Revisiting Statistical Laws of Semantic Shift in Romance Cognates

Yoshifumi Kawasaki
The University of Tokyo
ykawasaki@g.ecc.u-tokyo.ac.jp

Maëlys Salingre
Mejiro University
m.salingre@mejiro.ac.jp

Marzena Karpinska
University of Massachusetts Amherst
mkarpinska@cs.umass.edu

Hiroya Takamura
AIST
takamura.hiroya@aist.go.jp

Ryo Nagata
Konan University
nagata-coling2022@ml.hyogo-u.ac.jp

Abstract

This article revisits statistical relationships across Romance cognates between lexical semantic shift and six intra-linguistic variables, such as frequency and polysemy. Cognates are words that are derived from a common etymon, in this case, a Latin ancestor. Despite their shared etymology, some cognate pairs have experienced semantic shift. The degree of semantic shift is quantified using cosine distance between the cognates' corresponding word embeddings. In the previous literature, frequency and polysemy have been reported to be correlated with semantic shift; however, the understanding of their effects needs revision because of various methodological defects. In the present study, we perform regression analysis under improved experimental conditions, and demonstrate a genuine negative effect of frequency and positive effect of polysemy on semantic shift. Furthermore, we reveal that morphologically complex etyma are more resistant to semantic shift and that the cognates that have been in use over a longer timespan are prone to greater shift in meaning. These findings add to our understanding of the historical process of semantic change.

1 Introduction

The Romance languages, such as present-day French, Italian, and Spanish, are sister languages which evolved from Vulgar Latin (Alkire and Rosen, 2010). They share numerous cognates, that is, words derived from a common etymon: for instance, the Latin verb HABERE "to have" developed into *avoir* (fr), *avere* (it), and *haber* (es).[1] Whereas French and Italian still maintain the original meaning "to have", it is no longer used as such in Spanish, where it is primarily used as an auxiliary to form perfect tense, while the notion of possession is generally expressed with *tener*. Likewise, the meanings and functions of some cognate pairs have diverged despite their common etymology. In the current paper, we investigate factors at play in semantic shift[2] using a computational approach.

In recent years, the analysis of diachronic semantic change using computational methods, inter alia, distributed representation of words, has gained increasing research interest (Dubossarsky et al., 2016; Hamilton et al., 2016; Takamura et al., 2017; Kutuzov et al., 2018; Uban et al., 2019; Hengchen et al., 2021; Kutuzov et al., 2021; Montariol and Allauzen, 2021; Schlechtweg et al., 2021; Tahmasebi et al., 2021; Uban et al., 2021a). The degree of lexical semantic shift has been conventionally quantified using cosine distance between relevant words, represented in the form of embeddings. A number of researchers have investigated the relationships between semantic change and frequency, polysemy, or prototypicality (Dubossarsky et al., 2015; Hamilton et al., 2016; Uban et al., 2019, 2021a). However, these effects have since been called into question (Dubossarsky et al., 2017).

Following this line of research, the present study revisits statistical relationships across Romance cognates between semantic shift and six intra-linguistic variables, including frequency and polysemy. For this purpose, we apply regression analysis and compare partial regression coefficients of individual variables while controlling for the others. By employing cognates, we are able to investigate whether particular trends hold across all explored languages. For the sake of simplicity, we assume

[1] Hereinafter, "fr" stands for French, "it" for Italian, and "es" for Spanish. Latin words are put in small capitals.

[2] We refer to the resultant difference in meaning between cognates as *shift*, while the term *change* is reserved for transition of a word's sense to another as well as for more general usage.

that their common ancestor, Latin, was uniform at the primitive stage before it developed into the Romance languages, although in reality Latin was not immune to variation across time, geography, or social stratification (Adams, 2007, 2013).

Our principal contributions are three: (i) amending flaws in the past research, we demonstrate that the law of conformity and the law of innovation (Hamilton et al., 2016) both hold for Romance cognates; (ii) exploring hitherto unexplored attributes of Latin etyma, we reveal that cognates whose etyma are morphologically more complex are more resistant to semantic shift; and (iii) considering temporal gap between words in lexical incorporation to a language, we show that words that have been in use over a longer timespan are prone to diverge more in meaning.

The rest of the paper is organized as follows. In Section 2, we review related research and point out its shortcomings. Section 3 describes our improved methodology. In Section 4 we present the experimental setup and results, followed by discussion in Section 5. Section 6 concludes the paper, pointing to future research directions.

2 Related Work

Two studies deserve special mention; both of them tackled statistical analysis of diachronic semantic change using word embeddings. A milestone in the field was Hamilton et al. (2016), who analyzed historical corpora in several languages and proposed two laws concerning semantic change: one is the *law of conformity*, according to which frequency is negatively correlated with semantic change; the other is the *law of innovation*, stating that polysemy is positively correlated with semantic change. However, these laws, as well as that of prototypicality (Dubossarsky et al., 2015), were revised (Dubossarsky et al., 2017) on the grounds that (i) the effect of frequency turned up even under control conditions, where no correlation was to be expected; and (ii) polysemy and prototypicality as they were defined by that research were highly collinear with frequency.

Inquiry into Romance cognates was first pursued by Uban et al. (2019) and later extended in Uban et al. (2021a). In contrast to Hamilton et al. (2016), Uban et al. (2019, 2021a) witnessed a positive correlation between frequency and degree of semantic shift; words that underwent more semantic shift tend to be more frequent. However, the experiments were conducted under non-comparable setups. Unlike Hamilton et al. (2016), who used cosine distance, Uban et al. (2019, 2021a) quantified the magnitude of semantic shift with what they denominated "falseness score." This value was calculated by subtracting the similarity score for a non-cognate translation pair (e.g., *long* "long" (fr) – *largo* "id." (es)) from the one for a cognate pair judged as false friends (e.g., *long* (fr) – *luengo* "id. (erudite wording)" (es), both tracing back to LONGU)[3]. *False friend* refers to either of two cognates in different languages that have diverged semantically despite their common etymology (Penny, 2002). Focusing on false friend pairs inevitably translates into exclusion of the cognate pairs that preserve commonality in meaning; this would prevent us from gaining insights on general tendencies across overall cognate pairs. In addition, the use of falseness scores was hardly justified. Accordingly, it is imperative to settle the dispute over the diametric effect of frequency by implementing experiments under identical conditions, such that genuine cognate pairs as well as false friend pairs are covered.

Furthermore, there is still room for improvement. First, Uban et al. (2021a) defined polysemy of a Romance word as the number of synsets in WordNet that the word is part of. Thus, the polysemy score of a cognate pair was computed as the average of number of synsets for the two words in question. However, if the ultimate goal is to determine factors that could induce semantic shift, it is more appropriate to consider polysemy prior to eventual semantic shift. Second, Uban et al. (2021a) leveraged publicly available multilingual word embeddings (Conneau et al., 2018). Notwithstanding their usefulness, however, there is a risk that the embeddings are undesirably affected by the alignment algorithm employed; if a cognate pair is aligned in advance, then it naturally becomes more similar than it should be.

3 Methods

3.1 Overview

To circumvent the series of methodological defects adduced above, we took the following measures. (i) In the same spirit as Hamilton et al. (2016), we examined all the cognate pairs and measured degree of semantic shift with cosine distance score.

[3] *Id.* (idem) represents having parallel sense to its counterpart.

Then we applied linear regression analysis to explore statistical relationships between the degree of semantic shift and six intra-linguistic variables specified in Section 4.2. (ii) Independently from frequency counts, we defined polysemy as the number of word sense entries of Latin etyma in a dictionary. The use of Latin polysemy has the added advantage of yielding a polysemy score for the earliest stage of evolution, prior to semantic shift. (iii) Before undertaking our main analysis, we ensured the absence of a priori correlation between frequency and semantic affinity for random word pairs, which are presumed to present low similarity. (iv) We created our own cognates list from scratch to procure fit-for-purpose vector representation of words, as described in the next section[4].

3.2 Construction of multilingual embeddings

We limited the scope of study to three Romance languages: French, Italian, and Spanish. We used a Wikipedia dump as of December 2018 to acquire static word embeddings separately in each language. The relative uniformity of Wikipedia in style and topic is a suitable property to ensure, as much as possible, that any difference stemming from comparison of cognate embeddings is imputable to the nature thereof instead of the corpus on which training was performed. Data cleaning was carried out using `Wiki-cirrus`[5] for French and `Wikiextractor`[6] for Italian and Spanish. Then we performed lemmatization with `TreeTagger`[7]. No distinction was made between homonyms, due to their paucity.

Target words for analysis were restricted to nouns, adjectives, and verbs. To prevent low frequency from disturbing the reliability of the obtained embeddings, we opted to focus on highly frequent cognates. Specifically, a separate cognates list was created for each word class, such that at least one member of a cognate set was found among the most frequent 300 lemmas in any of the three languages; it was possible for a cognate to rank among the top 300 in one language while its counterpart did not in another. The number of cognate pairs added up to 487 for nouns, 477 for adjectives, and 493 for verbs, although the cognates were not always found in all three languages. Cognate identification was implemented by looking into etymological description in dictionaries[8].

Unilingual embeddings were learned by `word2vec` in the `gensim` library (Řehůřek and Sojka, 2010)[9]. The default hyperparameters were adopted, with the exception of `vector_size=600` and `min_count=50`. To compare word embeddings in various languages, we need to align the separately obtained embeddings in a common cross-lingual vector space. The alignment was executed using the supervised method of MUSE (Conneau et al., 2018)[10], that is, the linear mapping proposed by Mikolov et al. (2013a) with the orthogonality constraint, such that the sum of the squared errors across the inter-lingual seed pairs was as low as possible. The seed pairs with an acknowledged parallel meaning between the two languages were retrieved from MUSE bilingual dictionaries (Conneau et al., 2018)[11]; the cognate pairs that are targets of our analysis were eliminated therefrom. In any combination of languages, seed pairs amounted to about ten thousand. The choice of the language onto which the embeddings of other languages were mapped had marginal effect on the subsequent analysis. In the following, we report results obtained from French–Spanish cognate pairs aligned in French vector space for illustrative purposes. See Section 5.3 for discussion regarding different embedding spaces and language combinations.

3.3 Validation of embeddings

Following the commonplace procedure, the pairwise similarity between cognates was defined as the cosine similarity score between the corresponding embeddings. Based on the distributional hypothesis (Harris, 1954; Firth, 1957), word embeddings are claimed to capture lexical meaning to a certain degree. Before going any further, it is pivotal to assess the quality of the acquired embeddings to ensure the soundness of our principal analysis. To this end, we conducted the following two experiments.

[4] The data are available upon request.

[5] `https://www.mediawiki.org/wiki/Help:CirrusSearch`

[6] `https://github.com/attardi/wikiextractor`

[7] `https://www.cis.lmu.de/~schmid/tools/TreeTagger/`

[8] French: `https://www.cnrtl.fr`
Italian: `https://www.etimo.it`
Spanish: `https://dle.rae.es`

[9] `https://radimrehurek.com/gensim/models/word2vec.html`

[10] `https://github.com/facebookresearch/MUSE`

[11] `https://github.com/facebookresearch/MUSE`

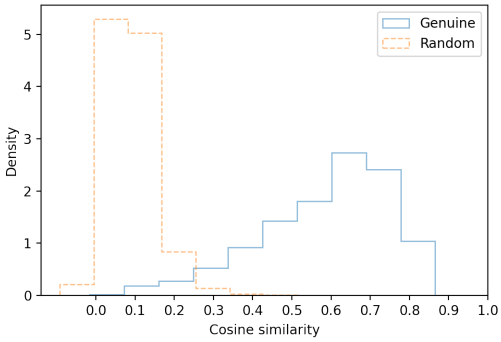

Figure 1: Density distribution of pair-wise similarity scores for ca. 4K genuine seed pairs and 1000K random word pairs.

First, we analyzed the pair-wise similarity distribution of the seed pairs that were employed when aligning word embeddings; it is essential that they present high similarity scores. For this, we chose seed pairs each word of which appears more than ten thousand times in the corpus of the relevant language. This threshold was set for fear that the embeddings of infrequent words might be under-learned. The number of seed pairs selected amounted to nearly four thousand for any language pair. The histogram with solid contour line in Figure 1 illustrates the density distribution of similarity scores for the frequent seed pairs. As expected, they show relatively high similarity, with mean of 0.66 and standard deviation of 0.15.

Next, we repeated the operation with 100,000 randomly generated pairs from among the seed words. These random pairs ought to show minimal similarity scores. The histogram with dotted contour line in Figure 1 illustrates the density distribution. Unsurprisingly, the mean similarity score that resulted was low at 0.10, with standard deviation of 0.06. It is worth noting that a similarity score of 0.1 can arise in our case, even when given two unrelated words.

Inasmuch as the validity of the obtained embeddings is confirmed, we can safely make use of them on the premise that cosine similarity reflects semantic affinities reasonably, if not completely. Table 1 provides examples of the five most and five least similar French–Spanish cognate pairs. A smaller similarity score means a greater degree of semantic shift, and vice versa.

French	Spanish	Sim.
construire “to construct”	constuir “id.”	0.87
provoquer “to provoke”	provocar “id.”	0.87
évêque “bishop”	obispo “id.”	0.86
détruire “to destroy”	destruir “id.”	0.86
féminin “feminine”	femenino “id.”	0.86
avoyer “to set saw”	aviar “to prepare”	0.05
atteindre “to reach”	atañer “to pertain”	0.05
mener “to take”	menar “to turn”	0.04
saison “season”	sazón “seasoning”	0.02
maire “mayor”	mayor “bigger, older”	0.00

Table 1: Most (upper half) and least (lower half) similar five French–Spanish cognate pairs.

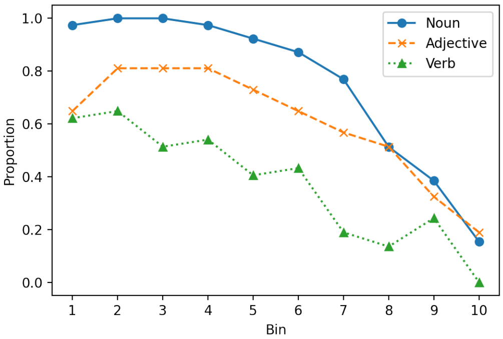

Figure 2: Proportion of cognate pairs enumerated in MUSE as translation pairs in ten equally divided bins for respective word classes. The leftmost bin corresponds to the top 10% translation pairs that are most similar, the second bin to the following 10%, and so on in the same way.

To make an overall evaluation of the computed affinity scores, we inspected the proportion of the cognate pairs that are enumerated in MUSE as translation pairs across ten equally divided bins for respective parts-of-speech. The result is shown in Figure 2. The leftmost bin corresponds to the top 10% translation pairs that are most similar, and the second bin to the following 10%, continuing in the same way. We can observe that, moving rightwards, the drop in proportion goes roughly in line with decreasing similarity across the word classes; the most similar cognate pairs are mostly among the inter-lingual translation pairs, whereas the dissimilar pairs, for the most part, are outside of them and thus regarded as false friend pairs.

Moreover, we notice that the proportion varies across the parts-of-speech. Specifically, the verbs consistently present the lowest proportion, which is indicative of this word class being more inclined to shift, followed by adjectives and nouns. This tendency partially conforms to the conclusion arrived at by Dubossarsky et al. (2016), except that they discovered an inverse order between nouns and adjectives.

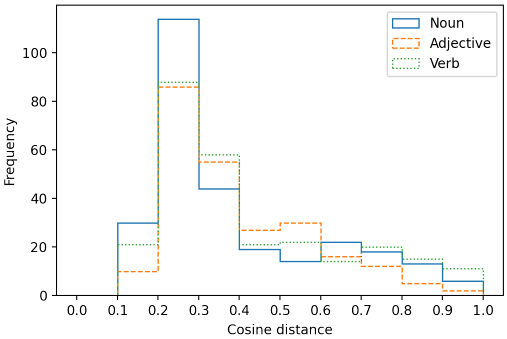

Figure 3: Distribution of cosine distance scores for French–Spanish cognate pairs in different word classes.

4 Regression Analysis

We applied linear regression analysis to explore statistical relationships between the degree of semantic shift and six intra-linguistic factors, detailed below. For the pair-wise scatter plots between the standardized variables, see Appendix A.

4.1 Dependent variable

The dependent variable $\textsc{dist}_{rom}$ is the log-transformed pair-wise cosine distance score between cognates, defined as $\log(dist_{cos}(cog1, cog2)) = \log(1 - \cos(cog1, cog2))$, where $\cos(cog1, cog2)$ stands for cosine similarity of a cognate pair $(cog1, cog2)$. This variable represents the magnitude of semantic shift that the cognate pair in question underwent. Figure 3 illustrates the distribution of cosine distance scores for French–Spanish cognate pairs in different word classes. The distribution is apparently skewed to the right in all classes, with mean of 0.39 and median of 0.29 for nouns, 0.40 and 0.34 for adjectives, and 0.42 and 0.33 for verbs. A rather long right tail implies a large number of dissimilar pairs.

4.2 Independent variables

We established six intra-linguistic factors as independent variables. The first three relate to Latin etyma and the remaining three to Romance descendants. Despite their potential contribution to semantic change, extra-linguistic variables such as socio-cultural, historical, political, technological factors as well as language contact were outside our scope (Penny, 2002; Newman, 2015; Hamilton et al., 2016; Dubossarsky et al., 2017).

4.2.1 Latin features

We leveraged three features concerning Latin etyma: frequency, polysemy, and word length. These are variables that relate to the phase prior to semantic shift and have been ignored in past studies (Uban et al., 2019, 2021a). For nouns and adjectives, we deemed the Latin accusative to be the etymological form, for it is the case from which Romance descendants are typically derived (Alkire and Rosen, 2010). As is conventionally done, word-final -M was omitted if applicable (e.g., we adopted ANNU for ANNUM "year"). For verbs, the infinitive was used for convenience, as the representative form for Latin as well as for the Romance languages. With respect to the infinitival form of Latin deponent verbs (Oniga, 2014), we employed reconstructed active forms of the corresponding conjugation class (e.g., JOCARI "to joke" was transformed into *JOCARE). As for irregular verbs including ESSE "to be", FERRE "to carry", and VELLE "to want", we used original infinitival forms.

Frequency of Latin etymon ($\textsc{freq}_{lat}$) This is defined as the log-transformed relative frequency of the Classical Latin etymon per 10,000 words retrieved from the online database on classical languages PhiloLogic4[12]. Even though the Romance languages did not evolve from Classical Latin (written language), but Vulgar Latin (spoken language) (Alkire and Rosen, 2010), since there was no resource available for the latter, we utilized the materials at hand on the literary language. The same applies to other Latin features below.

Latin polysemy ($\textsc{poly}_{lat}$) This is defined as the log-transformed number of word sense entries for the Latin etymon in the *Oxford Latin Dictionary* (Glare, 2012). Note, however, that taking logarithm did not remedy the right-skewed distribution.

Word length of Latin etymon ($\textsc{len}_{lat}$) This is defined as the number of characters in the Latin etymon. Longer words tend to be morphologically more complex due to affixation or derivation, which in turn helps restrict the semantic range of the base form. For example, PRAEVIDERE "to foresee", which is constructed from a base VIDERE "to see"

[12] http://artflsrv02.uchicago.edu/cgi-bin/perseus/LatinFrequency.pl?author=&title=&genre=&editor=&language=NOT+English&displaymorethan=10&displaylessthan=10000000&sortby=decreasingFreq&searchby=searchbylemma (as of January 28th, 2021).

and a prefix PRAE- "ahead", exhibits a more restricted sense than its base. In fact, the similarity score between *prévoir* (fr) and *prever* (es), derived from PRAEVIDERE, is as high as 0.72, while it is 0.35 between *voir* (fr) and *ver* (es), derived from VIDERE.

4.2.2 Romance features

The following four variables relate to the phase by which semantic shift occurred in relevant cases.

Mean frequency of Romance cognate pair (FREQ$_{rom}$) Defined as the log-transformed harmonic mean $\log \mathrm{Harmonic}(freq_{cog1}, freq_{cog2})$, where $freq_w$ is the relative frequency of a word w in the corresponding corpus, and $cog1$ and $cog2$ are cognate words. We found it more opportune to use the harmonic mean than the arithmetic mean in handling average ratio, thanks to its property of being biased toward the smaller value. The merit of using the harmonic mean was empirically verified in the following regression analysis.

To ensure that there is no a priori correlation between semantic shift and pair-wise mean frequency, we examined 100,000 randomly generated word pairs from among the seed words used in aligning embeddings. As expected, only a negligible correlation was seen, of -0.06, which signifies that our analysis is practically free from the spurious effect of frequency (Dubossarsky et al., 2017).

Mean embeddings' norm in Romance cognate pairs (NORM$_{rom}$) This is defined as the arithmetic mean over the embeddings' norm of the words that compose a cognate pair. We took into account the norm, which is a hitherto ignored attribute of embeddings, with the aim of assessing if its effect is detected independently from frequency and polysemy; the algorithm underlying the skip-gram model with negative sampling (Mikolov et al., 2013b) entails that the norm of a word embedding grows large when its frequency is high and is oligosemous, being used consistently in analogous contexts.

Mean edit distance between Latin etymon and Romance cognate pair (EDIT) This is defined as the arithmetic mean over the normalized edit distances between Latin etymon l and each word forming a cognate pair (r, r'), as follows:

$$\frac{1}{2}\left(\frac{edit(r,l)}{\frac{|r|+|l|}{2}} + \frac{edit(r',l)}{\frac{|r'|+|l|}{2}}\right), \qquad (1)$$

where $edit(\cdot)$ is the unnormalized edit distance, which is normalized with division by mean word length between l and r (Levenshtein, 1966), and $|\cdot|$ represents word length. Although it is desirable to make phonetic comparisons, we instead quantified graphical displacement for the sake of simplicity.

We argue that edit distance can be viewed as a proxy for how long the words have been used in a language or when they came to form the language's lexica; words with large edit distances with respect to Latin etymon are regarded as *inherited words*, which underwent typical phonetic alterations through oral transmission in the respective descendant languages and thus ended up having a distinct appearance from their ancestor (Penny, 2002). Conversely, the words with small edit distances correspond to *learned words*, which were relatively recently borrowed into the Romance languages with minimal phonetic alterations from the Medieval Latin (Penny, 2002). A dichotomous distinction between inheritance and recent borrowing was introduced by Uban et al. (2021b), who made the distinction by consulting dictionaries. Our novelty consists in quantifying the difference in an automatic manner using edit distance.

Prior to computing edit distance, all the diacritics were removed from the vowels. For example, *á* and *à* were both transformed into *a*. For sake of simplicity, the consonants were left intact despite the fact that some characters, such as Spanish *ñ* and French *ç*, did not exist in Latin, and that phonetic values of some letters are not identical between Latin and the Romance languages: for instance, CITARE [k] "to set in motion" in contrast to *citar* (es) [θ] "to cite".

Table 2 presents five French–Spanish cognate pairs with the largest (upper half) and the smallest (lower half) edit distance scores. We readily notice the aforesaid tendency; the cognate pairs with the largest edit distances correspond to inherited words with various modifications, while those with the smallest edit distances belong to learned vocabulary that virtually maintains its etymological form.

4.3 Setup

The dataset comprised those cognate pairs that bear Romance forms in the two languages of interest and for which information on frequency and polysemy of the Latin etymon was available; we ignored cognates whose etyma do not trace back to

Latin	French	Spanish	Dist.
CAPUT "head"	chef	jefe	1.00
VICE "time"	fois	vez	0.93
AURU "gold"	or	oro	0.93
RUSSU "red"	roux	rojo	0.89
EPISCOPU "bishop"	évêque	obispo	0.86
BASE "base"	base	base	0.00
ENORME "enormous"	énorme	enorme	0.00
SERIE "series"	série	serie	0.00
CELEBRE "busy"	célèbre	célebre	0.00
ANIMAL "animal"	animal	animal	0.00

Table 2: Five French–Spanish cognate pairs with largest (upper half) and smallest (lower half) edit distances.

Latin. This constraint almost halved the number of effective cognate pairs, down to 281 for nouns, 243 for adjectives, and 270 for verbs, hence, 794 in total.

Prior to performing regression analysis, all the variables were standardized to have zero mean and unit variance, allowing for directly comparing the scale of regression coefficients. Note that the log-transformed variables scale according to the power laws with the degree of semantic shift. We did not log-transform LEN$_{lat}$, NORM$_{rom}$, or EDIT, since these did not exhibit a right-skewed distribution; nor, at this point, did we distinguish among different word classes. See Section 5.2 for a survey discriminating them. In spite of a large correlation of 0.73 between FREQ$_{lat}$ and POLY$_{lat}$, we did not exclude either one, for their variance inflation factors were less than three. We implemented model selection methods in terms of AIC values (Akaike, 1974).

4.4 Results

The best model chosen, which only dropped NORM$_{rom}$, is summarized in Table 3. All the retained covariates were statistically significant at 0.05 level except FREQ$_{lat}$[13]. Standardized partial regression coefficients turned negative for FREQ$_{rom}$ (-0.54) and LEN$_{lat}$ (-0.21), and positive for POLY$_{lat}$ (0.10) and EDIT (0.13). The adjusted R-squared of 0.35 means that 35% of the total variance was accounted for by the model[14].

5 Discussion

5.1 Independent variables

Based on the outcome of the regression analysis, we discuss how individual explanatory variables correlate with the degree of semantic shift.

	Coef.	SE	t	$p > \|t\|$
Intercept	0.00	0.03	0.00	1.00
FREQ$_{lat}$	-0.08	0.04	-1.82	0.07
POLY$_{lat}$	**0.10**	**0.04**	**2.28**	**0.02**
LEN$_{lat}$	**-0.21**	**0.03**	**-6.29**	**0.00**
FREQ$_{rom}$	**-0.54**	**0.03**	**-18.40**	**0.00**
NORM$_{rom}$	–	–	–	–
EDIT	**0.13**	**0.03**	**4.07**	**0.00**

Table 3: Results of regression analysis on distance scores of French–Spanish cognate pairs ($N = 794$, $Adj.R^2 = 0.35$). NORM$_{rom}$ was kept out by model selection methods.

FREQ$_{lat}$ This predictor did not result statistically significant; its effect might have been absorbed by POLY$_{lat}$ and LEN$_{lat}$, which it is collinear with.

POLY$_{lat}$ A positive coefficient (0.10) denotes that the cognates with a more polysemous etymon tend to undergo a larger semantic shift. This finding agrees with the insights of Hamilton et al. (2016) and Uban et al. (2019, 2021a) that polysemy is positively correlated with semantic shift. Nonetheless, it should be highlighted that this finding hints at a causal effect, because we exploited the polysemy at the initial stage of linguistic development. That said, why polysemy leads to larger semantic shift requires explanation. We argue that, even when the rate of semantic change is constant per given time unit, polysemous words are more likely to digress into diverse directions in each language: for example, TRAHERE "to drag", which possessed 22 word sense entries, developed different senses in the daughter languages: *trarre* "to draw" (it), *traire* (fr) "to milk", and *traer* (es) "to bring". Thus, these cognates exhibit large distance scores, around 0.85.

LEN$_{lat}$ A negative coefficient (-0.21) backs up our hypothesis that the cognates with a longer Latin etymon tend to undergo smaller semantic shift. We conjecture that longer words are less susceptible to meaning shift, because they tend to have restricted senses, which indirectly supports the law of innovation (Hamilton et al., 2016). Actually, there is a negative correlation of -0.34 between word length and polysemy in Latin (Appendix A).

FREQ$_{rom}$ A negative coefficient (-0.54) sustains the alleged law of conformity (Hamilton et al., 2016) and, in turn, dismisses the opposite conclusion reached by Uban et al. (2019, 2021a). The largest absolute coefficient indicates its significant

[13] All subsequent significance tests are at $p < 0.05$.

[14] Henceforth, adjusted R-squared is abbreviated $Adj.R^2$.

contribution to the model's predictive power.

As to the negative correlation between frequency and semantic shift, we suppose that frequently used words are entrenched enough to resist semantic shift (Bybee, 2015). However, we cannot assert this with full confidence, because FREQ$_{rom}$ relates to a time point posterior to semantic shift, and so the observed frequency might well be the fruit of semantic shift.

It is also noteworthy that $Adj.R^2$ diminished considerably from 0.35 to 0.11 when using an arithmetic mean instead of a harmonic mean. The property of the harmonic mean of being biased toward the smaller value could be beneficial in cases where only one of the cognates has suffered semantic change, accompanied by considerable variation in frequency, and where consequently they became dissimilar.

NORM$_{rom}$ That the model kept out this predictor implies that its effect is incapable of being detected independently from frequency and polysemy.

EDIT A positive coefficient (0.13) confirms our hypothesis that inherited words tend to go through larger semantic shift.

5.2 Effect of word class

When fitting the model separately for separate word classes, we found a slight variation in $Adj.R^2$: 0.38 for nouns, 0.31 for adjectives, and 0.49 for verbs. This unequal behavior might imply that nouns are more susceptible than verbs, and adjectives more than nouns, to extra-linguistic factors disregarded in this study: socio-cultural circumstances, technological advances, language contact, metaphorical extensions, to name a few. As to the best models selected, LEN$_{lat}$ and FREQ$_{rom}$ were retained and were statistically significant across the parts-of-speech, exhibiting comparable coefficients; therefore, the effect of these variables appear to be extensive. In contrast, the remaining variables displayed differing behaviors from one class to another.

5.3 Consequences of aligned embedding spaces and language combinations

Thus far, for illustrative purposes, we have solely focused on French–Spanish cognate pairs aligned in French embedding space. It is indispensable also to assess the potential consequences that different embedding spaces and language combinations could bring about. There are nine settings in total; three language pairs for each of three embedding spaces. A close examination reveals that (i) different embedding spaces have practically null effect on the outcome of regression analysis; and (ii) different language pairings, in turn, slightly affect the composition of the best model, the scale of regression coefficients, and accordingly $Adj.R^2$.

Embedding space	Language pair	$Adj.R^2$	N
French	French–Italian	0.29	812
	French–Spanish	0.35	794
	Italian–Spanish	0.35	842
Italian	French–Italian	0.29	812
	French–Spanish	0.33	794
	Italian–Spanish	0.38	842
Spanish	French–Italian	0.27	812
	French–Spanish	0.35	794
	Italian–Spanish	0.39	842

Table 4: Adjusted R-squared for respective language pairs in different embedding spaces.

Table 4 is the summary of $Adj.R^2$ for respective language pairs in different embedding spaces in which the alignment was done. It is noteworthy that, in every embedding space, the score is consistently largest for the Italian–Spanish pair, followed by French–Spanish and French–Italian pairs. This suggests that some unconsidered variables are more at play for the French–Italian pair than for the French–Spanish pair, and more for French–Spanish pair than for Italian–Spanish pair.

Figure 4 depicts radar charts presenting regression coefficients of the best-fitted model for respective language pairs. The embeddings were aligned in the vector space that the legend indicates. A marker at the origin represents rejection of the variable in question. The dashed circle represents coefficients being equal to zero. The intercept was omitted for having a value almost equal to zero in every case. Regression analysis was run without parts-of-speech distinction. The almost overlapping lines demonstrate that the general trends commented above hold across the different settings, although not without exceptions: (i) NORM$_{rom}$ was retained for Italian–Spanish pairs in any embedding space, while FREQ$_{lat}$ was dropped. Considered in conjunction, the effects of FREQ$_{lat}$ might have been offset by that of NORM$_{rom}$, thereby exhibiting a negative correlation with semantic shift ; and (ii) EDIT was accidentally dropped in French–Spanish combinations in Italian embedding space for unknown reasons.

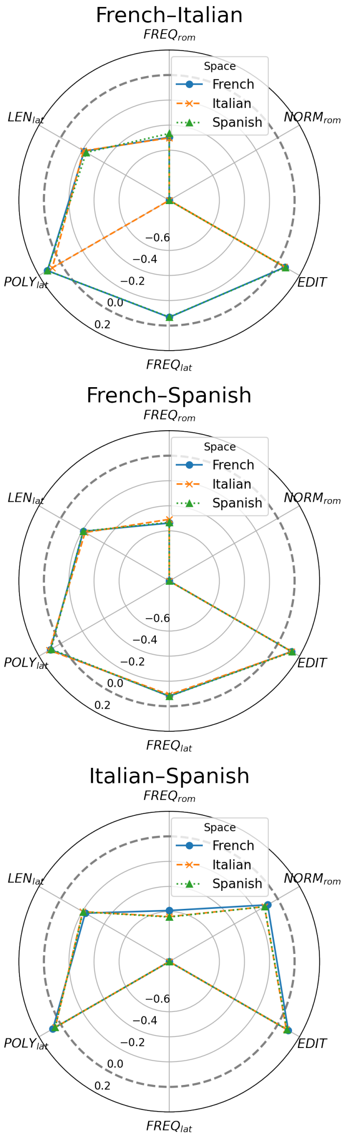

Figure 4: Regression coefficients of the best-fitted model for respective language pairs aligned in different embedding spaces.

5.4 A posteriori prediction at Latin era

It is an intriguing question how well one can make prediction on semantic shift. With this objective, we performed an additional regression analysis exploiting only Latin features. Model selection was performed with AIC. As Table 5 shows, FREQ$_{lat}$ and LEN$_{lat}$ were retained and became statistically significant. The fitted coefficients imply that frequent and long Latin etyma tend to undergo less semantic shift, which directly underpins the law of conformity and indirectly the law of innovation (Hamilton et al., 2016), if it is appropriate to associate word length with polysemy. Although barely 6% of the variance was explained, we found this phenomenon in accordance with the reflection that "small effects may be a priori more credible than large ones" (Dubossarsky et al., 2017).

	Coef.	SE	t	$p > \|t\|$
Intercept	0.00	0.03	0.00	1.00
FREQ$_{lat}$	**-0.10**	**0.04**	**-2.74**	**0.01**
POLY$_{lat}$	–	–	–	–
LEN$_{lat}$	**-0.27**	**0.04**	**-7.10**	**0.00**

Table 5: Results of regression analysis on distance scores of French–Spanish cognate pairs ($N = 794$, $Adj.R^2 = 0.06$). Only Latin features were employed. POLY$_{lat}$ was kept out by model selection methods.

6 Conclusions

In this study, we revisited statistical relationships between semantic shift and intra-linguistic variables across the Romance languages. Our principal contributions are three: (i) we demonstrated that the law of conformity and the law of innovation (Hamilton et al., 2016) both hold for Romance cognates by amending flaws in the past research; (ii) we revealed that cognates whose etyma are morphologically more complex are more resistant to semantic shift by exploring hitherto unexplored attributes of Latin etyma; and (iii) we showed that words that have been in use over a longer timespan are prone to diverge more in meaning by considering temporal gap in lexical incorporation to a language.

One limitation of our study is a lack of comparison with Latin. Without it, we would not have a complete picture of historical semantic change; we need to understand how much and in what direction it has evolved. To meet this aim, it is a sine qua non to obtain Latin embeddings trained upon Classical Latin sources, thereby allowing for direct comparisons with the Romance ones. Also, it will be an interesting extension to elucidate types of semantic change that have occurred in relevant cases (Kutuzov et al., 2018), such as specialization, generalization, melioration, and pejoration (Traugott and Dasher, 2005). In addition, we need to address qualitative difference in meaning, including typicality, hypernymy, and hyponymy, since we did not go further in this paper than to define polysemy as the number of word sense entries in the dictionary. Such an improvement could be achieved by taking advantage of information in WordNet.

Acknowledgements

This work was supported by JSPS KAKENHI Grant Number JP18K11456.

References

James Noel Adams. 2007. *The Regional Diversification of Latin 200 BC-AD 600*. Cambridge University Press, New York.

James Noel Adams. 2013. *Social Variation and the Latin Language*. Cambridge University Press, New York.

Hirotugu Akaike. 1974. A New Look at the Statistical Model Identification. *IEEE Transactions on Automatic Control*, 19(6):716–723.

Ti Alkire and Carol Rosen. 2010. *Romance Languages: A Historical Introduction*. Cambridge University Press, New York.

Joan Bybee. 2015. *Language Change*. Cambridge University Press, Cambridge.

Alexis Conneau, Guillaume Lample, Marc'Aurelio Ranzato, Ludovic Denoyer, and Hervé Jégou. 2018. Word Translation Without Parallel Data.

Haim Dubossarsky, Yulia Tsvetkov, Chris Dyer, and Eitan Grossman. 2015. A Bottom Up Approach to Category Mapping and Meaning Change. In *Word Structure and Word Usage*, pages 66–70, Pisa.

Haim Dubossarsky, Daphna Weinshall, and Eitan Grossman. 2016. Verbs Change More Than Nouns: A Bottom-up Computational Approach to Semantic Change. *Lingue e Linguaggio*, 15(1):5–25.

Haim Dubossarsky, Daphna Weinshall, and Eitan Grossman. 2017. Outta control: Laws of semantic change and inherent biases in word representation models. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, pages 1136–1145, Copenhagen, Denmark. Association for Computational Linguistics.

J.R. Firth. 1957. A Synopsis of Linguistic Theory 1930-55. *Studies in Linguistic Analysis: Special Volume of the Philological Society*, pages 1–32.

P.G.W. Glare, editor. 2012. *Oxford Latin Dictionary*, 2nd edition. Oxford University Press, Oxford.

William L. Hamilton, Jure Leskovec, and Dan Jurafsky. 2016. Diachronic word embeddings reveal statistical laws of semantic change. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 1489–1501, Berlin, Germany. Association for Computational Linguistics.

Zellig S. Harris. 1954. Distributional Structure. *WORD*, 10(2-3):146–162.

Simon Hengchen, Nina Tahmasebi, Dominik Schlechtweg, and Haim Dubossarsky. 2021. Challenges for Computational Lexical Semantic Change. In Nina Tahmasebi, Lars Borin, Adam Jatowot, Yang Xu, and Simon Hengchen, editors, *Computational Approaches to Semantic Change*, chapter 11, pages 341–372. Language Science Press, Berlin.

Andrey Kutuzov, Lilja Øvrelid, Terrence Szymanski, and Erik Velldal. 2018. Diachronic word embeddings and semantic shifts: a survey. In *Proceedings of the 27th International Conference on Computational Linguistics*, pages 1384–1397, Santa Fe, New Mexico, USA. Association for Computational Linguistics.

Andrey Kutuzov, Lidia Pivovarova, and Mario Giulianelli. 2021. Grammatical profiling for semantic change detection. In *Proceedings of the 25th Conference on Computational Natural Language Learning*, pages 423–434, Online. Association for Computational Linguistics.

Vladimir Iosifovich Levenshtein. 1966. Binary Codes Capable of Correcting Deletions, Insertions and Reversals. *Soviet Physics Doklady*, 10(8):707–710.

Tomás Mikolov, Quoc V. Le, and Ilya Sutskever. 2013a. Exploiting Similarities among Languages for Machine Translation. *CoRR abs/1309.4*.

Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg Corrado, and Jeffrey Dean. 2013b. Distributed Representations of Words and Phrases and Their Compositionality. In *Proceedings of the 26th International Conference on Neural Information Processing Systems*, volume 2 of *NIPS'13*, pages 3111–3119, Red Hook, NY, USA. Curran Associates Inc.

Syrielle Montariol and Alexandre Allauzen. 2021. Measure and evaluation of semantic divergence across two languages. 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)*, pages 1247–1258, Online. Association for Computational Linguistics.

John Newman. 2015. *Semantic Shift*, 1st edition, pages 266–280. Routledge, New York.

Renato Oniga. 2014. *Latin: A Linguistic Introduction*. Oxford University Press, New York.

Ralph Penny. 2002. *A History of the Spanish Language*. Cambridge University Press, Cambridge.

Dominik Schlechtweg, Barbara McGillivray, Simon Hengchen, Haim Dubossarsky, and Nina Tahmasebi. 2021. SemEval-2020 Task 1: Unsupervised Lexical Semantic Change Detection. In *Proceedings of the 14th International Workshop on Semantic Evaluation*, pages 1–23, Barcelona (Online).

Nina Tahmasebi, Lars Borin, and Adam Jatowt. 2021. Survey of Computational Approaches to Lexical Semantic Change Detection. In Nina Tahmasebi, Lars Borin, Adam Jatowot, Yang Xu, and Simon Hengchen, editors, *Computational Approaches to Semantic Change*, chapter 1, pages 1–91. Language Science Press, Berlin.

Hiroya Takamura, Ryo Nagata, and Yoshifumi Kawasaki. 2017. Analyzing semantic change in Japanese loanwords. In *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 1, Long Papers*, pages 1195–1204, Valencia, Spain. Association for Computational Linguistics.

Elizabeth Closs. Traugott and Richard B. Dasher. 2005. *Regularity in Semantic Change*. Cambridge University Press, Cambridge.

Ana Uban, Alina Maria Ciobanu, and Liviu P. Dinu. 2019. Studying laws of semantic divergence across languages using cognate sets. In *Proceedings of the 1st International Workshop on Computational Approaches to Historical Language Change*, pages 161–166, Florence, Italy. Association for Computational Linguistics.

Ana-Sabina Uban, Alina Maria Ciobanu, and Liviu P Dinu. 2021a. Cross-lingual Laws of Semantic Change. In Nina Tahmasebi, Lars Borin, Adam Jatowot, Yang Xu, and Simon Hengchen, editors, *Computational Approaches to Semantic Change*, chapter 7, pages 219–260. Language Science Press, Berlin.

Ana Sabina Uban, Alina Maria Cristea, Anca Dinu, Liviu P. Dinu, Simona Georgescu, and Laurentiu Zoicas. 2021b. Tracking semantic change in cognate sets for English and Romance languages. In *Proceedings of the 2nd International Workshop on Computational Approaches to Historical Language Change 2021*, pages 64–74, Online. Association for Computational Linguistics.

Radim Řehůřek and Petr Sojka. 2010. Software Framework for Topic Modelling with Large Corpora. In *Proceedings of the LREC 2010 Workshop on New Challenges for NLP Frameworks*, pages 45–50, Valletta. ELRA.

A Scatter plot and correlation of the variables

Figure 5 displays pair-wise scatter plots between the standardized variables for French–Spanish cognate pairs. On the top right of each panel is shown Pearson's correlation coefficient.

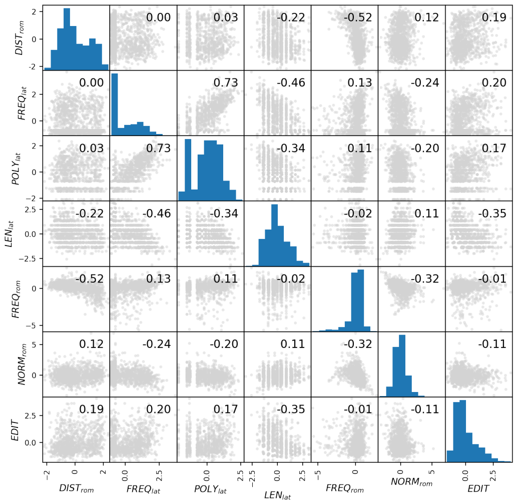

Figure 5: Scatter plot and correlation of the standardized variables for French–Spanish cognate pairs.