories of actualization? In pursuing this issue, the study has showcased the potential of a *cross-varietal diachronic variationist approach* to studies on language change.

Overall, the diachronic data show that the definite article in oblique relative clauses is conventionalizing in all the analyzed varieties. The results indicate that the same cognitive and usage-determined factors guide the course of actualization in the different varieties. Moreover, cross-dialectal comparisons of the courses of actualization reveal that trajectories are largely parallel varieties: The definite article is initially inserted in ORCs when it is most informative, that is, with plural, non-definite, and frequency-wise unpredictable and inaccessible antecedents. I argue that, under these conditions and in an initial phase, the definite article serves as an accessibility marker (cf. Ariel, 1990). However, during the actualization of the change, the innovative variant spreads to environments in which it is less informative in terms of accessibility. The order in which it spreads across environments reflects gradience in accessibility: from hapax legomena antecedents to mid-frequency antecedents and, only later, to high-frequency antecedents; from plural to singular antecedents; and from indefinite antecedents to zero-marked antecedents and, later on, definite antecedents. In other words, the order of diffusion is determined by the degree to which it is purposeful in terms of accessibility. This finding aligns well with the predicted course of change: innovative forms that are longer and costlier should first be used to convey less accessible meanings, and subsequently spread towards high-accessibility contexts (Jaeger, 2010; Levshina, 2022a:135; see also Marttinen Larsson, 2024).

Over time, the accessibility effect seems to partially grammaticalize (cf. Jaeger, 2010:49-50). Other indicators that the [*en* + DEFINITE ARTICLE + *que*] variant is, indeed, grammaticalizing are its increased frequency of use (in terms of token frequency; Hopper & Traugott, 2003 [1993]), the scope increase of the structure in terms of contexts of use (Tabor & Traugott, 1998:262), and the semantic change undergone by the definite article (i.e., accessibility is lost and only number and gender are maintained; cf. Girón Alconchel, 2004). Blas Arroyo and Vellón Lahoz (2018) examined the potential grammaticalization of *en* (*el*) *que* in European Spanish across three centuries, identifying only modest frequency increases along with enduring linguistic conditioning. However, the study's capability to ascertain whether grammaticalization has taken place is restricted by their methodology of using separate *Rbrul* models for each century rather than a single regression model incorporating time interactions.

Utilizing a real-time diachronic approach, the present study has offered empirical support substantiating the hypothesis that the [*en* + DEFINITE ARTICLE + *que*] variant is currently grammaticalizing (cf. Girón Alconchel, 2004, 2006; among others), as suggested by the innovation's scope increase, its higher frequency of use, and the semantic loss undergone by the accessibility marker. Moreover, this cross-varietal diachronic inquiry has also verified that numerous domain-general cognitive and usage factors steer actualization (e.g., De Smet, 2012b; Schmid, 2020; Wolk et al., 2013). Crucially, these factors operate similarly across varieties of a language, even when the grammatical change is taking place at different rates and at different times. This implies that the cognitive mechanisms that trigger and fuel variation and change are largely ubiquitous (cf. Langacker, 1977:99-100) and sheds light onto a fundamental principle underlying

https://doi.org/10.1017/S0954394525000055 Published online by Cambridge University Press

the analyzed process of change: while language varieties may actuate change at different times, cognitive processes guide the course of actualization uniformly across varieties, resulting in parallel pathways of change.

The cognitive corpus linguistic approach used in this study naturally presupposes that historical corpus data is capable of illuminating effects that are operative online in cognitive processing (Wolk et al., 2013). This is not entirely uncontroversial, given that language data from offline corpora can, strictly speaking, only provide indirect evidence of processes that occurred online in the mind of language users of past time periods, which raises the question of how well these offline sources mirror psychological reality and cognitive processes (Arppe, Gilquin, Glynn, Hilpert, & Zeschel, 2010; Kortmann, 2021; Wolk et al., 2013:414). If we regard the cognitive realism of historical corpus data as an empirical issue, this study has demonstrated that numerous predictions formulated on the basis of experimental psycholinguistic and cognitive research are borne out on historical corpus data (see also Wolk et al., 2013). This synergy between synchronic experimental research and diachronic linguistics provides a framework for effectively modeling how linguistic innovations become established in the grammar during the course of actualization and the cognitive underpinnings of this process.

With the above in mind, the findings presented in this study allow for the formulation a series of predictions for future diachronic, cognitivist, corpus-based research. For one, it is not only conceivable, but expected that, to the extent that different speech communities exhibit shifting input probabilities between structures (i.e., lexical items), micro level divergencies should be observed between regional varieties during the course of actualization (cf. Bresnan & Hay, 2008:255-256). The availability of different linguistic material in the mental lexicon is largely socially determined, and the mental representation of certain structures (or lexemes) hinges on how frequently the language user is exposed to it (Verhagen et al., 2018). Therefore, when an instance of change is underway across different varieties, it is predicted that the different lexical items that partake in, or are affected by, change (e.g., the lexical items that occupy a constructional slot) will vary according to their cognitive salience (accessibility) in the mental lexicon of a speaker and, by extension, their degree of activation in a specific speech community or social group.

Secondly, this study has identified that the actualization of change largely aligns cross-varietally in terms of the contexts that are affected by change and the order in which extensions take place (cf. De Smet, 2012b). In accordance with these findings, it is also predicted that micro level divergencies as the ones discussed above should, nevertheless, hinge on the same cognitive principles and thus align on the macro level. What this means is that, even if, for instance, the concrete lexical items that occupy certain constructional slots or that are otherwise affected by change may likely differ cross-varietally, the cognitive motivations that account for how change unfolds will align. Consequently, during changes that involve reduction (e.g., Marttinen Larsson, 2024), the specific structures or items that are omittable or reducible may differ between regional varieties, but these structures or items will all be highly accessible (predictable) in the respective speech communities or in certain social groups; conversely, cross-varietal changes that involve enhancement (such as the one analyzed in the present study) will most likely show gradience that is fundamentally conditioned by inaccessibility (see Levshina's, 2022a:230) *Hypothesis of Construction–Lexeme Accessibility and Formal Length*; see also Bresnan & Hay, 2008; Jaeger, 2010). Empirical verification of these predictions will be profitable lines of inquiry in future intervarietal diachronic studies on the cognitive underpinnings of actualization.

Acknowledgements. I acknowledge the financial support of the Swedish Research Council (grant no. 2022-00303, "Pathways of development around the Spanish-speaking world: diachronic variation and change in 19th–21st century Spanish morpho-syntax"). I am grateful to four anonymous reviewers for valuable comments and suggestions. I would also like to thank Miriam Bouzouita, Marta Garbelli, Laura Merino Hernández, and Antoine Primerano for their feedback, as well as the audiences at the 6th Variation and Language Processing Conference (Vigo, June 26–28, 2024), the 57th Annual Meeting of the Societas Linguistica Europaea (Helsinki, August 21–24, 2024), and Congreso Internacional de Lenguas Iberorrománicas (Helsinki, January 10–12, 2024). I am grateful to Johnatan Bonilla for his support with the spaCy annotation. The usual disclaimers apply.

Competing interests. The author declares none.

Note

[1] While the influence of factors related to register fall outside the scope of the present study, including the AUTHOR as a random effect can mitigate possible biases induced by register/genre (and, conceivably, corpus-specific) effects since individual authors can be used as a proxy for these possibly confounding variables (Marttinen Larsson, 2023:201). This matter will not be pursued further here.

References

Aaron, Jessi E. (2010). Pushing the envelope: Looking beyond the variable context. *Language Variation and Change 22*(1):1–36. https://doi.org/10.1017/S0954394509990226.

Andersen, Henning. (2001). Actualization and the (uni)directionality of change. In H. Andersen (ed.), *Actualization: Linguistic Change in Progress*. John Benjamins, 225–248.

Ariel, Mira. (1990). *Accessing Noun-phrase Antecedents*. Routledge.

Ariel, Mira. (1996). Referring expressions and the ±coreference distinction. In T. Fretheim, and J. K. Gundel (eds.), *Reference and Referent Accessibility*. John Benjamins, 13–36.

Arnold, Jennifer. (2010). How speakers refer: The role of accessibility. *Language and Linguistics Compass 4*(4):187–203. https://doi.org/10.1111/j.1749-818X.2010.00193.x.

Arppe, Anti, Gilquin, Gaëtanelle, Glynn, Dylan, Hilpert, Martin, & Zeschel, Arne. (2010). Cognitive Corpus Linguistics: Five points of debate on current theory and methodology. *Corpora 5*(1):1–27. https://doi.org/10.3366/cor.2010.0001.

Barking, Marie, Backus, Ad, & Mos, Maria. (2022). Individual corpus data predict variation in judgements: Testing the usage-based nature of mental representations in a language transfer setting. *Cognitive Linguistics 33*(3):481–519. https://doi.org/10.1515/cog-2021-0105.

Blas Arroyo, José Luis. (2021). Traces of the past in a lengthy change (still) in progress: Persistence and generalization in prepositional relative clauses in Peninsular Spanish. In M. Díaz-Campos (ed.), *The Routledge Handbook of Variationist Approaches to Spanish*. Routledge, 492–505.

Blas Arroyo, José Luis, & Vellón Lahoz, Javier. (2017). En los albores de un cambio lingüístico: Factores condicionantes y fases en la inserción del artículo en relativas oblicuas del siglo XVIII. *Zeitschrift Für Romanische Philologie 133*(2):492–529. https://doi.org/10.1515/zrp-2017-0024.

Blas Arroyo, José Luis, & Vellón Lahoz, Javier. (2018). On the trail of grammaticalization in progress: Has *el que* become a compound relative pronoun in the history of Spanish prepositional relative clauses? *Probus 30*(1):1–45. https://doi.org/10.1515/probus-2017-0010.

Bresnan, Joan, & Ford, Marilyn. (2010). Predicting syntax: Processing dative constructions in American and Australian varieties of English. *Language 86*(1):168–213. https://doi.org/10.1353/lan.0.0189.

Bresnan, Joan, & Hay, Jennifer. (2008). Gradient grammar: An effect of the animacy on the syntax of *give* in New Zealand and American English. *Lingua 118*(2):245–259. https://doi.org/10.1016/j.lingua.2007.02.007.

Bürkner, Paul-Christian. (2021). Bayesian item response modeling in R with brms and Stan. *Journal of Statistical Software 100*(5):1–54. https://doi.org/10.18637/jss.v100.i05.

Bybee, Joan. (2006). From usage to grammar: The mind's response to repetition. *Language 82*(4):711–733. https://doi.org/10.1353/lan.2006.0186.

Bybee, Joan. (2007). *Frequency of Use and the Organization of Language*. Oxford University Press.

Bybee, Joan. (2010). *Language, Usage, and Cognition*. Cambridge University Press.

Caravedo, Rocío. (2019). Reflexiones sobre la interrelación entre diacronía y diatopía. In V. Codita, and M. de la Torre (eds.), *Tendencias y perspectivas en el estudio de la morfosintaxis histórica*. Iberoamericana/Vervuert, 19–42.

Clark, Herbert H., & Sengul, C. J. (1979). In search of referents for nouns and pronouns. *Memory & Cognition 7*(1):35–41. https://doi.org/10.3758/BF03196932.

CORDE = Real Academia Española: Banco de datos (CORDE) [en línea]. Corpus diacrónico del español. http://www.rae.es [Consulted in 2023]

CREA = Real Academia Española: Banco de datos (CREA) [en línea]. Corpus de referencia del español actual. http://www.rae.es [Consulted in 2023]

De Smet, Hendrik. (2009). Analysing reanalysis. *Lingua 119*(11):1728–1755. https://doi.org/10.1016/j.lingua.2009.03.001.

De Smet, Hendrik. (2012a). *Spreading Patterns: Diffusional Change in the English System of Complementation*. Oxford University Press.

De Smet, Hendrik. (2012b). The course of actualization. *Language 88*(3):601–633. https://doi.org/10.1353/lan.2012.0056.

De Smet, Hendrik. (2016). How gradual change progresses: The interaction between convention and innovation. *Language Variation and Change 28*(1):83–102. https://doi.org/10.1017/S0954394515000186.

De Smet, Hendrik, & Fischer, Olga. (2017). The role of analogy in language change: Supporting constructions. In M. Hundt, S. Mollin & S. E. Pfenninger (eds.), *The Changing English Language: Psycholinguistic Perspectives*. Cambridge University Press, 240–268.

Dietrich, Nadine. (2024). The seamlessness of grammatical innovation: The case of *be going to* (revisited). *Folia Linguistica 58*(s45-s1):149–183. https://doi.org/10.1515/flin-2024-2004.

Dion, Nathalie. (2023). A question of change: Putting five complementary measures to the test with French polar interrogatives. *Language Variation and Change 35*(3):247–271. https://doi.org/10.1017/S0954394523000170.

Du Bois, John W. (1985). Competing motivations. In J. Haiman (ed.), *Iconicity in Syntax*. John Benjamins, 343–365.

Explosion. (2023). Spanish transformer pipeline. https://github.com/explosion/spacy-models/releases/tag/es_dep_news_trf-3.7.2

Gelman, Andrew, Jakulin, Aleks, Grazia Pittau, Maria, & Su, Yu-Sung. (2008). A weakly informative default prior distribution for logistic and other regression models. *The Annals of Applied Statistics 2*(4):1360–1383. https://doi.org/10.1214/08-AOAS191.

Girón Alconchel, José Luis. (2006). Las oraciones de relativo II. Evolución del relativo compuesto EL QUE, LA QUE, LO QUE. In C. Company (ed.), *Sintaxis Histórica de la Lengua Española. Segunda Parte: La Frase Nominal*. (Vol. 2, UNAM/FCE, 1477–1590.

Girón Alconchel, José Luis. (2004). Gramaticalización y estado latente. *Dicenda 22*:71–88.

Grafmiller, Jason. (2023). Visualizing grammatical similarities in comparative variationist analysis. In O. Schützler & L. Sönning (eds.), *Data Visualization in Corpus Linguistics: Critical Reflections and Future Directions*. VARIENG, 1–44. https://urn.fi/URN:NBN:fi:varieng:series-22-4

Grafmiller, Jason, Szmrecsanyi, Benedikt, Röthlisberger, Melanie, & Heller, Benedikt. (2018). General introduction: A comparative perspective on probabilistic variation in grammar. *Glossa*, *3*(1):1–10.

Greenberg, Joseph H. (1966). *Language Universals: With Special Reference to Feature Hierarchies*. Mouton.

Guasch, Marc, Ferré, Pilar, & Fraga, Isabel. (2016). Spanish norms for affective and lexico-semantic variables for 1,400 words. *Behavior Research Methods 48*:1358–1369. https://doi.org/10.3758/s13428-015-0684-y.

Guzmán Riverón, Martha. (2012). El artículo en las relativas oblicuas [prep. + (art. definido) + *que*] en textos americanos del siglo XVIII. *Cuadernos Dieciochistas*, *13*:171–204.

Harris, Alice C., & Campbell, Lyle. (1995). *Historical Syntax in Cross-linguistic Perspective*. Cambridge University Press.

Haspelmath, Martin. (2021). Explaining grammatical coding asymmetries: Form-frequency correspondences and predictability. *Journal of Linguistics 57*(3):605–633. https://doi.org/10.1017/S0022226720000535.

Hawkins, John A. (1999). Processing complexity and filler-gap dependencies across grammars. *Language*, *75*(2):244–285. https://doi.org/10.2307/417261.

Heine, Angela, Tamm, Sascha, Hofmann, Markus, Hutzler, Florian, & Jacobs, Arthur M. (2006). Does the frequency of the antecedent noun affect the resolution of pronominal anaphors? An ERP study. *Neuroscience Letters 400*(1):7–12.

Hopper, Paul J., & Traugott, Elizabeth C. (2003 [1993]). *Grammaticalization* (2nd ed.). Cambridge University Press.

Jaeger, T. Florian. (2010). Redundancy and reduction: Speakers manage syntactic information density. *Cognitive Psychology 61*(1):23–62. https://doi.org/10.1016/j.cogpsych.2010.02.002.

Jaeger, T. Florian, & Wasow, Thomas. (2005). Processing as a Source of Accessibility Effects on Variation. In R. T. Cover, & Y. Kin (eds.), *Annual Meeting of the Berkeley Lingusitics Society*. Linguistic Society of America, 169–180.

Jessen, Frank, Heun, Reinhard, Erb, Michael, Granath, Dirk Oliver, Klose, Uwe, Papassotiropoulos, Andreas, & Grodd, Wolfgang. (2000). The concreteness effect: Evidence for dual coding and context availability. *Brain and Language 74*(1):103–112. https://doi.org/10.1006/brln.2000.2340.

Kimball, Amelia E., Shantz, Kailen, Eager, Christopher, & Roy, Joseph. (2019). Confronting quasi-separation in logistic mixed-effects for linguistic data: A Bayesian approach. *Journal of Quantitative Linguistics 26*(3):231–255. https://doi.org/10.1080/09296174.2018.1499457.

Kortmann, Bernd. (2021). Reflecting on the quantitative turn in linguistics. *Linguistics 59*(5):1207–1226. https://doi.org/10.1515/ling-2019-0046.

Kruschke, John. (2015). *Doing Bayesian Data Analysis: A Tutorial with R, JAGS, and Stan* (2nd ed.). Elsevier.

Langacker, Ronald W. (1977). Syntactic reanalysis. In C. N. Li (ed.), *Mechanisms of Syntactic Change*. University of Texas Press, 59–139.

Langacker, Ronald W. (1987). *Foundations for Cognitive Grammar. Theoretical Prerequisites*. Stanford University Press.

Levshina, Natalia. (2018). Probabilistic grammar and constructional predictability: Bayesian generalized additive models of *help* + (to) Infinitive in varieties of web-based English. *Glossa: A Journal of General Linguistics 3*(1):55.

Levshina, Natalia. (2022a). *Communicative Efficiency: Language Structure and Use*. Cambridge University Press.

Levshina, Natalia. (2022b). Comparing Bayesian and frequentist models of language variation: The case of *help* + (*to*-)Infinitive. In O. Schützler, and J. Schlüter (eds.), *Data and Methods in Corpus Linguistics: Comparative Approaches*. Cambridge University Press, 224–258.

Lewis, Molly L., & Frank, Michael C. (2016). The length of words reflects their conceptual complexity. *Cognition 153*:182–195. https://doi.org/10.1016/j.cognition.2016.04.003.

Lievers, Francesca Strik, Bolognesi, Marianna, & Winter, Bodo. (2021). The linguistic dimensions of concrete and abstract concepts: Lexical category, morphological structure, countability, and etymology. *Cognitive Linguistics 32*(4):641–670. https://doi.org/10.1515/cog-2021-0007.

Long, Jacob A. (2019). *interactions: comprehensive, user-friendly toolkit for probing interactions*. https://cran.r-project.org/package=interactions

López García, Ángel. (1994). *Gramática del Español. La Oración Compuesta*. Arco Libros.

Makowski, Dominique, Ben-Shachar, Mattan S., & Lüdecke, Daniel. (2019). bayestestR: Describing effects and their uncertainty, existence and significance within the Bayesian framework. *Journal of Open Source Software 4*(40):1541.

Marttinen Larsson, Matti. (2023). Modelling incipient probabilistic grammar change in real time: The grammaticalisation of possessive pronouns in European Spanish locative adverbial constructions. *Corpus Linguistics and Linguistic Theory 19*(2):177–206. https://doi.org/10.1515/cllt-2021-0030.

https://doi.org/10.1017/S0954394525000055 Published online by Cambridge University Press

Marttinen Larsson, Matti. (2024). Probabilistic reduction and constructionalization: A usage-based diachronic account of the diffusion and conventionalization of the Spanish *la de* <noun> *que* construction. *Cognitive Linguistics 35*(4):579–602.

McKoon, Gail, & Ratcliff, Roger. (1980). The comprehension processes and memory structures involved in anaphoric reference. *Journal of Verbal Learning and Verbal Behavior 19*(6):668–682. https://doi.org/10.1016/S0022-5371(80)90355-2.

Naro, Anthony J. (1981). The social and structural dimensions of a syntactic change. *Language 57*(1):63–98. https://doi.org/10.1353/lan.1981.0020.

Nevalainen, Terttu, & Raumolin-Brunberg, Helena. (2017). *Historical Sociolinguistics: Language Change in Tudor and Stuart England*. Routledge.

Paananen, Topi, Bürkner, Paul, Vehtari, Aki, & Gabry, Jonah. (2024). Avoiding model refits in leave-one-out cross-validation with moment matching. https://mc-stan.org/loo/articles/loo2-moment-matching.html

Poplack, Shana, Lealess, Allison, & Dion, Nathalie. (2013). The evolving grammar of the French subjunctive. *Probus 25*(1):139–195. https://doi.org/10.1515/probus-2013-0005.

Poplack, Shana, & Malvar, Elisabete. (2007). Elucidating the transition period in linguistic change: The expression of the future in Brazilian Portuguese. *Probus 19*(1):121–169. https://doi.org/10.1515/PROBUS.2007.005.

Poplack, Shana, & Tagliamonte, Sali. (2001). *African American English in the Diaspora*. Blackwell.

R Core Team. (2022). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. https://www.R-project.org

Reilly, Jamie, Hung, Jinyi, & Westbury, Chris. (2017). Non-arbitrariness in mapping word form to meaning: Cross-linguistic formal markers of word concreteness. *Cognitive Science 41*(4):1071–1089. https://doi.org/10.1111/cogs.12361.

Reilly, Jamie, Westbury, Chris, Kean, Jacob, & Peelle, Jonathan E. (2012). Arbitrary symbolism in natural language revisited: when word forms carry meaning. *PLoS ONE* 7(8): e42286.

Röthlisberger, Melanie, Grafmiller, Jason, & Szmrecsanyi, Benedikt. (2017). Cognitive indigenization effects in the English dative alternation. *Cognitive Linguistics 28*(4):673–710. https://doi.org/10.1515/cog-2016-0051.

Santana Marrero, Juana. (2004). Preposición + (artículo) + que relativo: Análisis en la norma lingüística culta panhispánica. *Boletín de Lingüística 21*:66–91.

Schmid, Hans-Jörg. (2020). *The Dynamics of the Linguistic System: Usage, Conventionalization and Entrenchment*. Oxford University Press.

Schneider, Edgar W. (2023). Lexicosemantic diffusion in World Englishes: Variable meaning-form relations in prospective verbs. *English Language & Linguistics 27*(4):719–748. https://doi.org/10.1017/S136067432300014X.

Szmrecsanyi, Benedikt, Grafmiller, Jason, Heller, Benedikt, & Röthlisberger, Melanie. (2016). Around the world in three alternations: Modeling syntactic variation in varieties of English. *English World-Wide 37*(2):109–137. https://doi.org/10.1075/eww.37.2.01szm.

Tabor, Whitney, & Traugott, Elizabeth Closs. (1998). Structural scope expansion and grammaticalization. In A. G. Ramat, and P. J. Hopper (eds.), *Limits of Grammaticalization* 229–272.

Tagliamonte, Sali. (2013). *Roots of English: Exploring the History of Dialects*. Cambridge University Press.

Tagliamonte, Sali, & Baayen, Harald R. (2012). Models, forests, and trees of York English: *Was/were* variation as a case study for statistical practice. *Language Variation and Change 24*(4): 135–178. https://doi.org/10.1017/S0954394512000129.

Tagliamonte, Sali A., Durham, Mercedes, & Smith, Jennifer. (2014). Grammaticalization at an early stage: Future *be going to* in conservative British dialects. *English Language and Linguistics 18*(1):75–108. https://doi.org/10.1017/S1360674313000282.

Timberlake, Alan. (1977). Reanalysis and actualization in syntactic change. In C. Li (ed.), *Mechanisms of Syntactic Change*. University of Texas, 141–177.

Tizón-Couto, David, & Lorenz, David. (2021). Variables are valuable: Making a case for deductive modeling. *Linguistics 59*(5):1279–1309. https://doi.org/10.1515/ling-2019-0050.

Torres Cacoullos, Rena, & Walker, James A. (2009). The present of the English future: Grammatical variation and collocations in discourse. *Language 85*(2):321–354. https://doi.org/10.1353/lan.0.0110.

https://doi.org/10.1017/S0954394525000055 Published online by Cambridge University Press

Vellón Lahoz, Javier, & Moya Isach, Rosa A. (2017). Pervivencia de las relativas oblicuas sin artículo: Factores y contextos condicionantes. *Spanish in Context 14*(3):464–486. https://doi.org/10.1075/sic.14.3.06vel.

Verhagen, Véronique, Mos, Maria, Backus, Ad, & Schilperoord, Joost. (2018). Predictive language processing revealing usage-based variation. *Language and Cognition 10*(2):329–373. https://doi.org/10.1017/langcog.2018.4.

Weinreich, Uriel, Labov, William, & Herzog, Marvin. (1968). Empirical foundations for a theory of language change. In W. P. Lehmann, and Y. Malkiel (eds.), *Directions for Historical Linguistics*. University of Texas Press, 95–195.

Wolk, Christoph, Bresnan, Joan, Rosenbach, Anette, & Szmrecsanyi, Benedikt. (2013). Dative and genitive variability in Late Modern English: Exploring cross-constructional variation and change. *Diachronica 30*(3):382–419. https://doi.org/10.1075/dia.30.3.04wol.

Appendix

The tables found in this Appendix (Tables A1–A15) show the distribution of the data across three time periods in each regional variety, cross-tabulated by each predictor variable included in the multivariate analysis.

Table A1. Distribution of variants across time periods in Argentinean data

	Zero variant (*en que*)	Definite article variant (*en el que*)	Total
1839 – 1899	1143 (85.8%)	189 (14.2%)	1332
1900 – 1949	668 (80.6%)	161 (19.4%)	829
1950 – 2000	793 (57%)	599 (43%)	1392
Total	2604 (73.3%)	949 (26.7%)	3553

Table A2. Distribution of variants across time periods in Peruvian data

	Zero variant (*en que*)	Definite article variant (*en el que*)	Total
1839 – 1899	498 (79.8%)	126 (20.2%)	624
1900 – 1949	558 (84.5%)	102 (15.5%)	660
1950 – 2000	825 (62.2%)	502 (37.8%)	1327
Total	1881 (72%)	730 (28%)	2611

Table A3. Distribution of variants across time periods in Colombian data

	Zero variant (*en que*)	Definite article variant (*en el que*)	Total
1839 – 1899	504 (98.6%)	7 (1.4%)	511
1900 – 1949	517 (98.7%)	7 (1.3%)	524
1950 – 2000	697 (51.7%)	650 (48.3%)	1347
Total	1718 (72.1%)	664 (27.9%)	2382

Table A4. Distribution of variants across grammatical number in Argentinean data

	Zero variant (*en que*)	Definite article variant (*en el que*)	Total
Singular	2074 (75.1%)	688 (24.9%)	2762
Plural	530 (67%)	261 (33%)	791
Total	2604 (73.3%)	949 (26.7%)	3553

https://doi.org/10.1017/S0954394525000055 Published online by Cambridge University Press

Table A5. Distribution of variants across grammatical number in Peruvian data

	Zero variant (*en que*)	Definite article variant (*en el que*)	Total
Singular	1425 (72.9%)	529 (27.1%)	1954
Plural	456 (69.4%)	201 (30.6%)	657
Total	1881 (72%)	730 (28%)	2611

Table A6. Distribution of variants across grammatical number in Colombian data

	Zero variant (*en que*)	Definite article variant (*en el que*)	Total
Singular	1296 (75.1%)	496 (24.9%)	1792
Plural	422 (67%)	168 (33%)	590
Total	1718 (72.1%)	664 (27.9%)	2382

Table A7. Distribution of variants across restrictiveness in Argentinean data

	Zero variant (*en que*)	Definite article variant (*en el que*)	Total
Restrictive	2337 (78.7%)	632 (21.3%)	2969
Nonrestrictive	267 (45.7%)	317 (54.3%)	584
Total	2604 (73.3%)	949 (26.7%)	3553

Table A8. Distribution of variants across restrictiveness in Peruvian data

	Zero variant (*en que*)	Definite article variant (*en el que*)	Total
Restrictive	1707 (77.6%)	493 (22.4%)	2200
Nonrestrictive	174 (42.3%)	237 (57.7%)	411
Total	1881 (72%)	730 (28%)	2611

Table A9. Distribution of variants across restrictiveness in Colombian data

	Zero variant (*en que*)	Definite article variant (*en el que*)	Total
Restrictive	1537 (76.1%)	482 (23.9%)	2019
Nonrestrictive	181 (49.9%)	182 (51.1%)	363
Total	1718 (72.1%)	664 (27.9%)	2382

Table A10. Distribution of variants across definiteness in Argentinean data

	Zero variant (*en que*)	Definite article variant (*en el que*)	Total
Definite	1951 (82.4%)	418 (17.6%)	2369
Indefinite	361 (49.6%)	367 (50.4%)	728
Zero	292 (64%)	164 (36%)	456
Total	2604 (73.3%)	949 (26.7%)	3553

https://doi.org/10.1017/S0954394525000055 Published online by Cambridge University Press

Table A11. Distribution of variants across definiteness in Peruvian data

	Zero variant (*en que*)	Definite article variant (*en el que*)	Total
Definite	1341 (82.1%)	292 (17.9%)	1633
Indefinite	259 (48.4%)	276 (51.6%)	535
Zero	281 (63.4%)	162 (36.6%)	443
Total	1881 (72%)	730 (28%)	2611

Table A12. Distribution of variants across definiteness in Colombian data

	Zero variant (*en que*)	Definite article variant (*en el que*)	Total
Definite	1179 (79.2%)	309 (20.8%)	1488
Indefinite	319 (56.3%)	248 (43.7%)	567
Zero	220 (67.3%)	107 (32.7%)	327
Total	1718 (72.1%)	664 (27.9%)	2382

Table A13. Distribution of variants across syntactic function in Argentinean data

	Zero variant (*en que*)	Definite article variant (*en el que*)	Total
Non-subject	2075 (73.5%)	750 (26.5%)	2825
Subject/Root	529 (72.66%)	199 (27.3%)	728
Total	2604 (73.3%)	949 (26.7%)	3553

Table A14. Distribution of variants across syntactic function in Peruvian data

	Zero variant (*en que*)	Definite article variant (*en el que*)	Total
Definite	1435 (72.5%)	543 (27.5%)	1978
Zero	446 (70.5%)	187 (29.5%)	633
Total	1881 (72%)	730 (28%)	2611

Table A15. Distribution of variants across syntactic function in Colombian data

	Zero variant (*en que*)	Definite article variant (*en el que*)	Total
Definite	1278 (71%)	523 (29%)	1801
Zero	440 (75.7%)	141 (24.3%)	581
Total	1718 (72.1%)	664 (27.9%)	2382

Cite this article: Marttinen Larsson, Matti. (2025). Pathways of actualization across regional varieties and the real-time dynamics of syntactic change. *Language Variation and Change* **37**:31–57. https://doi.org/10.1017/S0954394525000055

https://doi.org/10.1017/S0954394525000055 Published online by Cambridge University Press