

Convergence by Shared Ancestry in Romance

Paul Widmer

Professor of Comparative Indo-European Linguistics, University of Zurich,
Zurich, Switzerland
paul.widmer@uzh.ch

Stefan Dedio

Research Assistant, (Indo-European Studies Group), Department of
Comparative Language Science, University of Zurich, Zurich, Switzerland
stefan.dedio@uzh.ch

Barbara Sonnenhauser

Professor of Slavonic linguistics, University of Zurich, Zurich, Switzerland
barbara.sonnenhauser@uzh.ch

Abstract

In many cases of apparent contact-induced change the contribution of genealogical correlation in the language sample and its interaction with processes such as matter and pattern replication are difficult to specify. In order to get a better sense of the relevance of shared ancestry, we quantify the change in similarity since the late Middle Ages in a sample of Romance and Germanic languages with data from a selected grammatical domain (expression of reflexivity). We compare their dynamics to patterns of change of similarity in two contact zones in Europe, namely the British Isles (Dedio et al., 2019) and the Balkans. Concerning the genealogical signal, the results indicate a maintenance and gain of similarity in Romance as opposed to a loss of similarity in Germanic. This hints at the importance of the inherited states, the time since the split from the common ancestor, and subsequent developments. We presume that these factors are likely to be at the origin of the maintenance and increase in similarity observed for the sampled Romance varieties. While this result cannot be generalized beyond the specific case study presented here, the basic approach will contribute to a better understanding of how contact, genealogy and culture interact in shaping the dynamics of linguistic similarity.

Keywords

convergence – divergence – Romance – Germanic – Balkan area – reflexive constructions

1 Introduction

Languages may be similar for various reasons, most prominently because similarity between languages tends to increase through language contact, because shared ancestry constrains evolutionary pathways such that potential divergences over time are slowed down, or because shared cultural practices of speaker communities foster gain or maintenance of similarity of specific linguistic properties (cf. the role of Latin in medieval Europe or the role of Pāli in the Buddhist sphere). Since these factors never occur in isolation it is obvious that change and maintenance of similarity are conditioned by an interplay of such drivers with a multitude of interacting socio-economic, spatio-temporal, and linguistic factors. Assessing and comparing evolutionary dynamics within and across groups of genealogically, spatially, and culturally related languages is therefore not a trivial undertaking, and telling apart the contribution of genealogical, spatial, and cultural drivers even less so.

The Balkans provide a major case in point for these challenges. At least since Trubetzkoy's introduction of the notion of 'sprachbund' (Trubetzkoy, 1928), the languages situated in the geographical area of the Balkan peninsula have been regarded as constituting a prototypical linguistic convergence area. This assumption has remained more or less unquestioned and in turn has led to the teleological interpretation of changes observed for exemplary features to result from language contact (see Introduction to this volume). Upon closer inspection, however, the Balkans constitute a prime example for the methodological challenges related to the identification of linguistic areas, in particular identifying the contributions of universal tendencies vs. inheritance vs. contact, assessing the dynamics of historical development, and avoiding the danger of circularity resulting from focusing on highly salient, emblematic examples. Since linguistic areas are assumed to be characterized by a trend of convergence that is significantly stronger than what could be expected from family-internal developments, the problem of how to estimate the diachronic dynamics (i.e., degree and amount) of convergence within and across (sub-) families necessarily gains center stage. In assessing the dynamics of convergence, it does not suffice to focus on features that have been changing towards increasingly similar patterns (see Introduction to this volume). Instead, further

information needs to be considered, in particular 1) information on features that resist change, 2) information on the directions of change, i.e., gain or loss of a feature, whereby both may have led to a state of indeterminacy, e.g., an attestation of various options. Obviously, this task is hard to accomplish by qualitatively oriented research.

The present paper aims to approach the above-mentioned challenges by applying an empirical method for assessing and comparing the dynamics of change of similarity across languages with variable genealogical relatedness, thus complementing qualitative research. In doing so, it takes a slight detour by focusing not on the Balkans in the first place, but on some Romance varieties of Europe and their development within their family and in the context of the Balkan languages. The analysis proposed in this paper is carried out on the basis of a set of morphosyntactic features involved in constructions expressing reflexivity. This construction is not commonly discussed in the context of the Balkan area, such as to avoid the above-mentioned danger of teleological interpretations of the data.

In addressing the interaction between genealogical, spatial, and cultural factors in language change, Romance varieties provide a good test case (see Joseph, 1999 for a similar point). They form a well-attested genealogical grouping with a thoroughly investigated internal history, stretching across various linguistic and geographical spaces from West to East in southern Europe, interacting with each other and quite a few other linguistic varieties across very diverse cultural settings (e.g., Catholic, Protestant, Orthodox, Jewish contexts, and/or under Ottoman and Arabic rule, etc.) since the early Middle Ages.

Because of the broad spatial distribution that offers opportunities for manifold contact events we expect the influence from various contact scenarios to manifest itself as a lineage-internal divergence irrespective of cultural and genealogical bonds, whereas a strong genealogical signal is expected to keep the divergence to a minimum or bring about a gain of similarity within the Romance varieties irrespective of space.

To get a better understanding of the properties of the genealogical signal in Romance, we propose to compare the change in similarity observed in a sample of Romance varieties to the changes in another lineage, namely Germanic, and to the changes in various configurations (see Dedio et al., 2019), i.e., geographical spaces with their own particular history and socio-cultural and socio-linguistic properties. To evaluate the dynamics of change, we apply the methods introduced by Dedio et al. (2019), which are based on the changes in similarity between linguistic varieties over time in predefined groups of languages, and make use of their set of morphosyntactic and morphological variables used to encode reflexivity, adding Balkan and Romance varieties to

their sample. In order to assess the contribution of Latin, the common cultural point of reference, to the change in similarity in Romance we add Classical and Medieval Latin, which remain, and are kept, unchanged during the period of interest. For practical reasons, the timespans between ancestor and successor varieties show more variation in our sample than in Dedio et al.'s; this may have an effect on the results we are not able to control for at the moment.

The article is structured as follows: We will introduce the data and methods in Section 2. The results concerning the dynamics of change observed for the different samples will be presented in Section 3 and discussed in Section 4, focusing in particular on the Romance and Balkan samples. Section 5 provides a conclusion and a more general embedding of the insights gained for assessing the role of contact for linguistic diversity and change.

2 Data and Methods

Teasing apart areal and genealogical signals in quantitative studies of linguistic diversity and change is a rather new enterprise that only started gathering pace in the last decade. The reliability of the methods proposed so far (e.g., Freckleton and Jetz, 2009; Nelson-Sathi et al., 2010; Willems et al., 2016; Kelly and Nicholls, 2017; Murawaki and Yamauchi, 2018) has not been thoroughly established yet, and although some of them seem promising, they are not suitable for this study, as they operate on a macro-level. We thus apply the method developed in Dedio et al. (2019) for identifying convergence within a group of languages and expand their data collection with data relevant for our purposes. Supplementing their sample of Gallo-Romance (French, Normand, Jèrriais) by Italian and Spanish and Balkan-Romance varieties (Romanian, Aromanian, Judezmo, cf. Table 1), we explore the dynamics of change in the Romance data against the backdrop of their sample from the British Isles, a sample of languages from the Balkans,¹ and another lineage, viz. Germanic. The method applied in Dedio et al. (2019) uses predefined areal groups, i.e., configurations, that are based on geographical, historical and sociological information to infer signals of convergence within these groups contrasted with the developments outside these predefined areas. As the languages of the Romance subphylum of Indo-European cannot in any way be interpreted as belonging to a single coherent area, we depart from their notion of configuration and apply the

1 Note that the documents serving as data basis for the ancestor varieties of Macedonian, Bulgarian, Serbian and Torlak all constitute mixtures of Church Slavonic and vernacular features of the respective modern varieties each. This is the closest one can get to the older stages of these languages.

TABLE 1 Language pairings added to the collection in Dedio et al. (2019)

ancestor variety	approximate date CE	modern variety
Old Castilian	1250	Judezmo
Old Castilian	1250	Spanish
Old Albanian	1500	Modern Albanian
Old Italian	1250	Modern Italian
Middle Greek	1250	Modern Greek
Old Romanian	1500	Aromanian
Old Romanian	1500	Modern Romanian
Early modern Macedonian	1550	Macedonian
Early modern Bulgarian	1600	Bulgarian
Serbian Church Slavonic	1250	Torlak
Serbian Church Slavonic	1250	Serbo-Croatian

method also to groups of languages defined by shared ancestry (i.e., Romance and Germanic).

As Dedio et al. (2019), we focus on the expression of reflexivity, i.e., constructions in which the A and P arguments share the same referent (excluding constructions with exclusively reciprocal or passive semantics, and constructions that lack overt marking). The expression of these constructions encompasses a set of morphological, morphophonological, and morphosyntactic devices that can be captured in terms of binary variables as defined in the following list (for a full description, see Dedio et al., 2019; abbreviations in parentheses refer to the variables in Table 2). Note that by taking the construction as a starting point, our perspective differs from approaches that identify reflexivity as one possible function of particular devices, such as pronouns (see Cennamo, 2014 for Italian, Fehrman et al., 2010 for Slavic).

Positional dependency (DEP): Is the position of the reflexive marker directly dependent on the position of the verb? I.e., is there a rule that the marker must be placed relative to the verb? This includes basic rules like ‘the reflexive marker is the innermost marker left of the verbal root’ or more complex ones like ‘with regular inflected verbs, the marker is in slot 3 of the verbal template, but in slot 1 with infinitives.’

Stress: Can the reflexive marker establish its own stress domain? We have opted for splitting stress and phonological interaction into two values as stress domains tend to be larger than other domains of phonological and prosodic interaction (Bickel et al., 2009: 72) and this distinction helps capturing variation in our data.

TABLE 2 States of the variables in each language pairing (represented by the modern language variety in the first column). The value on the left of each column represents the state in the older stage, the value on the right the state at the end of the investigated period; for Latin, both values of each variable are identical.

	dep.	stress	interac.	allom.	num.	pers.	case	pre	post	rec.	pass.	exp.	syncr.
French	F	T	T	F	T	T	T	T	T	T	F	T	T
Jerriais	F	T	T	F	T	T	T	T	T	T	F	T	T
Normand	F	T	T	F	T	T	NA	T	T	T	F	T	T
Picard	T	F	T	F	T	T	F	T	T	T	F	T	T
Italian	B	B	B	F	T	T	F	T	T	T	F	T	T
Spanish	T	B	T	F	T	T	F	T	T	T	F	T	T
Romanian	F	T	B	B	T	T	T	B	T	T	T	T	T
Aromanian	F	T	B	B	T	T	T	B	T	T	T	T	T
Judezmo	T	B	T	F	T	T	F	T	T	T	T	T	F
Classical Latin	B	T	B	T	T	T	B	B	T	F	B	T	B
Medieval Latin	B	B	B	T	T	T	B	B	T	B	T	T	T
Greek	B	B	B	B	T	T	B	B	T	B	B	T	F
Albanian	B	T	B	T	B	B	B	T	T	B	B	B	B
Macedonian	B	T	F	F	F	F	B	T	T	B	T	B	T
Bulgarian	T	F	F	F	F	F	T	T	T	T	T	T	T
Torlak	B	T	F	F	F	F	B	T	T	B	T	F	T
Serbo-Croatian	B	F	B	F	F	F	B	T	T	B	B	F	T

Interaction (INTERAC.): Does the reflexive marker interact phonologically with surrounding linguistic items (i.e., vowel harmony, liaison, mutations, etc.)? This includes the whole range of phonological integration like syllabification patterns, vowel harmony phenomena, liaison, etc., but excludes stress (see above).

Allomorphy (ALLOM.): Does the reflexive marker display phonologically, morphologically, syntactically or lexically conditioned allomorphy if all relevant features (i.e., number, person, case, TAM of the verb) remain the same?

Inflection (NUM., PERS., CASE): Is the reflexive marker specified for a) person, b) number, and/or c) case?

Positioning (PRE, POST): Where is the marker positioned with respect to the verbal root (PRE or POST)? The marker may be positioned to the left of the verbal root or to its right or, as in some cases, both.

Equivalence set (REC., PASS.): Does the marker form an equivalence set with passive or reciprocal? I.e., is the reflexive marker also used to express passives or reciprocals? Both functional overlaps are widely attested in the languages of the world and are present in our sample (e.g., the Old Norse “medio-passive” is used to form reflexives, reciprocals, and anti-passives). As with reflexive marking in general, we do not distinguish between “normal” or “unmarked” ways to express these two functions, but also include marginal strategies.

Expandable (EXP.): Can the reflexive construction be expanded with an intensifier or a similar formant for stress, clarification or similar ends? For most languages in our sample this is identical to the reflexive – intensifier distinction variable of SAE, but we wanted this variable to have a broader scope in the event one of the languages without this distinction developed the ability to use an additional reflexive/intensifier (e.g., ***I myself hurt myself*).

Third person number syncretism (SYNCR.): Does the reflexive marker distinguish number values in the third person? With this variable, we try to capture a common variation in our data, e.g., zero differentiation with Modern High German *sich*, full differentiation like in Modern Standard English, or is there one marker that is used for singular and plural, while there is another one that encodes plural only like in Fering, where the singular P-pronoun can also be used in plural constructions.

Any of the above variables can take the states TRUE, FALSE, and, as languages tend to make use of more than one construction to express reflexivity, BOTH. By comparing the values of ancestor and successor varieties such as Old Norman French and Normand, we get an impression of the direction in which a linguistic lineage develops (Section 3.1). The similarities between the feature values of contemporaneous varieties serve as the basis for assessing the trend and magnitude of convergences between individual languages or larger areal or phylogenetic groups (see Section 3.2).

For each variable we first explore the direction of change between ancestor and successor language and subsequently, following the method introduced in Dedio et al. (2019), we analyze the data as follows (refer to the supplementary information for technical details):

1. We compute the pairwise similarity of the languages in our sample (cf. Table 1) for two different points in time, namely $t_1 = 1400$ CE (\pm c. 200 years) and $t_2 =$ c. 1950 CE, using an adapted version of the simple matching coefficient (SMC; Cheetham and Hazel, 1969).
2. Change between ancestor and successor languages is implemented by connecting the sampled languages at t_1 and t_2 to their phylogenetic ancestors. Lineages may fork into new sub-lineages, and therefore an ancestor language can have multiple successor languages at a given point in time. For example, Old Romanian is the ancestor of both Aromanian and Modern Romanian, cf. Table 1.
3. We then first compute the pairwise similarity between all languages that belong to the same t (e.g., Old Castilian and Middle Greek), and subsequently the change over time between pairings of identical ancestry, e.g., between the pairing Old Castilian/Old Romanian and their successor pairing Modern Spanish/Modern Romanian, and so on.
4. To assess how the developments relate to possible area formation processes and lineage specific trends, we split the measurements of changes in similarity into groups according to whether both languages of a pair are part of the Balkan area or both languages of a pair belong to the same lineage. The development of similarity inside of each of the resulting groups are visualized as rainbow plots.
5. In order to estimate the trend of the development in the Romance and Germanic samples, we model the change in similarity as a Bernoulli process $Y \sim \text{Bernoulli}(p)$, where p is the probability of success, i.e., in our case gain in similarity. To explore the magnitude of the process in Romance and Germanic, we model the change in similarity as transformed beta distribution $Y \sim \text{Beta}(\alpha, \beta)$ in the interval $[-1, 1]$ and estimate the posterior predictive distribution given the observed pairwise change within Romance. For a full description, refer to Dedio et al. (2019) and Ranacher et al. (2019).

3 Results

We first qualitatively investigate individual developments within Romance and the Balkan area before turning to a quantitative evaluation of the areal and phylogenetic developments.

3.1 *Direction of Changes*

In this section we investigate the direction of the changes per variable within each pairing. The feature specifications for each variable are given in Table 2, with the left value showing the specification for the older, the right value that for the younger stage (with the exception of Classical and Medieval Latin that serve as control factors for the influence of cultural practices and are kept constant).

Based on these specifications, the following observations concerning the change patterns can be made:

Overall sample: Two variables, namely NUM. and PERS., have not changed at all across the whole language sample, hinting at more general extra- and/or intra-linguistic trends.

Romance: All Romance varieties fully converge on, or preserve, a specification with DEP., NUM., ALLOM., PERS., PRE., REC., and PASS. With POST., all converge except for Aromanian; with EXP., the two Balkan-Romance varieties Judezmo and Aromanian display a behavior different from the other Romance varieties. Concerning STRESS, Italian and Romanian are the only varieties within Romance that fail to converge.

Balkan sample: The development in the Balkan sample is much less uniform than within Romance, there is no obvious discernible pattern.

Note that the development of these features is meaningful only in the context of the overall construction and its overall 13 features, such that NUM. and PERS. need to be included in all analyses even though they remain constant for all varieties.

Since it is difficult to obtain an overview of the more general trends displayed by the specification given in Table 2, we apply a quantitative approach that helps to discern the change in similarity for all of these features and to compare different groups of languages according to these changes. We zoom in and compute the pairwise changes in similarity for the Balkan and the Romance sample varieties, the latter including Classical and Medieval Latin. Romanian, Aromanian and Judezmo are included in both samples in order to be able to assess their (non)compliance with the overall trends in both samples. The results are shown in Fig. 1.

Upon visual inspection, Romance displays more maintenance and convergence of similarity than divergence with the notable exceptions of the Judezmo-Aromanian pairing and Classical Latin, from which most Romance varieties slightly diverge. Romanian not only diverges from both Latin samples, but also from Spanish. The former might relate to its cultural embedding in the orthodox sphere, the latter to geographical distance. The Gallo-Romance varieties all slightly converge with Medieval Latin, while all other varieties keep the same distance from Medieval Latin or diverge to a minor degree.

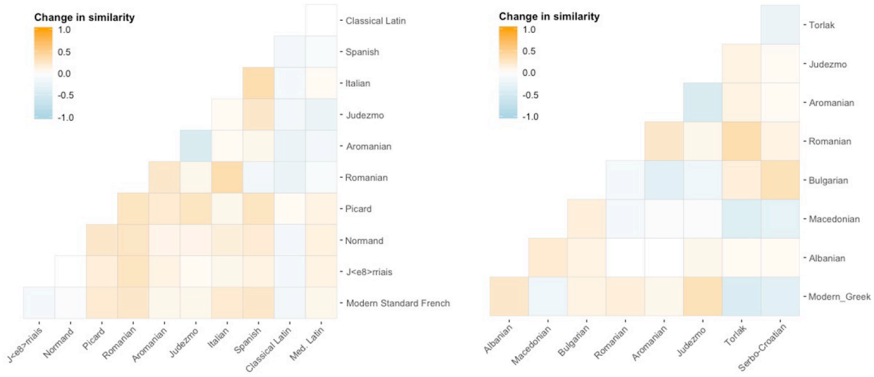


FIGURE 1 Pairwise change in similarity in the sample of Romance varieties including Classical and Medieval Latin (left panel) and the sample of varieties from the Balkans (right panel). Blue indicates divergence, orange convergence. The darker the color, the stronger the trend. White indicates no change.

Beyond their mutual divergence, Judezmo and Aromanian do not run counter to the overall dynamics of change in similarity within Romance. Within the Balkan sample they do not display an obvious increase in similarity. Rather, both converge with some Balkan varieties but diverge from others. Doing so, they fit in the overall picture in the Balkan sample, which lacks a uniform tendency in the direction of change.

In order to assess the significance of these observations, we provide kernel density estimates of the observed pairwise changes in five subsamples in Fig. 2, namely the Romance varieties (without Latin), all non-Romance varieties, all Germanic varieties, the sample of varieties from the British Isles (Dedio et al., 2019), and the sampled varieties from the Balkans. The median, maximal and minimal values, the median absolute deviation (MAD, a measure of variability) and the percentage of data points that are greater than 0 (i.e., show gain in similarity) are reported in Table 2.

The median trend towards gain of similarity is most pronounced in the sample from the British Isles (median .35, 80% of the data points > 0), followed by the Romance sample (median .18, 86% of the data points > 0), notably with the highest amount of data points > 0 and the smallest median absolute deviation (.08) across samples, hinting at a small, but robust trend towards gain in similarity. Unlike Romance, Germanic is dominated by divergence (median -.14, only 12% of the data points > 0). The Balkan sample has a quite flat distribution clustering around 0, indicating that there is ongoing, but unbiased change.

3.2 Trend and Magnitude of the Change of Similarity in Romance

The result of the Bernoulli model in Fig. 3 shows that the probability of gain in similarity (i.e., success) in 1000 random samples of Romance-Romance pairs

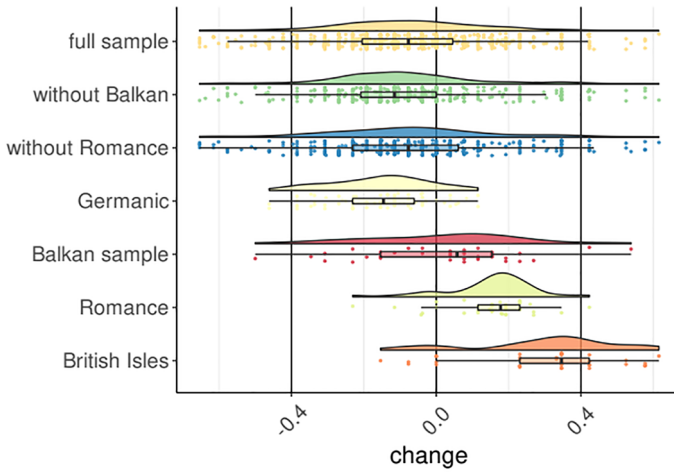


FIGURE 2 Raincloud plots and boxplots of the distribution of pairwise changes in the full sample and subsamples thereof; full sample = all languages in the sample; without Balkan = all languages except Balkan languages; without Romance = all languages except Romance languages; Germanic = only Germanic languages; Balkan sample = only Balkan languages; Romance = only Romance languages; British Isles = only languages from the British Isles.

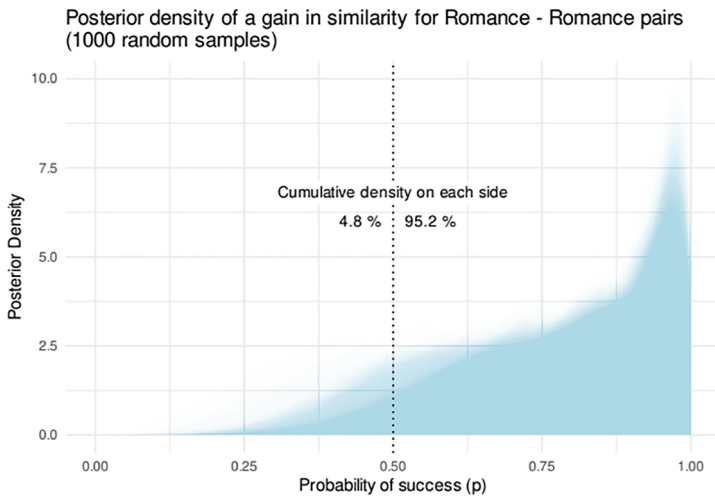


FIGURE 3 Posterior density of gain in 1000 random samples of Romance-Romance pairings. 95.2% of the samples display a gain of similarity.

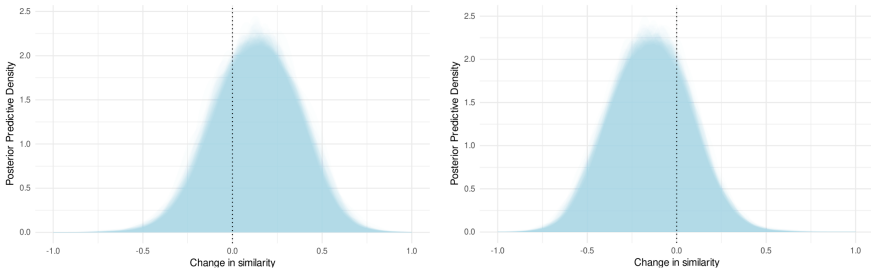


FIGURE 4 Posterior predictive distribution of changes in 1000 random samples. **Left panel:** Romance-Romance pairs (median = .11, 71% of the data > 0). **Right panel:** Germanic-Germanic pairs (median = -.15, 22% of the data > 0).

is 95.2%. This means that the Romance languages in our sample show a trend towards convergence in their expressions of reflexivity.

The magnitude of the process is estimated with a beta model. An overlay of all posterior predictive distributions of the beta model for Romance is shown in the left panel of Fig. 4 (median = 0.11; 71% of the data > 0) and for Germanic in the right panel (median = -0.15; 22% of the data > 0). The model predicts that in formerly unexplored Romance-Romance language pairs, there is an average gain of similarity of ca .1; for Germanic-Germanic pairs the average loss of similarity is .15. In practical terms, the posterior predictive distribution predicts that while there is a trend towards convergence, the actual average convergence in Romance amounts to a change in 1.5 variables out of 13.

4 Discussion

Towards identifying the input of the different factors contributing to changes in linguistic similarity, we examined samples of spatially and/or genealogically related language varieties between c. 800-500 ybp (ancestors) and present time (successors). We focused on Romance, a sample of languages which has a shared linguistic ancestry, stretches across contingent and non-contingent spaces, and partakes in a common cultural practice. To estimate the contributions of genealogy and contact in the Romance sample, we also contrasted a spatially defined sample of Balkan varieties with a sample from the British Isles.

4.1 Romance Sample

We found a small overall trend towards convergence across the varieties from the Romance lineage (see Section 3.1). The gain in similarity is smaller in

TABLE 3 Median, maximum, and minimum values of changes in all groupings, the MAD, and the percentage of data points with a positive value (i.e., gain); full sample = all languages in the sample; without Balkan = all languages except Balkan languages; without Romance = all languages except Romance languages; Germanic = only Germanic languages; Balkan sample = only Balkan languages; Romance = only Romance languages; British Isles = only languages from the British Isles.

grouping	median	maximum	minimum	MAD	data points > 0
full sample	-0.08	0.61	-0.65	0.19	30%
without Balkan	-0.11	0.61	-0.65	0.17	25%
without Romance	-0.08	0.61	-0.65	0.23	31%
Germanic	-0.14	0.11	-0.46	0.13	12%
Balkan sample	0.06	0.54	-0.5	0.20	61%
Romance	0.18	0.42	-0.23	0.08	86%
British Isles	0.35	0.61	-0.15	0.17	80%

our Romance sample than in the sample of varieties from the British Isles as reported in Dedio et al. (2019), and distinctly different from what we observe in Germanic as the other lineage-based sample. Also, the dynamics of change in Romance are quite different from the ones in our sample of non-Romance varieties, which has a rather flat distribution with more divergence than convergence.

Concerning the potential loss of similarity between non-Balkan and Balkan Romance, first sight evidence from Table 2 and Fig. 1 suggests that even though the Balkan Romance languages are not uniform in themselves, it is exactly these languages – in particular Aromanian and Judezmo – that spoil the otherwise quite uniform picture for Romance. It might thus be tempting to take this diverging behavior as attesting to the areal influence both have been exposed to. This is precisely what Friedman and Joseph (2014) suggest for Judezmo based on different data than ours, interpreting the divergences from Gallo- and Ibero-Romance as attesting to contact within the Balkan area. However, at least as concerns the part of grammar under consideration in this paper, the developments of features identified for Judezmo as compared to those identified for the Balkan sample do not reveal any similarities except for, maybe, the feature *POST*. This does not come as a surprise, given that Judezmo entered the Balkan area only in the 15th century, which also accounts for the considerable degree of divergence from Aromanian as the Romance variety that is most entrenched in the central Balkan area in our sample.

In a more general perspective, Romanian, Judezmo, and Aromanian, i.e., the Romance varieties that are not in direct contact with Gallo-, Italo-, and

Ibero-Romance, neither show significant convergences with the Balkan languages nor display a fundamentally aberrant behavior from the other Romance varieties (see also Joseph, 1999). This hints at lineage-internal contact not being the only relevant factor for the convergence in the Romance sample.

Moreover, the trend within Romance is stronger than in the non-Romance sample and in particular than in the subsample of varieties from the other lineage under consideration here, namely Germanic. It seems thus unlikely that these Romance developments attest to more general areal trends (e.g., those underlying larger presumed areas such as SAE). Unless the convergence in the sampled Romance varieties is mere coincidence or an artefact of our sampling choices, the findings suggests that either cultural and/or genealogical factors are at work and call for an explanation.

As for culturally induced convergence, it has been claimed that the literary national languages, which are well represented in our sample, were deliberately oriented toward Latin as prestigious role model (see Blatt, 1957; Pountain, 2010). This trend is assumed to have pertained to all evolving Romance literary languages in Europe. In our data and time span, this influence is not palpable in the case of Classical Latin, whose similarity to the Romance varieties decreases slightly or remains unchanged. Medieval Latin shows a minor gain in similarity with the Gallo-Romance varieties only, but even for Gallo-Romance this convergence is too small for it to provide conclusive evidence for a Latin influence.

As for the contribution of shared ancestry, it is notable that the Latin medio-passive conjugation was lost in the more or less simultaneous emergence of the Romance varieties in Late Antiquity and the early Middle Ages (Cennamo, 2016; Miller, 2010). The Romance varieties fell back on constructions that use pronominal elements of demonstrative or reflexive origin. During the span of time under investigation, these pronominals became ever more positionally dependent on the verb, lost the ability to establish their own stress domain, and started to increasingly interact with adjacent morphological material and assume passive functions, cf. Table 2. This shared evolutionary pathway corresponds to what is well known from grammaticalization of pronominals into markers of reflexivity (cf. Albanian *-u-* < **swe*, North Germanic *-s* < **sik*, Russian *-sja/-s'* < **sę* etc., cf. Matzinger, 2006: 120; Heine and Kuteva, 2002: 253; Vasmer, 1964: vol. 3: 823). On this cline of grammaticalization, the inherited Romance pronominals in reflexive functions lost syntactic and prosodic independence, cliticized to an already available host and became affixes eventually as in the fully grammaticalized and invariable reflexive prefix *se-* in Sursilvan Rumantsch (Spescha, 1989: 384); all of these developments are typical for grammaticalization processes (Lehmann, 2015). Concerning the hosting

site, Romance started out with a pre-/postverbal host position only (Wanner, 1987: 155) and it is in this inherited pre-/postverbal site that the Romance varieties subsequently hosted the cliticizing reflexive pronominals/markers. This inherited state also accounts for the differences in reflexive clitic placement between Romance and Slavic: unlike Romance, the Slavic varieties inherited both a Wackernagel and a pre-/postverbal host position for pronominal clitics and exploited them in different ways (Wackernagel position in BCMS, pre-/postverbal in Bulgarian, postverbal in Russian *-sja/-s'*; see Pancheva, 2005; Kuehnast, 2009; Franks, 2010).

Given the inherited state in Romance, it is not difficult to account for the emergence of passive functions in Gallo- and Italo-Romance either, since it is a well-known process of grammaticalization for reflexives to acquire passive functions as well (cf. Danish *-s*, Russian *-sja/-s'*, Heine and Kuteva, 2002: 253).

The shared ancestry of Romance with initially independent reflexive pronouns may thus represent an early stage of related (European and/or more general) grammaticalization clines (reflexive pronouns > reflexive marker; reflexive function > reflexive + passive function) that afforded systemic opportunities for converging developments. It seems that the time period covered by our investigation embraces exactly this cline of grammaticalization – which once again illustrates very clearly the relevance of the decisions made in the sampling choices and of the availability of data. Things are fundamentally different in the Balkan sample: the various lineages split long ago from their last common ancestor and there is no cross-lineage shared state at the beginning of the period under investigation.

To sum up, the observed lineage-internal convergence and maintenance of similarity in our sample of Romance varieties is most likely related to the rake-like speciation of the Romance languages. We assume that their shared inheritance from the common ancestor constrained the subsequent evolution so as to follow common paths of grammaticalization. To a yet unknown extent, this process was partly enhanced by contact induced innovation and spread (Gast and Auwera, 2012; Heine and Kuteva, 2003) mediated by mutual cultural influence, historical contingencies, and in part spatial proximity.

4.2 *Balkan Sample*

With regard to the sample of varieties from the Balkans, it is worth noting that the distribution of pairwise changes doesn't reveal a clear trend toward gain or loss of similarity, but there might be some temporal stratification. For example, no change into neither direction is found for Albanian and both Aromanian and Romanian, two Romance varieties attested long before the Ottoman rule, while with Judezmo, which came to the Balkans only during the Ottoman

times, a slight increase in similarity is observed. Overall, half of the data points center around 0, which hints at some degree of maintenance of similarity, and the dynamics add up to a rather low degree of overall change. Thereby, the Balkan sample clearly differs from the pattern observed in the sample from the British Isles (cf. Fig. 2). Note, however, that because of the contingencies of transmission, the time span covered in the Balkan sample is less uniform than in the other subsamples.

The kind of distribution detected here has been observed before and has variously led to conclude that the Balkan configuration embraces a zone of small-scale subareas of convergence with partially opposite trends across subareas (Joseph, 2010), or, alternatively, that the process of area formation on the Balkans is a phenomenon that belongs to the more remote past (Topolińska, 1995: 240). However, local convergences and Balkan-wide convergences are difficult to tell apart, and it has been argued that the invasion of the Avars and Slavs on the Balkans from the 6th c. onwards and the establishing of Ottoman rule from the mid-15th century onwards, both resulted in the dissolution of central administration and of administrative boundaries and provided a socio-linguistic situation favoring convergences and maintenance in the entire area (e.g., Lindstedt, 2000: 240; Friedman, 2011: 284). It would be premature to link the changes in our sample to any of the available theories – in fact, to some extent, our results are compatible with all of them – and after all, our sample and time span may simply be too small, or the features we looked at are not relevant for area formation in this case.

In more general terms, towards assessing the contributions and interactions of the various factors behind changes in similarity within a specific time span, for a specific language sample and a specific set of features, the state at the beginning of the period of time matters and needs to be identified for each language variety in the sample: it is an important confound and needs to be taken into account for estimating the role and degree of spatio-temporal and cultural factors and eventually getting a better sense of area formation processes.

5 Conclusion

Regarding the thirteen linguistic variables contributing to the expression of reflexivity considered in this paper, we discovered a maintenance and gain in similarity in the Romance sample in a period of time between the late Middle Ages and the present across ancestor and successor pairs. Given the manifold (family-external) contact situations provided by the wide geographical expansion of Romance, this convergence is rather unexpected. It also stands

in contrast to the development in another lineage, viz. Germanic, where we observe a decrease in similarity. However, it is strikingly similar to the development observed for the British Isles, a sample of languages of mixed genealogy, but becoming more similar because of contact.

On the basis of our data, the shared ancestry of the Romance languages emerges as the most likely source of the gain in similarity. The Romance sub-lineages branched off within a rather short span of time and inherited the same system of reflexive marking involving unbound pronominals. This shared inheritance constrained the development according to well-established pathways of grammaticalization, maintaining and even increasing the similarity, a development that runs counter to the divergence in other domains of grammar.

This case study leads us to conclude that when evaluating the impact of contact on the evolution of linguistic similarity, it is important to control for genealogical correlation between the sampled varieties and take into account the degree of relatedness in terms of time since speciation. Also, the dynamics of change in a given sample of languages depends on the state of each linguistic feature in each language at the beginning of the period of interest: completely different systems may converge due to contact, but so may very similar systems because of common pathways of grammaticalization.

As the data used in this study is a rather small set of variables from a rather small section of grammar, it is important to keep in mind that the obtained results cannot be used for generalizations without further investigation. What we hope to have provided is a proof of concept that when applied on a large set of linguistic data and a dense sample of language varieties, the approach taken here will likely enhance our understanding of the manifold interactions between drivers of gain, loss, and maintenance of linguistic similarity.

References

- Bickel, Balthasar, Kristine A. Hildebrandt, and René Schiering. 2009. The distribution of phonological word domains: A probabilistic typology. In Janet Grijzenhout and Barış Kabak (eds.), *Phonological Domains: Universals and Deviations*, 47–75. Berlin: De Gruyter.
- Blatt, Franz. 1957. Latin influence on European syntax. *Classica et Mediaevalia* 18: 133–178.
- Cennamo, Michela. 2014. Passive and impersonal reflexives in the Italian dialects. In Paola Benincà, Adam Ledgeway, and Nigel Vincent (eds.), *Diachrony and Dialects: Grammatical Change in the Dialects of Italy*, 71–95. Oxford: Oxford University Press. DOI: 10.1093/acprof:oso/9780198701781.003.0004.

- Cennamo, Michela. 2016. Voice. In Adam Ledgeway and Martin Maiden (eds.), *The Oxford Guide to the Romance Languages*, 967–980. Oxford: Oxford University Press. DOI: 10.1093/acprof:oso/9780199677108.003.0060.
- Cheetham, Alan H. and Joseph E. Hazel. 1969. Binary (presence-absence) similarity coefficients. *Journal of Paleontology* 43(5): 1130–1136.
- Dedio, Stefan, Peter Ranacher, and Paul Widmer. 2019. Evidence for the British Isles as a linguistic area. *Language* 95(3): 498–522.
- Fehrmann, Dorothee, Uwe Junghanns, and Denisa Lenertová. 2010. Two reflexive markers in Slavic. *Russian Linguistics* 34(3): 203–238. DOI: 10.1007/s11185-010-9062-7.
- Franks, Steven. 2010. Clitics in Slavic. *Glossos* 10: 1–157. https://slaviccenters.duke.edu/sites/slaviccenters.duke.edu/files/media_items_files/10franks.original.pdf.
- Freckleton, Robert P. and Walter Jetz. 2009. Space versus phylogeny: Disentangling phylogenetic and spatial signals in comparative data. *Proceedings of the Royal Society of London B: Biological Sciences* 276(1654): 21–30. DOI: 10.1098/rspb.2008.0905.
- Friedman, Victor A. 2011. The Balkan languages and Balkan linguistics. *Annual Review of Anthropology* 40(1): 275–291.
- Friedman, Victor A. and Brian D. Joseph. 2014. Lessons from Judezmo about the Balkan Sprachbund and contact linguistics. *International Journal of the Sociology of Language* (226): 3–23. DOI: 10.1515/ijsl-2013-0073.
- Gast, Volker and Johan van der Auwera. 2012. What is contact-induced grammaticalization?: Examples from Mayan and Mixe-Zoquean languages. In Björn Wiemer, Bernhard Wälchli, and Björn Hansen (eds.), *Grammatical Replication and Borrowability in Language Contact*, 381–426. Berlin: De Gruyter.
- Heine, Bernd and Tania Kuteva. 2002. *World Lexicon of Grammaticalization*. Cambridge: Cambridge University Press.
- Heine, Bernd, and Tania Kuteva. 2003. On contact-induced grammaticalization. *Studies in Language* 27(3): 529–572.
- Joseph, Brian D. 1999. Romanian and the Balkans: Some comparative perspectives. In Sheila Embleton, John E. Joseph, and Hans-Josef Niederehe (eds.), *The Emergence of the Modern Language Sciences. Studies on the transition from historical-comparative to structural linguistics in honour of E.F.K. Koerner, Volume 2: Methodological perspectives and applications*, 217–235. Amsterdam: John Benjamins.
- Joseph, Brian D. 2010. Language contact in the Balkans. In Raymond Hickey (ed.), *The Handbook of Language Contact*, 618–633. Oxford: Wiley-Blackwell.
- Kelly, Luke J. and Geoff K. Nicholls. 2017. Lateral transfer in Stochastic Dollo models. <https://arxiv.org/abs/1601.07931v3> (accessed March 16, 2017).
- Kuehnast, Milena. 2009. Processing clitic pronouns in Bulgarian. Evidence from normal and agrammatic comprehension. *Poznan Studies in Contemporary Linguistics* 45(4): 473–493. DOI: 10.2478/v10010-009-0029-z.

- Lehmann, Christian. 2015. *Thoughts on Grammaticalization*. 3rd ed. Berlin: Language Science Press.
- Lindstedt, Jouko. 2000. Linguistic balkanization: Contact-induced change by mutual reinforcement. *Studies in Slavic and General Linguistics* 28: 231–246.
- Matzinger, Joachim. 2006. *Der altalbanische Text Mbsuame e krështerë (Dottrina cristiana) des Lekë Matrënga von 1592: Eine Einführung in die albanische Sprachwissenschaft*. Dettelbach: Röhl.
- Miller, D. Gary. 2010. *Language Change and Linguistic Theory, Vol. 2: Morphological, Syntactic and Typological Change*. Oxford: Oxford University Press.
- Murawaki, Yugo and Kenji Yamauchi. 2018. A statistical model for the joint inference of vertical stability and horizontal diffusibility of typological features. *Journal of Language Evolution* 3(1): 13–25. DOI: 10.1093/jole/lzx022.
- Nelson-Sathi, Shijulal, Johann-Mattis List, Hans Geisler, Heiner Fangerau, Russell D. Gray, William Martin, and Tal Dagan. 2010. Networks uncover hidden lexical borrowing in Indo-European language evolution. *Proceedings of the Royal Society of London B: Biological Sciences*. DOI: 10.1098/rspb.2010.1917.
- Pancheva, Roumyana. 2005. The rise and fall of second-position clitics. *Natural Language and Linguistic Theory* 23: 103–167.
- Pountain, Christopher. 2010. Latin and the structure of written Romance. In Martin Maiden, John Charles Smith, and Adam Ledgeway (eds.), *The Cambridge History of the Romance Languages*, vol. 1, 606–659. Cambridge: Cambridge University Press. DOI: 10.1017/CHOL9780521800723.015.
- Ranacher, Peter, Stefan Dedio, and Paul Widmer. 2019. *Evidence for Britain and Ireland as linguistic Area: Supporting online material 2: R-script*. <https://drive.switch.ch/index.php/s/Gptpalz78TGcd8f/download>.
- Spescha, Arnold. 1989. *Grammatica sursilvana*. Cuera: Casa Editura per Mieds d’Instrucziun.
- Topolińska, Zuzanna. 1995. Convergent evolution, creolization, and referentiality. *Prague Linguistic Circle Papers* 1: 239–247.
- Trubetzkoy, Nikolai S. 1928. Proposition 16. *Actes du premier Congrès International de Linguistes: à La Haye du 10–15 avril 1928*, 17–18. Leiden: Sijthoff.
- Vasmer, Max. 1964. *Étimologičeskij slovar’ russkogo jazyka*. Moscow: Progress.
- Wanner, Dieter. 1987. *The Development of Romance Clitic Pronouns: From Latin to Old Romance*. Berlin: Mouton de Gruyter.
- Willems, Matthieu, Etienne Lord, Louise Laforest, Gilbert Labelle, François-Joseph Lapointe, Anna Maria Di Sciullo, and Vladimir Makarencov. 2016. Using hybridization networks to retrace the evolution of Indo-European languages. *BMC Evolutionary Biology* 16(1): 180. DOI: 10.1186/s12862-016-0745-6.