Long-term preferences and patterns of instability in the evolution of Romance verbal morphology

7735+37+455 (15/16/248/5) Total words

Abstract

Romance languages contain irregular paradigms that display unpredictable formal variation (e.g., Spanish ***teng****o* 'I have', ***tien****es* 'you have') as well as regular ones that do not (e.g., ***mand****o* 'I send', ***mand****as* 'you send'). While the diachronic development and persistence of regularity and irregularity is well studied, the processes that underlie the distribution of irregular forms are not fully understood. Specifically, it is not clear whether the distribution of irregular and regular forms is driven by differential patterns of instability, with greater volatility on the part of infrequent items, or preferences rooted in general pressures toward greater predictability on the part of less frequent forms. We address this question using a novel phylogenetic model that allows us to assess the effect of different predictors on two different facets of evolution in lexical paradigms — speed of change, representing the overall instability of a system, and stationary probability, representing the long-term preference for different irregular pattern types — while controlling for a range of other factors. We find that more frequent verbs show a greater long-term preference for irregularity, but we find no relationship between frequency and instability. This suggests that the distribution of irregularity is chiefly driven by general term pressures, possibly rooted in processig and learning biases, and not by the higher historical volatility of low-frequency forms.

1   Introduction

Speakers of the world's languages are regularly confronted with the task of realizing forms from inflectional paradigms, at times in lexemes which they rarely encounter. In many cases, they must additionally contend with multiple possible realizations of a single unit of meaning (e.g. English *feed* $\sim$ *fed* but *seed* $\sim$ *seed-ed*), a common cross-linguistic phenomenon known as allomorphy. Speakers have the capacity to produce some (e.g. unencountered) forms on the basis of others within the same or a different paradigm. The mechanisms involved in solving this Paradigm Cell Filling Problem (1, 2) and related issues (3, 4) are central to the study of morphological production and comprehension as well as morphological change.

Within inflectional paradigms, as elsewhere is language, differences in form typically correspond to differences in meaning, e.g., English present tense *drive* versus past tense *drove*.

In the case of allomorphy, however, differences in form are semantically moot. Sometimes, allomorphs are phonologically predictable: in certain Sino-Tibetan languages of Nepal, the verb stem can be different depending on whether the following suffix starts with a consonant or with a vowel, e.g., Limbu *nuːŋ-ʔɛ* 'I return' vs. *nuːks-ige* 'we return' (5, 6); similarly, in Western Armenian, definiteness is marked on vowel-final nouns with the suffix -*n*, and on consonant-final nouns with a formally highly differentiated, suppletive counterpart -*ə* (7:252). Some other times, allomorphy is unpredictable (sometimes called 'morphomic' or 'eidemic') (8, 9), which results in irregular paradigms for which the distribution of allomorphs must simply be memorized by language users.

It is worth noting that irregularity is a poorly defined phenomenon (10, 11) and that some examples of phonologically motivated allomorphy can be classified as irregular because they are uncommon or no longer productive. In this paper, we use the term to refer to Romance verb stem alternations, which are morphomic, and comparatively uncommon and unproductive compared to the absence of alternation.

Irregular paradigms in general and unpredictable allomorphy in particular are central to a number of questions regarding the various pressures that drive language change. The analogical processes involved in paradigm evolution are a central object of study for diachronic linguistics. Allomorphic variation introduces many-to-one mappings between form and meaning, whereas form-meaning isomorphy is traditionally held to facilitate language learning and communication (12, 13, 14, 15, 16); hence, changes affecting morphological paradigms are predicted to involve the regularization of allomorphy (known as leveling) to an overwhelming degree. Other research argues that allomorphy is in fact beneficial for communication (in particular comprehension), as it often aids in teasing apart the meanings of competing inflected forms (17, 18, 2, 19); a listener hearing the English words *go* and *went* will be able to narrow down the set of tenses compatible with each form at a very early point in time. It has been argued that the relative frequency with which regularizing changes such as paradigm leveling occur is an epiphenomenon of the fact that leveled paradigms tend to be more frequent in the first place, and that such changes are a mere extension of a frequent pattern and involve no drive towards form-meaning isomorphy (20), and in a similar vein, that there is no inherent drive towards irregulization in morphological change (21). A large number of studies (e.g., 22) report that, across languages, irregular forms tend to be more frequent, and less frequent forms tend to be regular. However, many aspects of the way in which frequency impacts morphological change remain at issue.

In this paper, we explore the drivers of morphological change in Romance languages, an Indo-European subgroup where patterns of inflectional allomorphy are complex and diverse. We focus on changes in verb stem allomorphy, also known as 'stem alternation', a phenomenon that has been the object of abundant qualitative research (e.g., 23, 24, 25, 26, 27, 28, 29, 30) but has thus far not enjoyed much in the way of quantitative analysis across the family. We do this through a novel phylogenetic methodology inspired by recent advances in hierarchical Bayesian modeling (31) as well as existing work on evolutionary rate variation in biology and linguistics (32, 33, 34, 35). Our approach makes it possible to assess the effect of predictor variables on different components of language change, namely speed of change

(irrespective of direction, representing the overall volatility of a system) and stationary probability (interpretable as the long-term preference for a feature), and allows us to control for lineage-level trends that could potentially be conflated with relationships of interest between different variables. One or more of these components of change may underlie the distribution of irregular forms, and are arguably rooted in pressures involved in speech planning and production (36, 37) and/or listener-oriented considerations in language use (38, 39, 40, 41), as well as dynamics of lexical replacement (33).

We find that while overall speeds of change vary across lexemes, this variation is not explained by frequency of usage. Instead, we find that frequency has an effect on the stationary probability of an irregular inflectional pattern in a lexeme; more frequent lexemes are likely to display a long-term preference for irregularity, while less frequent ones are likely to display a long-term preference for regularity. We find also pronounced variation in speeds of change and preferences for regularity among different Romance lineages, indicating that sociolinguistic factors probably play a role in shaping the synchronic distribution of stem allomorphy across Romance languages. We conclude by outlining future directions for research into Romance stem allomorphy and for development of the method we employ.

2 Background

2.1 Regularization and irregularization and morphological change

The distribution of irregular forms within languages is a much-discussed issue in linguistics. There is widespread consensus that frequency and irregularity tend to be positively correlated (42, 43, 44, 45, 46). More frequent words tend to be more irregular than less frequent ones, or stated conversely, irregular forms tend to be more frequent than the average regular form. This is generally believed to be because irregular forms must be deeply entrenched in use in order to survive across generations (43:10). Hence, less frequent irregular forms should be more prone to regularization, which is what preserves the correlation.

More recent research has explored the evolutionary dynamics of this relationship, asking whether the lexical distribution of irregularity is such that it maximizes its diachronic stability (47), or whether regularity or irregularity are selected for in the course of language evolution or are mostly due to stochastic drift (48). An aspect which is of great importance for the evolution of frequency and irregularity is whether less frequent words are less stable (i.e. change faster) than more frequent ones. **(author?)** (49) show that low-frequency English irregular verbs are more prone to regularization. Trends of this sort may exist because they serve to keep high-frequency verbs irregular, which would keep higher-frequency forms more discriminable in processing through, for example, earlier uniqueness points (50, 51, 52).

At the same time, the association of frequency with irregularity could be the result of the greater stability of high frequency items, as is claimed for the dynamics of lexical replacement (33). This would lead to higher entrenchment of any inherited irregularity in high-frequency verbs, while low-frequency verbs would be more prone to change in general, in any direction, including regularization.

Despite a growing body of research on the topic, the role of frequency in morphological change is still not fully understood. Most studies reporting an effect of frequency on regularization or analogical change more generally are restricted to single languages. Furthermore, some studies do not find that frequent forms are less likely to undergo analogical change (53), and some Romance languages have been argued to not even display a correlation between frequency and irregularity (54).

The Romance subgroup is a particularly intriguing natural laboratory for the study of the effect of frequency on analogical change generally and regularization in particular. Irregular stem alternations emerged through regular sound change in most Romance verbs. At the same time, such alternations are often found where they are not historically expected, or are absent from where they would be expected. This implies that abundant change must have taken place. Specialized literature (e.g., 29), however, rarely mentions frequency as a factor predicting which verbs lose or gain stem alternations. The Romance languages are also convenient to study because they are generally well documented, both synchronically (55, 56), and diachronically (i.e., through Latin). Here we use phylogenetic comparative methods to determine more precisely what role frequency has played in the evolution of Romance verb stem allomorphy.

2.2   Phylogenetic modeling

A number of quantitative tools, many of them borrowed from computational biology, have made inroads into diachronic linguistic research in recent decades. A common and highly visible use of phylogenetic methods in linguistics has involved inferring phylogenies of language families (57, 58, 59). Phylogenetic comparative methods (PCMs) are a family of methods that model the evolutionary dynamics of various biological features. Many of these methods can be directly imported from biology applied to their linguistic analogs. In general, these methods rely on an existing phylogeny of the organisms or languages under study, on which dynamics of change are conditioned.

PCMs often assume that the evolution of categorical linguistic data undergo state changes over a phylogeny according to a continuous-time Markov (CTM) process. The transition rates of the process can be estimated, usually via Bayesian inference, and estimated values can be used to test hypotheses regarding the dynamics of change of the features in question. Under the standard view of a continuous-time Markov process for binary data, a feature arises and is lost according to a gain rate and a loss rate. Assuming a speed of change s and stationary probability π, the gain rate and loss rate can be rewritten as $s\pi$ and $s(1 - \pi)$, respectively (see Figure 1). This is the binary case of a general time-reversible model (60), which parameterizes changes between multiple states in a continuous-time Markov chain according to stationary probabilities of state presence and exchange rates or speeds of change between each pair of states. Simulated trajectories of change under binary CTM processes with different speeds and stationary probabilities are found in Figure 2.

This simple model has been expanded in a number of ways in order to allow rates of change to vary both across features and across lineages or branches of a phylogeny. Most of these are models of heterotachy, which allow speeds of change to vary (61, 32, 62), with
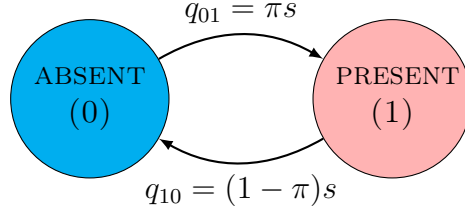
Figure 1: Binary continuous-time Markov process representing changes between presence and absence of a feature. Transitions are annotated according to the gain rate (q_{01}) and loss rate (q_{10}) of the feature, with alternative parameterizations according to speed of change (s) and stationary probability of feature presence (π) provided as well.
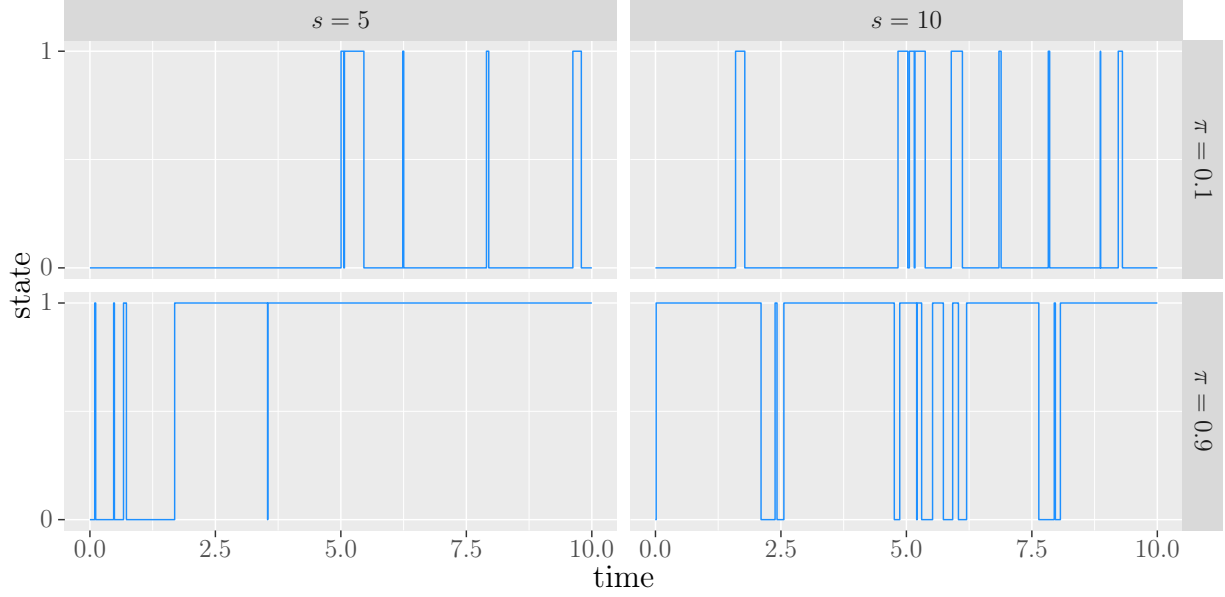

Figure 2: Simulated CTM processes showing transitions between states of a binary feature under different speeds ($s \in \{5, 10\}$) and stationary probabilities ($\pi \in \{0.1, 0.9\}$).

Table 1: Partial paradigms of Spanish *perder* 'lose' and *decir* 'say'

	N			L		
	Pres. ind.	Pres. sbjv.	Ipf. ind.	Pres. ind.	Pres. sbjv.	Ipf. ind.
1SG	pierd-o	pierd-a	perd-ía	dig-o	dig-a	dec-ía
2SG	pierd-es	pierd-as	perd-ías	dic-es	dig-as	dec-ías
3SG	pierd-e	pierd-a	perd-ía	dic-e	dig-a	dec-ía
1PL	perd-emos	perd-amos	perd-íamos	dec-imos	dig-amos	dec-íamos
2PL	perd-éis	perd-áis	perd-íais	dec-ís	dig-áis	dec-íais
3PL	pierd-en	pierd-an	perd-ían	dic-en	dig-an	dec-ían

a minority of models accounting for variation not only in speed but in long-term biases in change (63, 64, 65, 66). We build upon models of rate variation in order to explore the effect of frequency and other predictors on the dynamics of morphological change of different Romance verbs while controlling for branch-specific trends.

3 Data

Romance philological literature (e.g., 29) identifies three main types of stem alternations in verbs, referred to by the arbitrary labels N, L, and P(YTA), which can co-occur within verbal paradigms. The emergence of the N and L patterns (see Table 1) occurred as a result of sound changes after Classical Latin but (largely) before the break-up of Romance into different languages.

Thus, the alternation between *e* and *ie* in the stem of Spanish *perder* is due to different trajectories of change affecting stressed syllables (leading to *ie*) and unstressed syllables (leading to *e*). The alternation between *g* and *c* (phonetically [g] and [θ], respectively) in the stem of Spanish *decir* results in turn from the palatalization of Latin *c* ([k]) before the vowels *e* and *i*.

Unlike the other two patterns, the P pattern (e.g. a vowel *i* vs *a* in the stem of Spanish *hacer* is inherited from Classical Latin (see Table 2), where it used to mark a semantic distinction that is no longer present in modern Romance languages, leading to alternations that are nowadays arbitrary from the perspective of meaning. While the presence of stem alternations in these lexemes is expected given regular sound changes and the shape of the ancestral Latin paradigms, other lexemes have lost or acquired their alternations in analogy with other verbs (e.g. *caigo, caes* ← *cadō, cadis* 'fall' vs. expected †*cao, caes*).

The Oxford Online Database of Romance Verb Morphology (67, 56) contains information regarding verb paradigms in a large number of Romance speech varieties (see Figure 3 for the languages included in this study). Paradigms are given for each lemma in each language; each lemma is also coded according to the Latin etymon from which it descends, if inherited from Latin. Among Latin-descended lemmata, 231 cognate sets are found in the database. For each lemma in each variety, we manually assessed whether it contained a reflex of the stem alternation patterns (N, L, and P) described above. We gloss over language-specific

Table 2: Partial paradigms of Spanish *hacer*, and Latin *faciō* 'do' illustrating P

|	Spanish			Latin		
	Ipf. ind.	Pret. ind.	Ipf. sbjv.	Ipf. ind.	Perf. ind.	Plup. sbjv.
1SG	hac-ía	hic-e	hic-iese	fac-iēbam	fēc-ī	fēc-issem
2SG	hac-ías	hic-iste	hic-ieses	fac-iēbās	fēc-istī	fēc-issēs
3SG	hac-ía	hi[θ]-o	hic-iese	fac-iēbat	fēc-it	fēc-isset
1PL	hac-íamos	hic-imos	hic-iesemos	fac-iēbāmus	fēc-imus	fēc-issēmus
2PL	hac-íais	hic-isteis	hic-ieseis	fac-iēbātis	fēc-istis	fēc-issētis
3PL	hac-ían	hic-ieron	hic-iesen	fac-iēbant	fēc-ērunt	fēc-issent
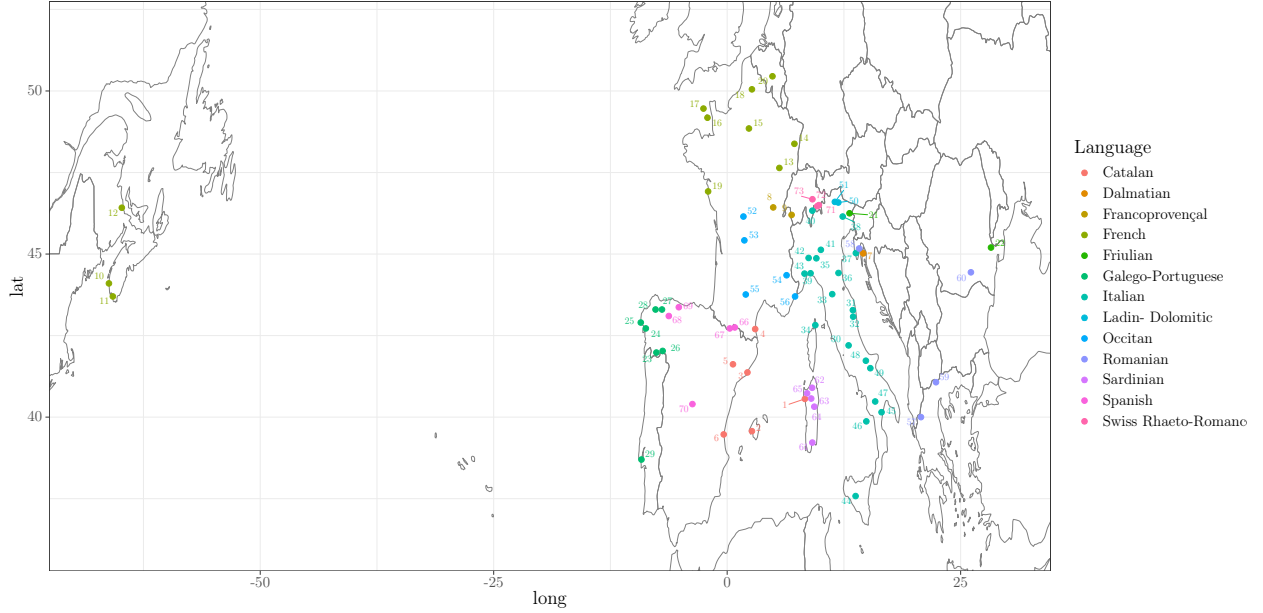
Figure 3: Speech varieties in our sample, plotted according to the locations in which they are spoken. Speech varieties denoted by numerical identifiers can be found in Table 3.

idiosyncrasies in the inherited distribution of alternants in the paradigm for the purposes of this study (e.g., the L pattern affects the 3PL present indicative of Italian *dire* 'say' but is absent from the 1–2PL present subjunctive, in contrast to its Spanish cognate in Table 1), coding patterns as PRESENT/ABSENT/MISSING DATA. We target items with wide coverage, excluding cognate sets that are limited to dialects of a single language or attested in fewer than five varieties. In total, we analyze 171 lemma-pattern pairs involving 67 different lemmas from 66 speech varieties.

We collect data for a number of additional variables. We take the frequency for each lemma in Latin texts (68). We also note the conjugation class of each Latin lemma. The stem of each Latin lemma is converted to a broad phonemic transcription (including characters indicating stem boundaries) and trigram phone sequences are extracted. We exclude trigram phone sequences whose presence within a word is entirely predictable on the basis of preceding

LID	Speech Variety
1	CatalanEasternAlgueresfromLAlgherAlghero
2	CatalanEasternBalearMallorquifromPalma
3	CatalanEasternCentralBarcelonifromBarcelona
4	CatalanEasternRossellonesfromCanetdeRosselloCanet-en-Roussillon
5	CatalanWesternfromLleida
6	CatalanWesternfromValencia
7	DalmatianVegliotefromVegliaKrk
8	FrancoprovencalLyonnaisfromVaux
9	FrancoprovencalValaisanValdIlliezfromValdIlliez
10	FrenchAcadianfromBaieSainte-Marie
11	FrenchAcadianfromPubnico
12	FrenchAcadianSouth-EastNewBrunswickfromMonctonenvirons
13	FrenchFranc-ComtoisfromPierrecourt
14	FrenchLorrainfromRanrupt
15	FrenchModernStandard
16	FrenchNormanfromJersey
17	FrenchNormanGuernsey
18	FrenchPicardfromMesnil-Martinsart
19	FrenchPoitevin-SaintongeaisVendeenfromBeauvoir-sur-Mer
20	FrenchWallonfromNamur
21	Friulian
22	FriulianWesternManiagofromGreci
23	Galego-PortugueseGalicianfromCualedro
24	Galego-PortugueseGalicianfromDodro
25	Galego-PortugueseGalicianfromFisterraFinisterra
26	Galego-PortugueseGalicianfromLubian
27	Galego-PortugueseGalicianfromVilanovadeOscos
28	Galego-PortugueseGalicianfromXermade
29	Galego-PortuguesePortuguese
30	ItalianCentralLazialeNorth-centralfromAscrea
31	ItalianCentralMarchigianofromMacerata
32	ItalianCentralMarchigianofromServigliano
33	ItalianCentralModernStandard
34	ItalianCentralTuscanCorsicanfromSisco
35	ItalianNorthernIEmilianfromTravo
36	ItalianNorthernIEmilianRomagnolfromLugo
37	ItalianNorthernIIVenetoIstriotoValledIstriafromValledIstria

LID	Speech Variety
38	ItalianNorthernIIVenetoNorthernfromAlpago
39	ItalianNorthernILigurianGenoesefromGenova
40	ItalianNorthernILombardAlpinefromValCalanca
41	ItalianNorthernILombardCremonesefromCremona
42	ItalianNorthernIPiedmonteseBassofromCascinagrossa
43	ItalianNorthernIPiedmontesefromCairoMontenotte
44	ItalianSouthernIISicilianCentralfromMussomeli
45	ItalianSouthernILucanoArchaicfromNovaSiri
46	ItalianSouthernILucanoCalabriafromPapasidero
47	ItalianSouthernILucanoCentralfromCalvello
48	ItalianSouthernIMolisanofromCasacalenda
49	ItalianSouthernIPuglieseDaunofromLucera
50	Ladin-DolomiticAtesinoValBadia
51	Ladin-DolomiticAtesinoValGardena
52	OccitanNorthernAuvergnatGartempefromGartempeCreuse
53	OccitanNorthernLimousinfromSaint-Augustin
54	OccitanNorthernVivaro-AlpinfromSeyne
55	OccitanSouthernLanguedocienfromGraulhet
56	OccitanSouthernProvencalfromNice
57	RomanianAromanian
58	RomanianIstro-RomanianfromSusnjevicaValdarsa
59	RomanianMegleno-Romanian
60	RomanianModernStandard
61	SardinianCampidanese
62	SardinianGallurese
63	SardinianLogudorese
64	SardinianNuorese
65	SardinianSassarese
66	SpanishAragoneseAnsotanofromAnsoFago
67	SpanishAragonesefromPanticosa
68	SpanishAsturo-LeoneseAsturianfromSomiedo
69	SpanishAsturo-LeoneseAsturianSudestedefromParres
70	SpanishModernStandard
71	SwissRaeto-RomancePuterUpperEngadine
72	SwissRaeto-RomanceSurmiranfromBivio-Stalla
73	SwissRaeto-RomanceSurselvan

Table 3: Speech varieties in sample

trigrams. As an example, the trigrams *#bi*, *bib*, and *ib#* are restricted within our data set to the stem of the lemma ***bib****ere* 'to drink'. Accordingly, these trigrams predict the presence of each other within a stem with total certainty. In order to reduce the risk of multicollinearity in the models we describe below, we retain only *#bi*, which effectively stands in for the longer sequence *#bib#*. Additionally, we generate word embeddings for the 3rd person singular present indicative form of each Latin lemma from LatinBERT, a transformer-based language model that generates contextualized and uncontextualized latent representations (69), representing the semantic profile of each Latin verb (Figure 4 displays lemmas in embedding space).

3.1 Tree

The models we employ require a phylogenetic representation of the languages in our sample, in the form of a tree sample. Tree samples for related languages are generally generated on the basis of lexical cognacy data using Bayesian inference, which allows for the use of various priors (e.g., over tree topologies), as well as clade constraints and calibration points for dating. In order to infer a tree for the Romance languages in our sample, we first match each languoid in the Oxford Romance Database to its closest correspondent in the Automated Similarity Judgment Program (70) and the Sound Comparisons database (71). We use lexical cognacy characters automatically extracted from the 40-word ASJP wordlist (72). Because there is little variation among Romance languages in terms of the cognacy of basic vocabulary items, we also use phonological characters extracted from the Sound Comparisons database, which provides narrow phonetic transcriptions for basic vocabulary items in a number of Romance speech varieties; we code languages according to whether or not a given segment is present across their basic vocabulary. These features serve to capture properties of lower-level diversification among the languages in our sample. Because phonological characters of this sort have the potential to represent areal rather than genetic signal, we impose uncontroversial clade constraints on the tree topology along with lower and upper bounds for each clade's date calibration (73), as shown in Table 4.

We infer the phylogeny using RevBayes (74). We use a Birth-Death tree prior (75) and a General Time-Reversible model of character evolution (60). We use a relaxed clock with log-normally distributed branch-level rate multipliers and gamma-distributed variation across rate classes; we fix the number of rate classes at 4. We allow for separate rate multipliers and rate classes for the two data types we use (lexical characters and phonological characters). All priors are found in Table 5. We run 1,000,000 iterations of Markov chain Monte Carlo over 4 chains, discarding the first half of samples as burn-in. We monitor convergence by comparing the log posteriors of the chains. We concatenate the samples from each chain and thin the resulting set of trees to 50 trees. A DensiTree visualization (76) of the tree sample prior to thinning can be seen in Figure 5.
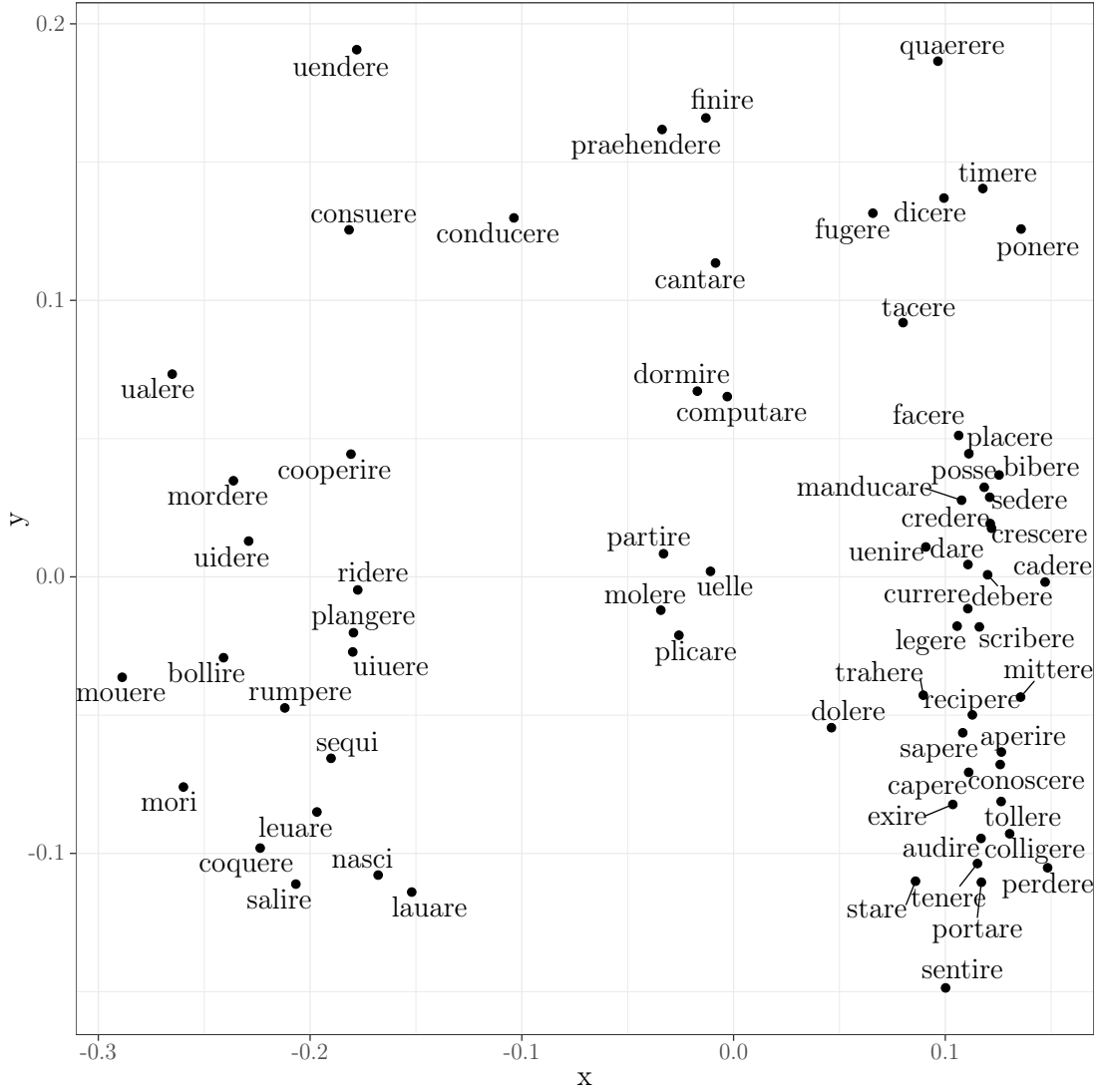
9

Figure 4: Lexemes in sample plotted according to LatinBERT word embeddings for the 3rd person singular present indicative form. Cosine distances between lexemes are scaled to two dimensions using multidimensional scaling.

Clade	Max. age	Min. age
Latin	2600	2400
Proto-Romance	2150	1850
Sardinian	1350	1050
Balkan Romance	1750	1350
Romanian	1450	1250
Italo-Western Romance	1800	1650
Italian Romance	1650	1250
Rhaetic	1450	1150
Ladin	1050	850
Friulian	1050	850
Swiss	1050	850
Gallo-Romance	1550	1450
Languedoc	1450	1450
Franco-Provençal	1250	1050
Occitan	1250	1050
Catalan	1050	950
Spanish	1050	950
Ibero-Romance	1650	1250
Western Ibero-Romance	1250	1050

Table 4: Clade constraints and date ranges imposed upon the inferred phylogeny

Parameter	Intepretation	Prior
Birth-death tree prior		
λ	Speciation rate	LogNormal$(-7, 10)$
μ	Extinction rate	LogNormal$(-7, 10)$
ρ	Taxon sampling fraction	Beta$(1, 1)$
rootAge	Root age	Unif$(2400, 2600)$
General Time-Reversible model		
π_0, π_1	Stationary frequencies	Dirichlet$(1, 1)$
r_{01}, r_{10}	Exchange rates	Dirichlet$(1, 1)$
Rate variation		
$\boldsymbol{\tau}$	Branch rate multiplier	$\tau_{ij} \sim \mathrm{Exp}(1/\lambda_j^\tau) : i \in \{1, ..., 2N - 2\}, j \in \{1, 2\}$ $\lambda_j^\tau \sim \mathrm{LogNormal}(\mu^\tau, \sigma) : j \in \{1, 2\}$ $\mu^\tau \sim \mathrm{Normal}(-7, 10), \sigma \sim \mathrm{Exp}(1)$
$\boldsymbol{\kappa}$	Rate class multiplier	$\kappa_{kj} \sim \mathrm{Gamma}(\alpha_j, \alpha_j) : k \in \{1, .., 4\}, j \in \{1, 2\}$ $\alpha_j \sim \mathrm{LogNormal}(\ln 5, 0.587405)$

Table 5: Priors used for phylogenetic inference with RevBayes

Dalmatian Vegliote from Veglia Krk
Romanian Istro-Romanian from Susnjevica Valdarsa
Romanian Aromanian
Romanian Megleno-Romanian
Romanian Modern-Standard
Sardinian Campidanese
Sardinian Logudorese
Sardinian Nuorese
Sardinian Gallurese
Sardinian Sassarese
Friulian Western Maniago from Greci
Friulian
Ladin- Dolomitic Atesino Val Badia
Ladin- Dolomitic Atesino Val Gardena
Swiss Raeto-Romance Puter Upper Engadine
Swiss Raeto-Romance Sursilvan
Swiss Raeto-Romance Surmiran from Bivio-Stalla
Italian Northern II Venetic Istrioto Valle distria from Valle dIstria
Italian Central Marchigiano from Servigliano
Italian Central Laziale North-central from Ascrea
Italian Central Marchigiano from Macerata
Italian Northern II Ligurian Genoese from Genova
Italian Northern I Piedmontese from Cairo Montenotte
Italian Northern II Veneto Northern from Alpago
Italian Northern I Piemontese Basso from Castelnagrossa
Italian Northern I Emilian from Lugo
Italian Northern II Emilia Romagna from Lugo
Italian Northern I Lombard Cremonese from Cremona
Italian Northern I Lombard Alpine from Val Calanca
Italian Central Modern Standard
Italian Central Tuscan Corsican from Sisco
Italian Southern II Lucano Central from Mussomeli
Italian Southern I Lucano Calabria from Papasidero
Italian Southern I Pugliese from Casamicciola
Italian Southern Pugliese Dauno from Lucera
Italian Southern I Lucano Central from Calvello
Italian Southern I Lucano Archaic from Nova Siri
Spanish Asturo-Leonese Asturian from Somiedo
Spanish Asturo-Leonese Asturian Sudeste de from Parres
Spanish Aragonese from Panticosa
Spanish Modern Standard
Spanish Aragonese Ansotano from Anso Fago
Galego-Portuguese Portuguese
Galego-Portuguese Galician from Vilanova de Oscos
Galego-Portuguese Galician from Lubian
Galego-Portuguese Galician from Dodro
Galego-Portuguese Galician from Xermade
Galego-Portuguese Galician from Fisterra Finisterra
Galego-Portuguese Galician from Coaledro
Catalan Eastern Alguares from L'Algher Alghero
Catalan Eastern Balear Mallorqui from Palma
Catalan Western from Valencia
Catalan Eastern Rossellones from Canet de Rossello Canet-en-Roussillon
Catalan Eastern Central Barceloni from Barcelona
Catalan Western from Lleida
Occitan Northern Vivaro-Alpin from Vaux
Francoprovençal Valaisan Val dIlliez from Val dIlliez
Occitan Southern Languedocien from Grauhet
Occitan Southern Provençal from Nice
Occitan Northern Vivaro-Alpin from Seyne
Occitan Northern Auvergnat Gartempe from Gartempe Creuse
Occitan Northern Limousin from Saint-Augustin
French Wallon from Namur
French Poitevin-Saintongeais Vendeen from Beauvoir-sur-Mer
French Lorrain from Rantrupt
French Modern Standard
French Norman from Jersey
French Norman Guernsey
French Picard from Mesnil-Martinsart
French Franc-Comtois from Pierrecourt
French Acadian from Pubnico
French Acadian South-East New Brunswick from Moncton environs
French Acadian from Baie Sainte-Marie

3000   2500   2000   1500   1000   500   0

Figure 5: Thinned tree sample for languages in data set

4 Method

We use phylogenetic comparative methods to explore the temporal dynamics of stem alternation patterns in reflexes or descendants of Latin verbs. We treat each pattern (i.e. N, L, and P) in each reflex as a binary trait that can express the values PRESENT, ABSENT, or MISSING DATA. The dilemma posed by missing data is analogous to the well-known "red tail/blue tail/no tail" problem in biology (77). Solutions to this problem include the standard approach to missing data in phylogenetic models, namely setting the likelihoods of both presence and absence to 1 (78:255), or using a three-state character state encoding the states PRESENT, ABSENT, and ENVIRONMENT ABSENT. We adopt the former approach, since less-studied speech varieties may be less thoroughly etymologized when compared to better known languages, and the absence of a reflex in the database is often due to incomplete data coverage rather than the actual absence of a reflex in a given language.

Each binary trait evolves according to a CTM process, parametrized in terms of the overall SPEED of change, irrespective of the direction, and the STATIONARY PROBABILITY of trait presence, interpretable as the long-term preference for the trait — i.e., the proportion of time it will have been observed, as time approaches infinity (cf. Figure 1). Inspired by HIERARCHICAL DISTRIBUTIONAL MODELS from Bayesian statistics which treat both the expected location and scale in a regression model as a function of one or more predictors (31), we model the speed and stationary probability of each binary trait according to combinations of predictors. Under our model, the gain and loss rate of a lemma-pattern pair with index $d \in \{1, ..., D\}$ are $\pi_d s_d \rho$ and $(1 - \pi_d) s_d \rho$. Here, $s_d \rho$ represents the speed of change for feature d, s_d being a multiplier of the global speed ρ. The parameter π_d is the stationary probability for the lemma-pattern pair in question. Speed multipliers $\boldsymbol{s}$ and stationary probabilities $\boldsymbol{\pi}$ are logit-normally distributed, which constrains values to fall within the interval $(0, 1)$:

$$\text{logit } s_d \sim \text{Normal}(\alpha_s + \boldsymbol{\beta}_s^\top \boldsymbol{X}_s, \sigma_s); \tag{1}$$

$$\text{logit } \pi_d \sim \text{Normal}(\alpha_\pi + \boldsymbol{\beta}_\pi^\top \boldsymbol{X}_\pi, \sigma_\pi); \tag{2}$$

Above, α represents intercepts, $\boldsymbol{\beta}^\top \boldsymbol{X}$ represents model predictors and associated coefficients, and σ represents the dispersion in speed and stationary probability that are not explained by the predictors included.

The likelihood of the gain and loss rates for feature d, $\pi_d s_d \rho$ and $(1 - \pi_d) s_d \rho$ (or vectors of gain and loss rates, in cases where they vary from branch to branch) under a phylogeny Ψ, $P(\boldsymbol{x}_d | \pi_d, s_d, \rho, \Psi)$, can be calculated using Felsenstein's Pruning algorithm (78). The joint model probability under a single phylogeny is the following:

$$P(\boldsymbol{X}, \boldsymbol{\pi}, \boldsymbol{s}, \rho | \Psi) = \prod_{d=1}^{D} P(\boldsymbol{x}_d | \pi_d, s_d, \rho, \Psi) P(\boldsymbol{\pi}) P(\boldsymbol{s}) P(\rho)$$

Phylogenetic uncertainty can be integrated out by aggregating posterior parameters across a sample of trees $\mathbb{T}$:

$$\int P(\boldsymbol{X}, \boldsymbol{\pi}, \boldsymbol{s}, \rho | \Psi) P(\Psi) d\Psi \approx \frac{1}{|\mathbb{T}|} \sum_{\Psi \in \mathbb{T}} P(\boldsymbol{X}, \boldsymbol{\pi}, \boldsymbol{s}, \rho | \Psi)$$

We fit a number of models to assess the effects of different predictors (and interactions of predictors) on each component of change. We take into account the following variables:

- **Lemma frequency**: We model the effect of lemma frequency as a categorical predictor with levels that increase monotonically according to each lemma's frequency in Latin texts. This parameterization involves a combination of a coefficient representing the effect of moving from the lowest to the highest frequency and a simplex parameter representing the effect of moving along the cline of frequency (79). This monotonic function is denoted by $mo(\textsc{lemma.freq})$.

- **Pattern type**: We model the effect of pattern type with two dummy-coded factors, comparing the levels L and P to N. This predictor has the effect of capturing differences in speed and long-term preference for each pattern type across all possible lemma-pattern combinations in the dataset.

- **Conjugation class**: We model the effect of conjugation class according to three dummy-coded factors, comparing the levels 2 (verbs ending in $-\bar{e}re$), 3 (verbs ending in $-ere$), and 4 (verbs ending in $-\bar{i}re$) to 1 (verbs ending in $-\bar{a}re$). We capture differences in speed and long-term preference for each pattern type via main effects; we capture long-term preferences on the part of a given conjugation class for specific pattern types by modeling interactions between conjugation class and pattern type.

- **Phonological similarity**: We model the effects of phonological profiles of lemmata on response variables as a linear function of unique trigrams found in each lemma's Latin stem — i.e., as a sum of coefficients associated with each trigram present in a given stem. Concerning this variable, we are interested primarily on the effect that phonological properties of lemmata have on long-term preferences for the three pattern types (N, L and P), rather than speed or a more general long-term preference for irregularity.

- **Semantic similarity**: We model the contribution of semantic similarity via a Gaussian process with a squared exponential kernel function of the cosine distance between word embeddings generated for the 3SG.PRES form of each Latin lemma in the data set using LatinBERT (69). Under a Gaussian process, the relative contribution to a response variable of a predictor is distributed according to a multivariate normal distribution with a mean of zero and covariance that is a function of distances between predictor values for each pair of data points. This has the effect of capturing non-linear and non-monotonic relationships between predictors and responses. Under a squared-exponential covariance kernel function, lexemes i and j have the following covariance, where α is a positive dispersion parameter, λ is the length scale (a positive parameter indicating the cutoff in distance at which two data points do not influence each other), and $\delta_{i,j}$ is the cosine distance between the embeddings for lexemes i and j:

$$\Sigma_{i,j} = \alpha \exp\left(-.5\frac{\delta_{i,j}^2}{\lambda}\right)$$

A positive parameter σ is added to diagonal cells of the variance-covariance matrix, which ensures that the matrix is positive semi-definite. Generally speaking, large length scales indicate that a predictor variable lacks explanatory power: if λ is large, then $\Sigma_{i,j}$ will be very close to α regardless of the magnitude of $\delta_{i,j}$, rather than exhibiting high covariance for proximate values and low covariance for more distant values.

We assume that the coefficients representing each lemma's propensity towards each pattern type (N, L, and P) are generated by three separate Gaussian Processes. We place HalfNormal$(0, 1)$ priors over all GP parameters.

We model the effects of these predictors on speed of change and/or long-term preference for different stem alternation types over the entire phylogeny of Romance. In addition to these overall effects, we can also allow the overall speed of change and stationary probability of presence of any irregular pattern to vary from branch to branch, analogous to varying intercepts in hierarchical regression models, and can also allow the effects of different predictors on different components of change to vary in the same fashion, analogous to random slopes. Our rationale here is to capture trends displayed by individual lineages in the tree, including variation in speed of change as well as general branch-level preferences or dispreferences for irregular morphological patterns, effectively accounting for branch-specific idiosyncrasies that might point to contact or events coinciding with other linguistic changes (e.g., loss of stress in Gallo-Romance, the conserving effects of standardization in national standard varieties, etc.).

We place a Uniform$(0, 10)$ prior over the global speed of change ρ, preventing changes from happening more frequently than ten times per millennium. We place Dirichlet$(1, ..., 1)$ priors over simplex parameters involved in monotonic predictors. We place Normal$(0, 1)$ priors on all unconstrained parameters (coefficients associated with predictors and involved in monotonic predictors) and HalfNormal$(0, 1)$ priors over parameters constrained to be positive (standard deviations along with dispersion parameters and length scales of Gaussian Process terms).

We assess the impact of the predictors enumerated above on different components of change via model comparison. We cannot assess the effect of all possible subsets of predictors on each component of change given the computational infeasibility of comparing such a large number of models. We instead fit a series of models in an incremental fashion, as specified below. Our principle is to start with a relatively simple model and build in additional predictors retaining them if they improve model fit, as measured by Pareto-smoothed importance sampling leave-one-out cross-validation (PSIS-LOO-CV, (80), see also below). To reduce the number of models to compare, we choose not to model the effects of predictors on components of change if there is not a good rationale for hypothesizing such a relationship; e.g., there may be an association between a particular phonological profile and a given stem alternation pattern, but not necessarily the speed of change. We fit each model with the CmdStanR interface (81) to the Stan programming language (82) using 50 trees from the tree sample described above. For each tree, we run four chains of 1000 each of the No-U-Turn sampler with a thinning interval of 10, and aggregate posterior samples across trees. Code accompanying this paper can be found at `https://osf.io/yxhzr/?view_only=ceb404ac0aaf4b72b14bb68084c1b27f`.

Models under comparison in this study are found below. Variables include lemma frequency

(LF), pattern type (PT), conjugation class (CC), and branch (B). $\Phi \times$ PT denotes a linear predictor indicating the association between a lexeme's phonological profile and a given pattern type, comprising coefficients for each unique trigram in our dataset and each pattern type; $f_{\mathrm{GP}}(\Sigma; \mathrm{PT})$ denotes a smooth term based on semantic properties of each lexeme (i.e., based on cosine distances between word embeddings), associated with different pattern types.

- `lemma+pattern+conj+pattern×conj`

  $s \sim mo(\mathrm{LF}) + \mathrm{PT} + \mathrm{CC}$

  $\pi \sim mo(\mathrm{LF}) + \mathrm{PT} + \mathrm{CC} + \mathrm{PT} \times \mathrm{CC}$

  Speed varies according to lemma frequency, pattern type, and conjugation class. Long-term preference varies according to lemma frequency, pattern type, conjugation class and interactions between pattern type and conjugation class.

- `vary_branch+lemma+pattern+conj+pattern×conj`

  $s \sim mo(\mathrm{LF}) + \mathrm{PT} + \mathrm{CC} + (1|\mathrm{B})$

  $\pi \sim mo(\mathrm{LF}) + \mathrm{PT} + \mathrm{CC} + \mathrm{PT} \times \mathrm{CC} + (1|\mathrm{B})$

  Speed varies according to lemma frequency, pattern type, and conjugation class. Long-term preference varies according to lemma frequency, pattern type, conjugation class and interactions between pattern type and conjugation class. Overall speed and long-term preference for irregularity vary at the branch level.

- `vary_branch+lemma+pattern+conj+pattern×conj+phon+sem`

  $s \sim mo(\mathrm{LF}) + \mathrm{PT} + \mathrm{CC} + (1|\mathrm{B})$

  $\pi \sim mo(\mathrm{LF}) + \mathrm{PT} + \mathrm{CC} + \mathrm{PT} \times \mathrm{CC} + \Phi \times \mathrm{PT} + f_{\mathrm{GP}}(\Sigma; \mathrm{PT}) + (1|\mathrm{B})$

  Model structure equivalent to `vary_branch+lemma+pattern+conj+pattern×conj`, but with the addition of a linear term representing each lemma's long-term preference for a given pattern based on its phonological profile and a GP smooth term representing each lemma's long-term preference for a given pattern based on its semantic profile.

- `vary_branch+lemma+pattern+conj+pattern×conj+phon`

  $s \sim mo(\mathrm{LF}) + \mathrm{PT} + \mathrm{CC} + (1|\mathrm{B})$

  $\pi \sim mo(\mathrm{LF}) + \mathrm{PT} + \mathrm{CC} + \mathrm{PT} \times \mathrm{CC} + \Phi \times \mathrm{PT} + (1|\mathrm{B})$

  Model structure equivalent to `vary_branch+lemma+pattern+conj+pattern×conj`, but with the addition of a linear term representing each lemma's long-term preference for a given pattern based on its phonological profile.

- `vary_branch+lemma+pattern+conj+pattern×conj+sem`

$$s \sim mo(\text{LF}) + \text{PT} + \text{CC} + (1|\text{B})$$
$$\pi \sim mo(\text{LF}) + \text{PT} + \text{CC} + \text{PT} \times \text{CC} + \Phi \times \text{PT} + f_{\text{GP}}(\Sigma; \text{PT}) + (1|\text{B})$$

Model structure equivalent to `vary_branch+lemma+pattern+conj+pattern×conj`, but with the addition of a GP smooth term representing each lemma's long-term preference for a given pattern based on its semantic profile.

- `vary_branch+lemma+pattern+conj`

$$s \sim mo(\text{LF}) + \text{PT} + \text{CC} + (1|\text{B})$$
$$\pi \sim mo(\text{LF}) + \text{PT} + \text{CC} + (1|\text{B})$$

Model structure equivalent to `vary_branch+lemma+pattern+conj+pattern×conj`, but without the interaction term between pattern type and conjugation class. This model serves to determine if inclusion of the interaction term is justified.

5 Results

We compare fitted models via their leave-one-out expected log pointwise density (ELPD) values (80) and stacking (83), which averages predictive distributions of different models to generate weights representing their relative predictive power. Figure 6 gives ELPD values in decreasing order and stacking weights for each model, and plots differences in ELPD between each model and the model with the largest ELPD along with error bars representing ± 2 standard errors of each difference (the standard cutoff for determining whether the model with higher ELPD is a decisively better fit to the data). As is clear, not all absolute differences between the ELPD of the model with the largest value and other models are greater than two standard errors; the three models which include phonological and/or semantic predictors exhibit comparable ELPD values. This indicates that the combined predictions of these three models are slightly better than their individual predictions. Additionally, while a single model (`vary_branch+lemma+pattern+conj+pattern×conj+phon`) exhibits the highest stacking weight (.48), two other models (`vary_branch+lemma+pattern+conj+pattern×conj+sem` and `vary_branch+lemma+pattern+conj`) show non-negligible stacking weights as well (.23 and .273). Additionally, the stacking weights do not necessarily follow the same order as ELPD values. It is often the case that a model with a relatively high ELPD value may have stacking weights close to zero if estimates for some model coefficients are very small and the model makes similar predictions to a simpler model under comparison (84). Both ELPD and stacking weights show that there is no model improvement brought about by including phonological and semantic predictors in the same model, possibly due to a relationship between these variables. Additionally, the relatively high stacking weight of the model that excludes an interaction term modeling the effect of conjugation class on the preference for individual pattern types indicates that conjugation class is not a particularly good predictor of pattern preference (this is confirmed by an inspection of posterior parameters below). This is a surprising result, since Romance philologists have frequently observed that first conjugation -$\bar{a}re$ verbs tend to resist stem allomorphy (see e.g. (85) and Chapter 9.2 in (29)).
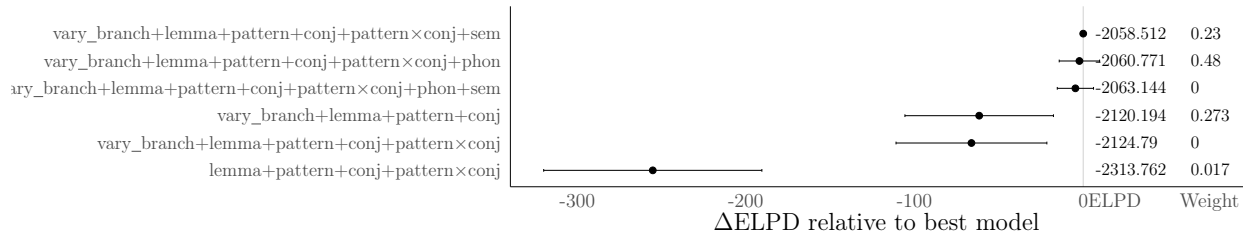
Figure 6: Differences between the expected log pointwise predictive density (ELPD) of `vary_branch`+`lemma`+`pattern`+`conj`+`pattern`×`conj`+`sem` (the best fitting model) and all remaining models. Error bars represent ±2 standard errors of each difference. Raw ELPD values and stacking weights are also provided.

In reality, semantics, phonology and conjugation class are likely interrelated, as several studies show that aspects of a word's semantics can be predicted from its phonological profile, and vice versa (86, 87), and that inflectional classes of lexical items can be predicted on the basis of their phonological profiles (88). It is possible that these predictors explain portions of the same variance in components of change because they are not truly orthogonal, and this phenomenon is reflected in lower stacking weights for certain models under comparison.

Posterior distributions for selected model parameters (for visibility, we exclude phonological and semantic predictors) are given in Figure 7, along with the proportion of samples that are greater than or less than zero (depending on whether posterior medians are positive or negative). Because multiple models display comparable performance, we sample parameters shared by all models proportional to their stacking weights rather than inspecting the parameters of the best fitting model alone. Following conservative criteria for hypothesis evaluation, we infer decisive evidence for the effect of a predictor if the 95% credible interval (representing values contained by the [0.025, 0.975] quantile range) of the corresponding parameter excludes zero, and strong evidence in cases where the 95% CI overlaps with zero but the 85% CI (representing values contained by the [0.075, 0.925] quantile range) does not.

Most model predictors do not have a decisive effect on features' speed of change (s; top of figure), though there is strong evidence that the L pattern shows a higher speed of change (representing greater instability) than the N pattern. This may be attributable to the fact that the L pattern involves paradigm cells that are less frequent in usage than the N pattern, giving it the opportunity to be gained and lost by verb paradigms more frequently than the N pattern, which is relatively stable and productive. In contrast, we see decisive and strong evidence for effects of a number of key predictors on long-term pattern preferences, i.e. the stationary probabilities π. The 95% credible interval for the parameter representing the effect of lemma frequency on stationary probability is positive, indicating that more frequent lexemes are more likely to display a stem alternation pattern in the long term and less frequent lexemes are likelier not to. Additionally, we see strong evidence that the long-term preference for N is greater than that of P, and weak evidence that the long-term preference for N is greater than that of L. The greater productivity of N relative to other patterns of stem alternation has been established in other recent publications as well (89). Turning to
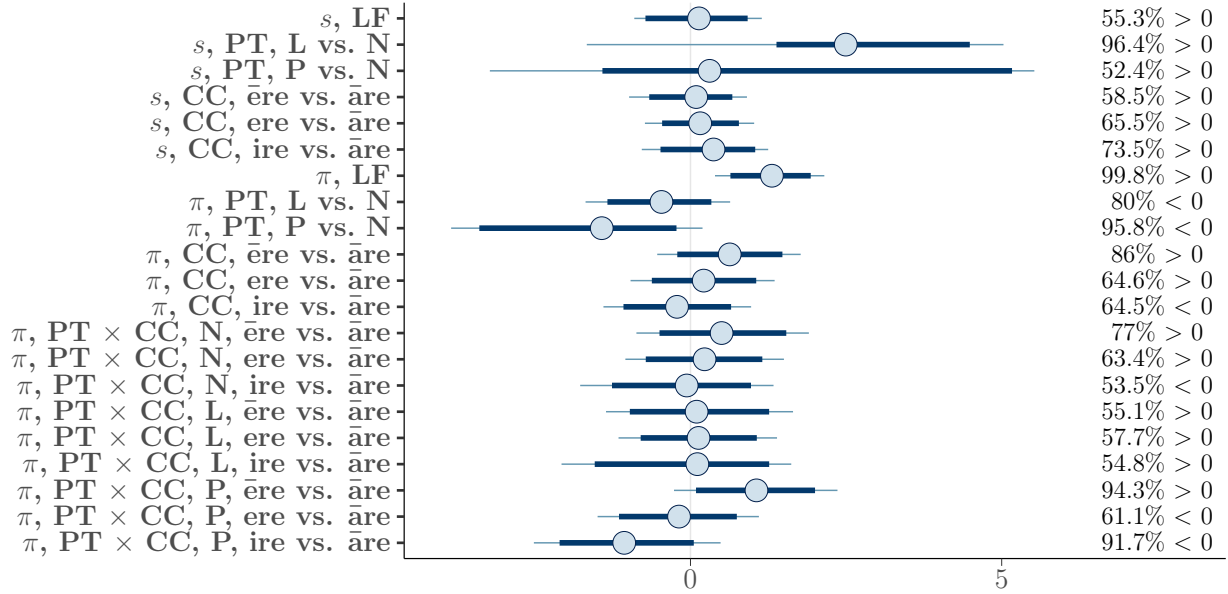
Figure 7: Medians and posterior 95% and 85% (shaded) credible intervals for model parameters of interest on the logit scale.

the effect of inflection class membership, there is strong evidence that class 2 verbs (those ending in *-ēre*) display a greater long-term preference for the P pattern than class 1 verbs (those ending in *-āre*), which in turn show a greater long-term preference for this pattern than class 4 verbs (those ending in *-īre*), in 91.7% of samples, just below the threshold for strong evidence. Conjugation, however, appears to have no effect on the proclivity towards patterns N and L. Thus, while some sources have mentioned that "the first conjugation resists and expels root allomorphy, [while] other conjugations retain and attract it" (29:277), we find no such effect when other predictors like frequency are taken into account.

5.1 Phonological profile

In addition to the predictors discussed above, we can assess the extent to which the phonological profile of each Latin lemma contributes to its long-term preference for a given pattern type. Figure 8 provides posterior distributions for coefficients representing the association between trigrams present in the stems of Latin etyma and their effect on long-term preference for different patterns, ordered according to their posterior medians (for visibility, only coefficients with the 5 highest and 5 lowest medians are given) and accompanied by the frequency of the trigram in question in our data set. Posterior samples are taken from the model `vary_branch+lemma+pattern+conj+pattern×conj+phon`.

A few patterns emerge from our results. While certain trigram coefficients for the N and L patterns show decisive differences from baseline levels (as indicated by posterior credible intervals that do not overlap with zero), this is not the case for coefficients associated with the

P pattern. This may have something to do with a general dispreference for P, with lexemes of all phonological profiles showing a trend away from this pattern. It is worth noting that even though very few coefficients have posterior credible intervals that do not overlap with zero, the inclusion of this variable as an omnibus predictor still improves predictive accuracy as there may be decisive evidence for contrasts between predictor levels that we have not coded for. Results also show that certain trends generalize across multiple lexemes, and do not necessarily reflect lexeme-specific idiosyncrasies in change. For instance, the trigram $\#mo$ shows the strongest association with the N pattern, and is found in a number of lexemes in our dataset (4). Finally, we see a strong association between individual patterns and the phonological environments in which they first emerged. For example, a number of trigrams favoring the N pattern contain o, whereas high, low, and long vowels, as well as vowelless trigrams are overrepresented among those that disfavor the same pattern. This is perhaps a consequence of the fact that lexemes with the stem vowels e and o are where N alternations appeared initially as a result of sound changes (see *perder* in Table 1).

5.2   Semantic profile

In addition to phonological predictors, we are able to inspect model coefficients representing the extent to which the semantic profile of each Latin lemma contributes to its long-term preference for a given pattern type.

Figure 10 provides posterior distributions for coefficients representing the association between lexemes in our data set and long-term preferences for different patterns based on their semantic profiles, ordered according to their posterior medians. These coefficients are based on the semantic profiles (specifically, LatinBERT embeddings) of the 3sg present form of each Latin lemma. Coefficients are generated by a Gaussian processes (we assume separate processes for each pattern type) with a squared-exponential covariance kernel based on pairwise cosine distances between embeddings. Posterior samples are taken from the model `vary_branch+lemma+pattern+conj+pattern×conj+sem`. Further inspection of the coefficients of the Gaussian Processes used to generate these parameters reveals high posterior values for the length scales $\boldsymbol{\lambda}$ (see Figure 9), indicating that our model ignores the semantic distances between lemmas, learning roughly equal inter-lemma covariance values and generating lexeme-specific coefficients that do not necessarily represent semantic profiles. This behavior potentially leads to overfitting and results in a lower stacking weight than for the more restrictive model `vary_branch+lemma+pattern+conj+pattern×conj+phon`. The fact that our models essentially ignored these semantic representations is surprising, since the coefficients appear to show decisive or strong effects on long-term preference for certain pattern types for semantically related forms (e.g., *mordere* 'bite' and *bibere* 'drink' on N, *dicere* 'say' and *sapere* 'know' on L, etc.). At the same time, the coefficients representing associations with the P pattern exhibit higher variance than those for N and L, with most 95% and 85% credible intervals overlapping with 0). This is perhaps due to the general obsolescence of the P pattern.

There are several possible explanations for this particular result. On one hand, it is possible that LatinBERT embeddings are a realistic numerical representation of Latin verbal semantics,
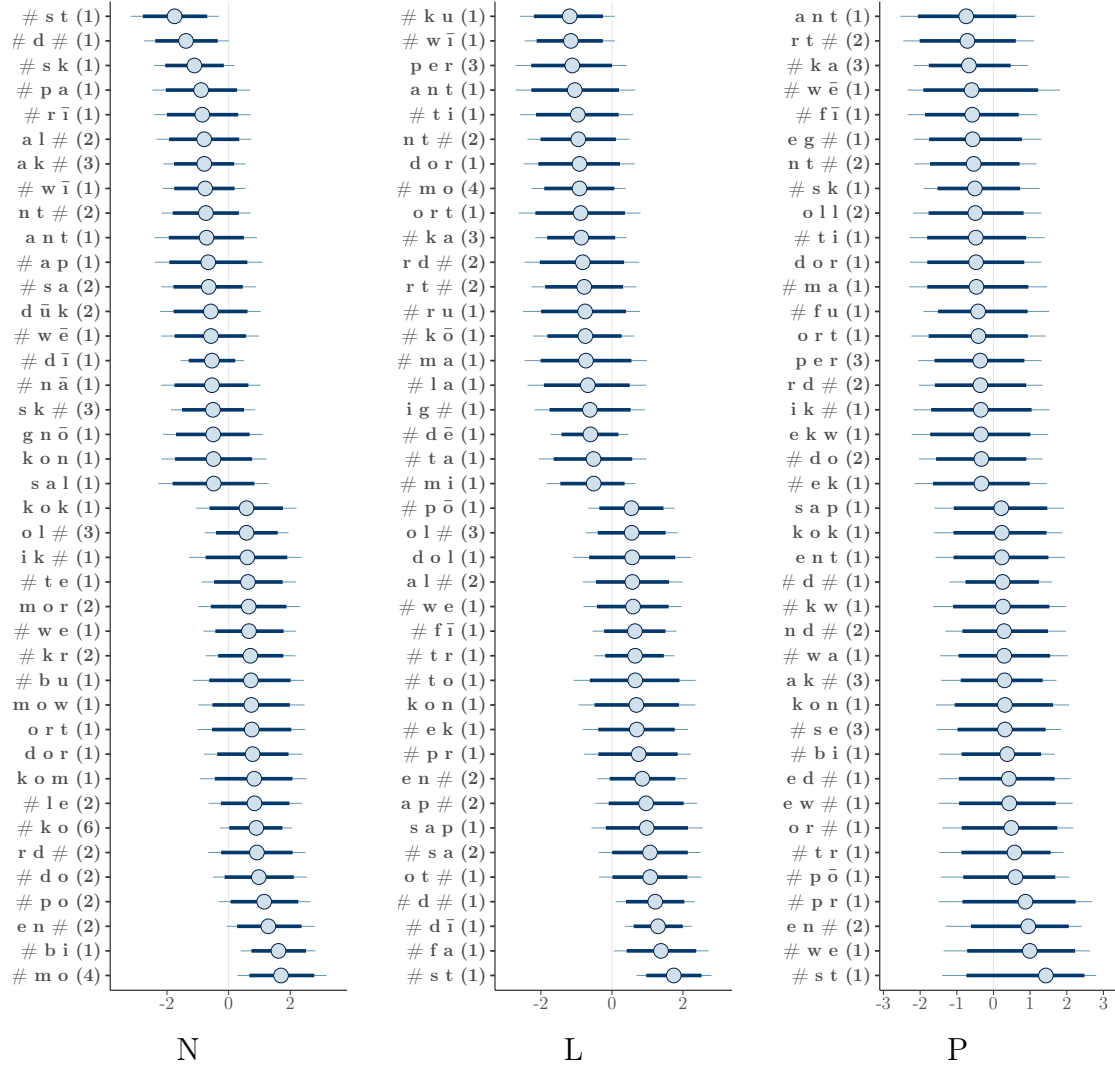
Figure 8: Posterior distributions for coefficients representing the association between trigrams present in the stems of Latin etyma and their effect on long-term preference for N (left), L (center), and P (right) patterns, ordered according to their posterior medians. For visual clarity, we give only coefficients with the 20 highest and 20 lowest medians. Trigram frequencies across lemma stems in our sample are given in parentheses. Figures displaying all distributions can be found in the repository associated with this paper.
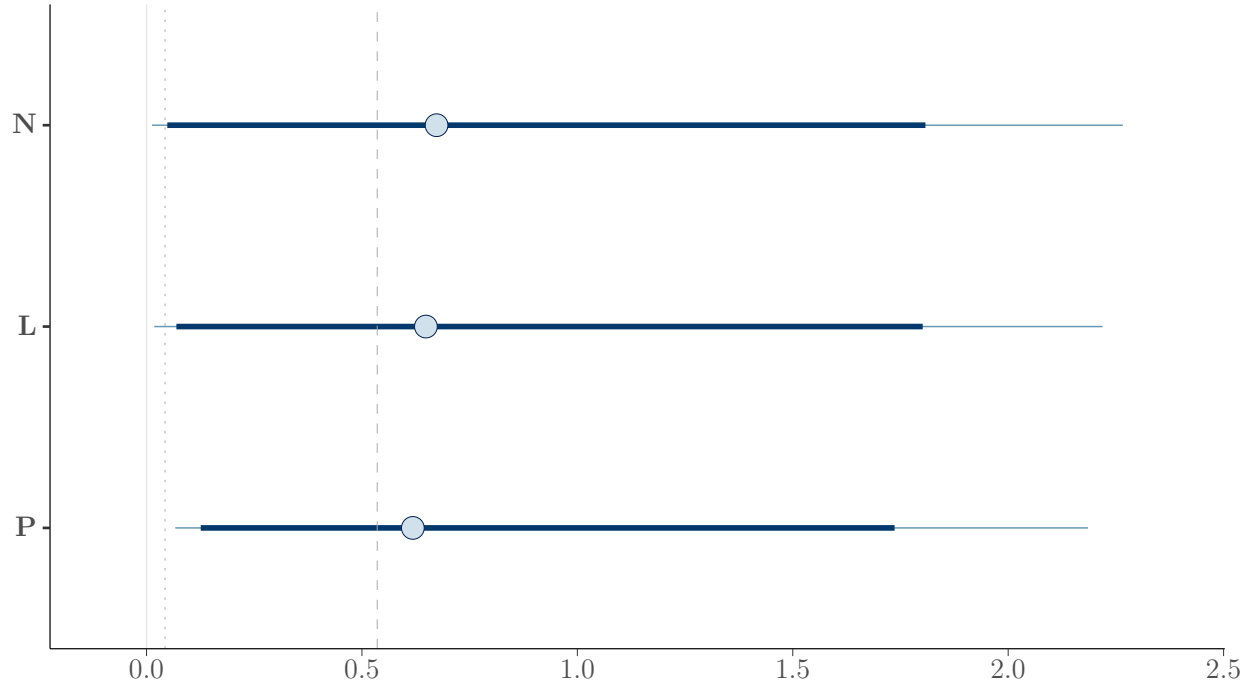
Figure 9: Medians and posterior 95% and 85% (shaded) credible intervals for length scale parameters $\boldsymbol{\lambda}$ of Gaussian processes underlying coefficients representing the association between lexemes in our data set and long-term preferences for different patterns based on their semantic profiles. The dashed and dotted lines represent the maximum and minimum cosine distance between lexemes.

but that our model is somehow misspecified for the purpose of capturing verbal semantic profiles. This issue could potentially be resolved by placing stronger, more informative priors over the parameters involved in the Gaussian processes used to generate the coefficients in Figure 10, though this could have unintended consequences, such as a knock-on effect on values of other parameters in the model.

Another possibility is that LatinBERT embeddings are a reliable model of semantics and that our model is in fact properly parameterized, but that semantic profiles are not particularly active or detectable in morphological change. This interpretation would likely strike most readers as controversial, given the wide range of literature arguing for the role of semantics in morphological change (90, 91).

Finally, it is possible that our models failed to infer coefficients representing semantic profiles of verbs because of the model of vector semantics that we employed. We used LatinBERT, a transformer-based model, to generate static word embeddings, although the primary purpose of such models is to produce contextual word embeddings. Methods for generating higher-quality static word embeddings from transformer-based models have been proposed and may prove relevant in future work (92). In general, the visualization in Figure 4 shows that while certain contextually related verbs are found in proximity to one another (e.g., *ridere* 'laugh' vs. *plangere* 'cry', *legere* 'read' vs. *scribere* 'write'), other proximate pairs show no clear semantic relation (e.g., *nasci* 'be born' vs. *lauare* 'wash').

5.3 Branch level variation

Model comparison supports the hypothesis that speeds and general preferences for irregularity differ from branch to branch. Since we run our model over multiple trees in our tree sample, it is challenging to visualize these trends in aggregate, as different trees have different branching structure. To confront this problem, we fit models on a maximum clade credibility (MCC) tree generated from the entire tree sample. Below, we analyze branch level variation for `vary_branch+lemma+pattern+conj+pattern×conj+sem`, the model with the highest ELPD.

Figure 11 shows MCC trees annotated according to branch-level differences in speed and general preference for irregular patterns. Branches are colored blue if branch-level median posterior coefficients are less than zero, and red if greater. Branches are solid if 95% or more of samples are either greater than or less than zero, and dotted otherwise.

As is clear, a relatively small number of branches show a definitive deviation from baseline speed levels. Interestingly, among these branches, we see a clear tendency towards decreases in speed in branches leading to standard language varieties (standard French, Portuguese, Spanish, Italian, and Romanian all show blue branches in Figure 11). This is in line with the notion that the centripetal forces involved in the codification of standard languages have the effect of restraining the normal rate of diversification and change found in other circumstances (94, 95). We see branches with credibly higher than baseline speeds of change within the French clade of the tree as well as in Alpine Romance. This may be linked to sociolinguistic factors such as intense language contact or 'catastrophic' events hypothesized to speed up language change (96, 97).
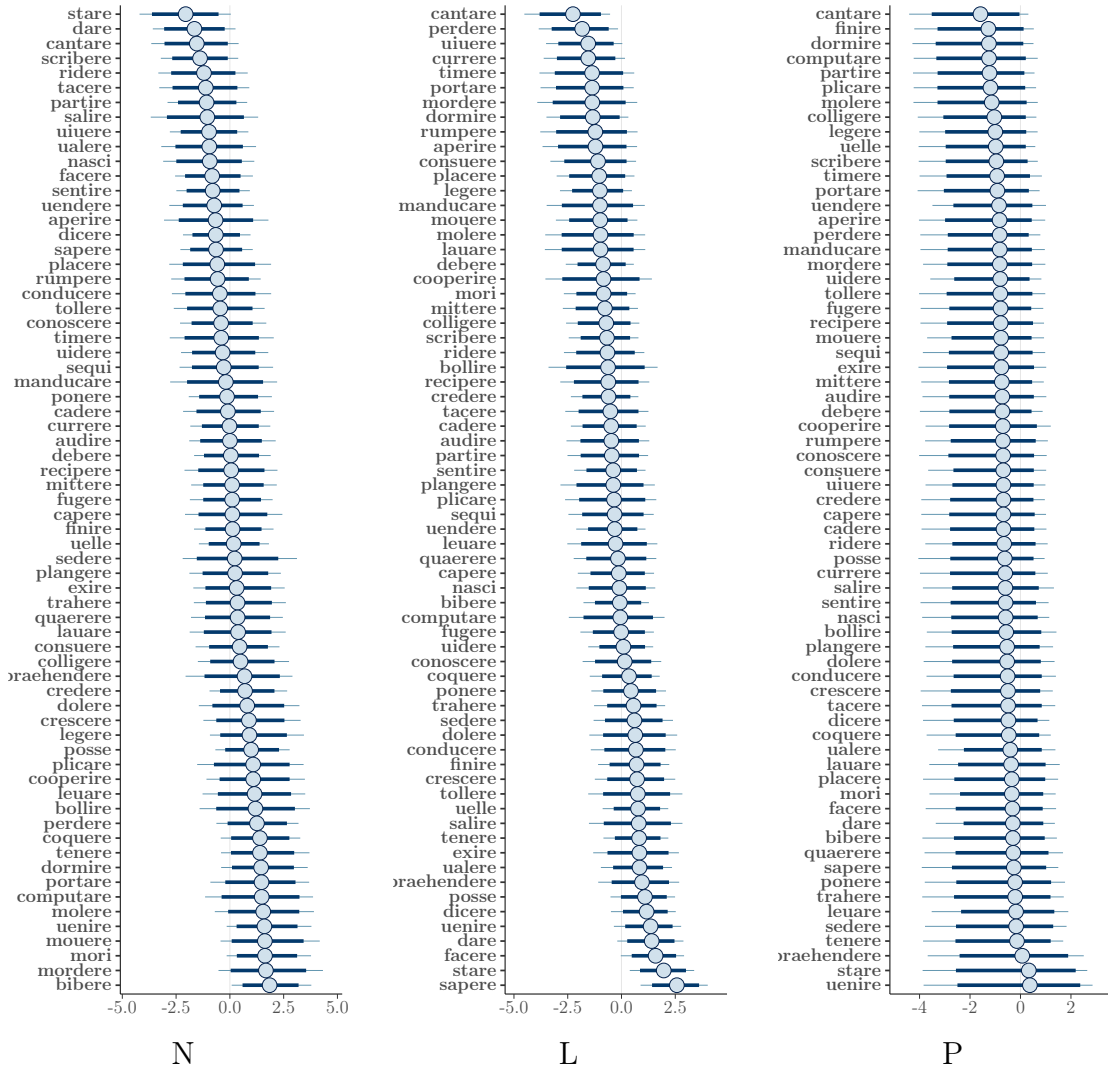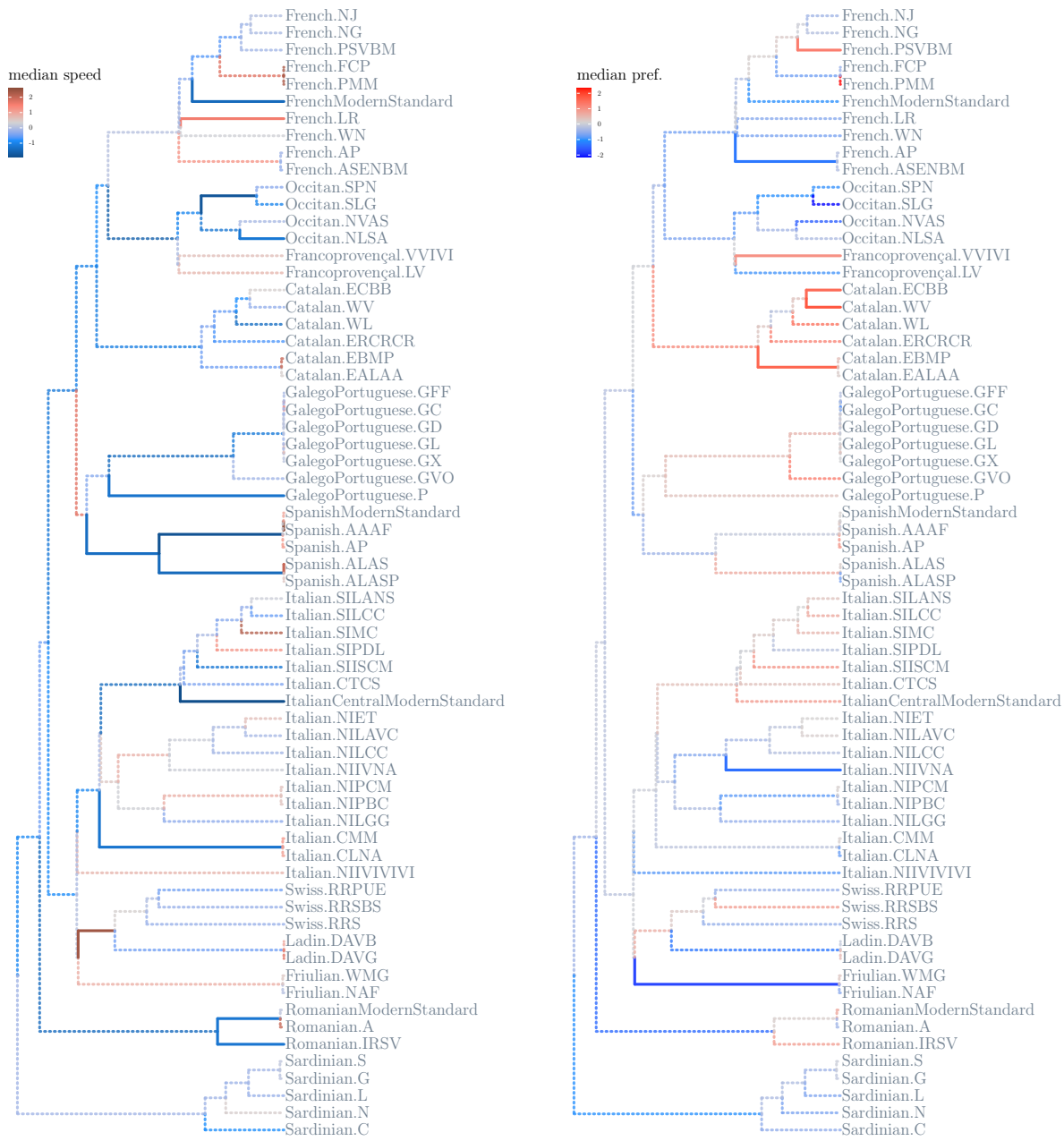
Figure 10: Posterior distributions for coefficients representing the inferred association between semantic profiles of lexemes and long-term preferences for N (left), L (center), and P (right) patterns, ordered according to their posterior medians.

Speed

Preference for irregularity

Figure 11: MCC tree showing branch-level variation in speed and general preference for irregularity for `vary_branch+lemma+pattern+conj+pattern×conj+sem`, the model with the highest ELPD, generated with the R package `ggtree` (93). Branch-level median posterior coefficients are either less than (blue) or greater than (red) zero. Branches are solid if 95% or more of samples are either greater than or less than zero, or dotted otherwise.

Concerning the general preference for irregularity, we see a general trend towards the loss of the inherited stem alternations in Gallo-Romance branches (i.e., French and Occitan broadly) as well as in Friulian and other northern Italian dialects. These results fit the overall profile of these varieties, which are generally innovative not only morphologically, but also phonologically and syntactically (98, 99). Catalan, by contrast, shows a preference for more stem alternations. This is probably the result of the L pattern fusing with P in these varieties and becoming very productive (100), and hence really semi-regular.

It is likely that the patterns identified here represent a complex causal scenario involving multiple linguistic and extralinguistic variables, and more work is needed to properly disentangle these factors. The framework we have presented here provides a principled approach for doing so.

6   Discussion

We presented the results of a series of phylogenetic models that isolate the effects of different predictors on orthogonal facets of linguistic evolution, namely speed (representing overall stability) and stationary probability (representing long-term preferences and proclivities). While speeds of change and stationary probabilities vary across the features in our data set, we found that several variables of interest explain variation in stationary probability but not speed. Importantly, our models allowed us to assess the effects of lemma frequency on speed as well as stationary probability while controlling for a number of other predictors as well as lineage-specific variation in patterns of change.

The observed effect of frequency is limited to stationary probabilities (long-term preferences), and does not affect how fast verbs move between states (i.e. how often they acquire or lose a particular stem alternation pattern). This appears to go against the effects of frequency that have been observed in lexical replacement ((33), though see also (101)) and against the expectations derived from population size differences in genetics, according to which the time to fixation of a variable trait is longer the larger the population/sample/number of tokens (102, 103). From this perspective one would expect lower frequency verbs to exhibit more variability and change. The very few verbs for which regular and irregular forms are both admissible in Spanish, for example, tend to be very low frequency (e.g., 1st person *roo*, *roigo*, *royo* for *roer* 'gnaw'; 1st person *yazo*, *yazgo*, *yago* for *yacer* 'lie dead', etc.), which plausibly stems from the fact that speakers simply do not hear these forms frequently enough to learn whether they have stem alternation or not. Yet, this is not what our model suggests: frequency does not impact speed of change.

Instead, we find that frequency primarily impacts long-term stationary probabilities. Asymmetries in these facets of evolution, rather than in speed, are indicative of general pressures on linguistic evolution (104, 105, 106, 107), here most likely reflecting biases induced by processing and/or learning, such as higher demands on discriminability (17, 18, 2, 19).

Additional reasons for this result are open for discussion: if at any given point in time low frequency items tend to be regular, and if analogical irregularization is less common than analogical regularization, infrequent lemmas might have fewer opportunities to change than

higher frequency items, which might compensate for a greater "inherent instability" due to infrequent usage (47). Similarly, while higher frequency items often behave in morphologically idiosyncratic ways, low-frequency items tend to cluster in large inflection classes (11), which may provide a certain "herd immunity" to their members that compensates for their rarity at the individual lemma level.

Our results show, in addition, that factors such as phonological and semantic similarity also play a role in shaping long-term trends. This is in line with the finding that language users deploy inter-lexeme semantic and phonological similarity when realizing inflected forms (108, 2, 87).

Our general framework can be expanded in a number of ways. A potential objection to our modeling decisions is that we have based variables regarding lemma frequency and conjugation, as well as the phonological and semantic profiles of verbs, on values taken from Latin, rather than allowing these variables to change/coevolve over the phylogeny. Our decision is largely one of convenience, as implementing a more complex model in which multiple variables coevolve and influence each other is computationally challenging (but could draw upon recently proposed frameworks, e.g., (109)). Regardless of this simplification, we believe our results are still meaningful, as we have shown how certain variables influence the evolution of linguistic features on a broad scale.

Our approach can also be used to address a number of outstanding questions in morphological change. For instance, recent work highlights the role of repulsion dynamics in shaping inflectional class systems (110), making the prediction that sufficiently dissimilar lexemes should display negative covariance in their affinity for the same pattern types. The Gaussian process squared-exponential kernel function that we use to model the effects of semantic similarity cannot model negative covariance, but other priors over correlation matrices (111) and kernel functions (112, 113) are in theory capable of doing so, and could be employed to probe the role of phonological and semantic similarity and dissimilarity in driving morphological change. Beyond this, our distributional approach can be integrated with models of continuous trait evolution (114) in order to explicitly test hypotheses regarding signatures of drift or selection. We leave these possibilities for future research.

7   Conclusion

In this paper, we contribute to the understanding of the mechanisms which underlie the distribution of irregular forms in Romance languages. As we explained in previous sections, the maintenance and loss of irregularity could be the result of two different logical components of change: long-term preference for irregularity in certain environments, and/or greater degree of change (both regularization and irregularization) in forms with certain distributional properties. In the concrete phenomenon that we explore (presence of irregular stem alternation in Romance verbs), the greater irregularity of more frequent verbs could be the result of conservatism (i.e., slower change), or of a long-term preference for irregularity/stem flexivity in these verbs. Both options are plausible and compatible. Our results, however, suggest that long-term probability (and not relative stability) is the main factor upon which our

predictors (in particular frequency) exert their influence.

Data availability: Code accompanying this paper can be found at `https://osf.io/yxhzr/?view_only=ceb404ac0aaf4b72b14bb68084c1b27f`.

References

[1] Ackerman F, Malouf R. Morphological organization: The low conditional entropy conjecture. Language. 2013:429-64.

[2] Blevins JP, Milin P, Ramscar M. The Zipfian paradigm cell filling problem. In: Perspectives on morphological organization. Brill; 2017. p. 139-58.

[3] Boyé G, Schalchli G. Realistic data and paradigms: The paradigm cell finding problem. Morphology. 2019;29:199-248.

[4] Erdmann A, Elsner M, Wu S, Cotterell R, Habash N. The paradigm discovery problem. arXiv preprint arXiv:200501630. 2020.

[5] Van Driem G. A grammar of Limbu. In: A Grammar of Limbu. De Gruyter Mouton; 2011. .

[6] Herce B. Stem alternations in Kiranti and their implications for the morphology–phonology interface. Journal of Linguistics. 2020:1-43.

[7] Vaux B. The Phonology of Armenian. Oxford: Clarendon Press; 1998.

[8] Aronoff M. Morphology by itself: Stems and inflectional classes. Cambridge, MA: MIT Press; 1994.

[9] Bickel B. In the Vestibule of Meaning: Transitivity Inversion as a Morphological Phenomenon. Studies in Language. 1994;19:73-127.

[10] Stolz T, Otsuka H, Urdze A, van der Auwera J. Irregularity in Morphology (and beyond). Akademie Verlag; 2012.

[11] Herce B. Deconstructing (ir)regularity. Studies in Language. 2019;43(1).

[12] Vennemann T. Sound change and markedness theory. In: Stockwell RKS R P Macaulay, editor. Linguistic change and generative theory. Bloomington: Indiana University Press; 1972. p. 230-74.

[13] Anttila R. Analogy. vol. 10. The Hague: Mouton; 1977.

[14] Wurzel WU. Flexionsmorphologie und Natürlichkeit: ein Beitrag zur morphologischen Theoriebildung. vol. 21. Walter de Gruyter GmbH & Co KG; 1984.

[15] Clark EV. The principle of contrast: A constraint on language acquisition. In: MacWhinney B, editor. Mechanisms of language acquisition: the 20th annual Carnegie Symposium on Cognition. Lawrence Erlbaum; 1987. p. 1-33.

[16] Dressler WU. Natural morphology. In: Oxford Research Encyclopedia of Linguistics; 2019. .

[17] Nübling D. The development of "junk". Irregularization strategies of have and say in the Germanic languages. In: Yearbook of Morphology 1999. Springer; 2001. p. 53-74.

[18] Nübling D. How do exceptions arise? On different paths to morphological irregularity. In: Simon H, Wiese H, editors. Expecting the unexpected: exceptions in grammar. Berlin: de Gruyter; 2011. p. 139-62.

[19] Marzi C, Ferro M, Pirrelli V. A Processing-Oriented Investigation of Inflectional Complexity. Front Comm. 2019;4:1-23.

[20] Garrett A. Paradigmatic uniformity and markedness. In: Good J, editor. Explaining linguistic universals: Historical convergence and universal grammar. Oxford: Oxford University Press; 2008. p. 124-43.

[21] Gaeta L. Is analogy economic? In: Montermini FGBNH, editor. Selected Proceedings of the 5th Décembrettes: Morphology in Toulouse. Somerville, MA: Cascadilla Press; 2007. p. 20-33.

[22] Wu S, Cotterell R, O'Donnell TJ. Morphological irregularity correlates with frequency. arXiv preprint arXiv:190611483. 2019.

[23] Malkiel Y. New problems in Romance interfixation (I): the velar insert in the present tense (with an excursus on-zer/-zir verbs). Romance Philology. 1974;27(3):304-55.

[24] Maiden M. Irregularity as a determinant of morphological change. Journal of linguistics. 1992;28(2):285-312.

[25] R Anderson S. Phonologically conditioned allomorphy in the morphology of Surmiran (Rumantsch). Word Structure. 2008;1(2):109-34.

[26] Maiden M. The Latin 'third stem'and its Romance descendants. Diachronica. 2013;30(4):492-530.

[27] O'Neill P. The origin and spread of velar allomorphy in the Spanish verb: a morphomic approach. Bulletin of Hispanic Studies. 2015;92(5):489-518.

[28] Esher L. Morphomes and predictability in the history of Romance perfects. Diachronica International Journal for Historical Linguistics Founded by EFK Koerner, General Editor, 1984–2001. 2015;32(4):494-529.

[29] Maiden M. The Romance verb: Morphomic structure and diachrony. Oxford University Press; 2018.

[30] Herce B. Morphome interactions. Morphology. 2019;29(1).

[31] Bürkner PC. Bayesian distributional non-linear multilevel modeling with the R package brms. arXiv preprint arXiv:170511123. 2017.

[32] Huelsenbeck JP, Suchard MA. A nonparametric method for accommodating and testing across-site rate variation. Systematic Biology. 2007;56(6):975-87.

[33] Pagel M, Atkinson QD, Meade A. Frequency of word-use predicts rates of lexical evolution throughout Indo-European history. Nature. 2007;449(7163):717-20.

[34] Heath T, Holder M, Huelsenbeck J. A Dirichlet Process Prior for estimating lineage-specific substitution rates. Molecular biology and evolution. 2011;29:939-55.

[35] Greenhill SJ, Wu CH, Hua X, Dunn M, Levinson SC, Gray RD. Evolutionary dynamics of language systems. Proceedings of the National Academy of Sciences. 2017;114(42):E8822-9.

[36] Gahl S, Yao Y, Johnson K. Why reduce? Phonological neighborhood density and phonetic reduction in spontaneous speech. Journal of memory and language. 2012;66(4):789-806. Publisher: Elsevier.

[37] Jacobs CL, MacDonald MC. A chimpanzee by any other name: The contributions of utterance context and information density on word choice. Cognition. 2023 Jan;230:105265. Available from: https://linkinghub.elsevier.com/retrieve/pii/S0010027722002530.

[38] Aylett M, Turk A. The Smooth Signal Redundancy Hypothesis: A Functional Explanation for Relationships between Redundancy, Prosodic Prominence, and Duration in Spontaneous Speech. Language and Speech. 2004 Mar;47(1):31-56. Available from: http://journals.sagepub.com/doi/10.1177/00238309040470010201.

[39] Jaeger TF. Redundancy and reduction: Speakers manage syntactic information density. Cognitive psychology. 2010;61(1):23-62.

[40] Mahowald K, Fedorenko E, Piantadosi ST, Gibson E. Info/information theory: Speakers choose shorter words in predictive contexts. Cognition. 2013;126(2):313-8.

[41] Wedel AB, Hall KC, Jaeger TF, Hume E. The Message Shapes Phonology. 2019.

[42] Bybee JL. Natural morphology: The organization of paradigms and language acquisition. Crosscurrents in second language acquisition and linguistic theories. 1991;2:67-92.

[43] Pinker S. Words and Rules. New York. NY: Basic Books. 1999.

[44] Haspelmath M, Sims A. Understanding morphology. Routledge; 2013.

[45] De Smet I, Van de Velde F. Reassessing the evolution of West Germanic preterite inflection. Diachronica. 2019;36(2):139-80.

[46] Sims-Williams H. Token frequency as a determinant of morphological change. Journal of Linguistics. 2021:1–37.

[47] Cotterell R, Kirov C, Hulden M, Eisner J. On the diachronic stability of irregularity in inflectional morphology. arXiv preprint arXiv:180408262. 2018.

[48] Newberry MG, Ahern CA, Clark R, Plotkin JB. Detecting evolutionary forces in language change. Nature. 2017;551(7679):223-6.

[49] Lieberman E, Michel JB, Jackson J, Tang T, Nowak MA. Quantifying the evolutionary dynamics of language. Nature. 2007;449(7163):713-6.

[50] Luce PA. A computational analysis of uniqueness points in auditory word recognition. Perception & Psychophysics. 1986;39(3):155-8.

[51] Radeau M, Mousty P, Bertelson P. The effect of the uniqueness point in spoken-word recognition. Psychological Research. 1989;51:123-8.

[52] Balling LW, Baayen RH. Probability and surprisal in auditory comprehension of morphologically complex words. Cognition. 2012;125(1):80-106.

[53] Fertig DL. Variation and change in verbal inflection in late medieval and early modern Nuremberg; 1994.

[54] Fratini V, Acha J, Laka I. Frequency and morphological irregularity are independent variables. Evidence from a corpus study of Spanish verbs. Corpus Linguistics and Linguistic Theory. 2014;10(2):289-314.

[55] Maiden M. Oxford online database of Romance verb morphology. 2010.

[56] Beniamine S, Maiden M, Round E. Opening the Romance Verbal Inflection Dataset 2.0: a CLDF Lexicon. In: Proceedings of The 12th Language Resources and Evaluation Conference; 2020. p. 3027-35.

[57] Bouckaert R, Lemey P, Dunn M, Greenhill SJ, Alekseyenko AV, Drummond AJ, et al. Mapping the origins and expansion of the Indo-European language family. Science. 2012;337(6097):957-60.

[58] Chang W, Cathcart C, Hall D, Garrett A. Ancestry-constrained phylogenetic analysis supports the Indo-European Steppe Hypothesis. Language. 2015;91(1):194-244.

[59] Sagart L, Jacques G, Lai Y, Ryder RJ, Thouzeau V, Greenhill SJ, et al. Dated language phylogenies shed light on the ancestry of Sino-Tibetan. Proceedings of the National Academy of Sciences. 2019;116(21):10317-22.

[60] Tavaré S. Some probabilistic and statistical problems in the analysis of DNA sequences. Lectures on Mathematics in the Life Sciences. 1986;17(2):57-86.

[61] Tuffley C, Steel M. Modeling the covarion hypothesis of nucleotide substitution. Mathematical Biosciences. 1998;147(1):63-91.

[62] Heath T, Holder M, Huelsenbeck J. A Dirichlet Process Prior for estimating lineage-specific substitution rates. Molecular Biology and Evolution. 2011;29:939-55.

[63] Lartillot N, Poujol R. A phylogenetic model for investigating correlated evolution of substitution rates and continuous phenotypic characters. Molecular biology and evolution. 2011;28(1):729-44.

[64] Lartillot N. Phylogenetic patterns of GC-biased gene conversion in placental mammals and the evolutionary dynamics of recombination landscapes. Molecular biology and evolution. 2013;30(3):489-502.

[65] Horvilleur B, Lartillot N. Monte Carlo algorithms for Brownian phylogenetic models. Bioinformatics. 2014;30(21):3020-8.

[66] Beaulieu JM, O'Meara BC. Hidden Markov Models for Studying the Evolution of Binary Morphological Characters. In: Garamszegi LZ, editor. Modern Phylogenetic Comparative Methods and Their Application in Evolutionary Biology: Concepts and Practice. Heidelberg, New York, Dordrecht, London: Springer; 2014. p. 395-408.

[67] Maiden M, Smith JC, Cruschina S, Hinzelin MO, Goldbach M. Oxford Online Database of Romance Verb Morphology. University of Oxford. Online at http://romverbmorph.clp.ox.ac.uk/; 2010.

[68] Delatte L, Evrard É, Govaerts S, Denooz J. Dictionnaire fréquentiel et index inverse de la langue latine. 1981.

[69] Bamman D, Burns PJ. Latin BERT: A Contextual Language Model for Classical Philology. arXiv preprint arXiv:200910053. 2020.

[70] Wichmann S, Holman EW, Brown CH. The ASJP Database (version 18); 2018. http://asjp.clld.org/.

[71] Heggarty P, Shimelman A, Abete G, Anderson C, Sadowsky S, Paschen L, et al.. Sound Comparisons: Exploring Diversity in Phonetics across Language Families; 2019. Available from: https://soundcomparisons.com.

[72] Jäger G. Global-scale phylogenetic linguistic inference from lexical resources. Scientific Reports. 2018;5.

[73] Hall RA. Comparative Romance Grammar. New York/London/ Amsterdam: Elsevier; 1974.

[74] Höhna S, Landis MJ, Heath TA, Boussau B, Lartillot N, Moore BR, et al. RevBayes: Bayesian phylogenetic inference using graphical models and an interactive model-specification language. Systematic Biology. 2016;65(4):726-36.

[75] Yang Z, Rannala B. Bayesian phylogenetic inference using DNA sequences: a Markov Chain Monte Carlo method. Molecular biology and evolution. 1997;14(7):717-24.

[76] Bouckaert RR. DensiTree: making sense of sets of phylogenetic trees. Bioinformatics. 2010;26(10):1372-3.

[77] Maddison WP. Missing Data Versus Missing Characters in Phylogenetic Analysis. Systematic Biology. 1993;42(4):576-81. Available from: `http://www.jstor.org/stable/2992490`.

[78] Felsenstein J. Inferring phylogenies. Sunderland, Mass.: Sinauer Associates; 2004.

[79] Bürkner PC, Charpentier E. Modelling monotonic effects of ordinal predictors in Bayesian regression models. British Journal of Mathematical and Statistical Psychology. 2020;73(3):420-51.

[80] Vehtari A, Gelman A, Gabry J. Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC. Statistics and computing. 2017;27(5):1413-32.

[81] Gabry J, Češnovar R. CmdStanR: R Interface to 'CmdStan'; 2021. Https://mc-stan.org/cmdstanr/, https://discourse.mc-stan.org.

[82] Carpenter B, Gelman A, Hoffman MD, Lee D, Goodrich B, Betancourt M, et al. Stan: A probabilistic programming language. Journal of statistical software. 2017;76(1)(1):1-32.

[83] Yao Y, Vehtari A, Simpson D, Gelman A. Using stacking to average Bayesian predictive distributions. Bayesian Analysis. 2017.

[84] Vehtari A, Gabry J. Bayesian Stacking and Pseudo-BMA weights using the loo package; 2022. Tutorial available online at `https://mc-stan.org/loo/articles/loo2-weights.html`.

[85] Montermini F, Boyé G. Stem relations and inflection class assignment in Italian. Word Structure. 2012;5(1):69-87.

[86] Baayen RH, Chuang YY, Blevins JP. Inflectional morphology with linear mappings. The Mental Lexicon. 2018;13(2):230-68.

[87] Baayen RH, Chuang YY, Shafaei-Bajestan E, Blevins JP. The discriminative lexicon: A unified computational model for the lexicon and lexical processing in comprehension and production grounded not in (de) composition but in linear discriminative learning. Complexity. 2019;2019.

[88] Guzmán Naranjo M. Analogical classification in formal grammar. Language Science Press; 2019.

[89] Herce B. Morphological autonomy and the long-term vitality of morphomes: stem-final consonant loss in Romance verbs and paradigmatic analogy. Morphology. 2023;33(2):153-87.

[90] Barðdal J. Productivity: Evidence from case and argument structure in Icelandic. vol. 8. John Benjamins Publishing; 2008.

[91] Nesset T, Makarova A. Testing the Semantic Homogeneity Constraint: Analogical Change and Russian Verbs. Journal of Historical Linguistics. 2014;4(2):161-91.

[92] Gupta P, Jaggi M. Obtaining Better Static Word Embeddings Using Contextual Embedding Models. In: Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers); 2021. p. 5241-53.

[93] Yu G, Lam TTY, Zhu H, Guan Y. Two methods for mapping and visualizing associated data on phylogeny using ggtree. Molecular Biology and Evolution. 2018;35:3041-3. Available from: https://doi.org/10.1093/molbev/msy194.

[94] Ferguson CA. Diglossia. word. 1959;15(2):325-40.

[95] Joseph JE. Eloquence and power: The rise of language standards and standard languages. Basil Blackwell; 1987.

[96] Labov W. The intersection of sex and social class in the course of linguistic change. Language variation and change. 1990;2(2):205-54.

[97] Gaeta L, Seiler G. A specter is haunting Europe: the Alps as a linguistic area? STUF-Language Typology and Universals. 2021;74(1):1-16.

[98] Kerkhof PA. Language, law and loanwords in early medieval Gaul: language contact and studies in Gallo-Romance phonology. Leiden University; 2018.

[99] Simonenko A, Crabbé B, Prévost S. Agreement syncretization and the loss of null subjects: quantificational models for Medieval French. Language Variation and Change. 2019;31(3):275-301.

[100] Badal Bertolín M. Morfologia verbal catalana: anàlisi diacrònica dels verbs velaritzats. 2021.

[101] Wilson FM, Pappas PA, Mooers AO. The role of frequency of use in lexical change: Evidence from Latin and Greek. Diachronica. 2019;36(4):584-612.

[102] Otto SP, Whitlock MC. The probability of fixation in populations of changing size. Genetics. 1997;146(2):723-33.

[103] Reali F, Griffiths TL. Words as alleles: connecting language evolution with Bayesian learners to models of genetic drift. Proceedings of the Royal Society B: Biological Sciences. 2010;277(1680):429-36.

[104] Maslova E. A dynamic approach to the verification of distributional universals. Linguistic Typology. 2000;4:307-33.

[105] Cysouw M. Understanding Transition Probabilities. Linguistic Typology. 2011;15:415-31.

[106] Bickel B. Distributional Typology: Statistical Inquiries into the Dynamics of Linguistic Diversity. In: Heine B, Narrog H, editors. The Oxford Handbook of Linguistic Analysis, 2nd Edition. Oxford: Oxford University Press; 2015. p. 901-23.

[107] Jäger G, Wahle J. Phylogenetic Typology. Front Psychol. 2021;12. Available from: https://www.frontiersin.org/article/10.3389/fpsyg.2021.682132.

[108] Bonami O, Beniamine S. Joint predictiveness in inflectional paradigms. Word Structure. 2016;9(2):156-82.

[109] Ringen E, Martin JS, Jaeggi A. Novel phylogenetic methods reveal that resource-use intensification drives the evolution of "complex" societies. EcoEvoRxiv. 2021.

[110] Round ER, Beniamine S, Esher L. The role of attraction-repulsion dynamics in simulating the emergence of inflectional class systems. arXiv preprint arXiv:211108465. 2021.

[111] Lewandowski D, Kurowicka D, Joe H. Generating random correlation matrices based on vines and extended onion method. Journal of multivariate analysis. 2009;100(9):1989-2001.

[112] Gasic M, Jurcicek F, Keizer S, Mairesse F, Thomson B, Yu K, et al. Gaussian processes for fast policy optimisation of pomdp-based dialogue managers. In: Proceedings of the SIGDIAL 2010 Conference; 2010. p. 201-4.

[113] Wilson A, Adams R. Gaussian process kernels for pattern discovery and extrapolation. In: International conference on machine learning. PMLR; 2013. p. 1067-75.

[114] Beaulieu JM, Jhwueng DC, Boettiger C, O'Meara BC. Modeling stabilizing selection: expanding the Ornstein–Uhlenbeck model of adaptive evolution. Evolution. 2012;66(8):2369-83.