

Towards a Map of Related Words in Romance Languages

Liviu P. Dinu^{♣,♥} Ana Sabina Uban^{♣,♥} Bogdan Iordache^{♣,♥}

Claudia Vlad^{♣,♥} Simona Georgescu^{♣,♥} Laurentiu Zoicas^{♣,♥} Anca Dinu^{♣,♥}

University of Bucharest, [♣] Faculty of Mathematics and Computer Science,

[♣] Faculty of Foreign Languages and Literatures, [♥] HLT Research Center

{ldinu, auban}@fmi.unibuc.ro, ioan.iordache@s.unibuc.ro,

{claudia.vlad, simona.georgescu, laurentiu.zoicas, anca.dinu}@fls.unibuc.ro

Abstract

We propose a map of cognates and borrowings usage in Romance languages, having as a starting point the pairs of cognates and borrowings between any two of these idioms from RoBoCoP, the largest database built upon electronic dictionaries containing etymological information for Portuguese, Spanish, French, Italian and Romanian. Having in mind that words are used and evolve in language communities over time, on the basis of the pairs extracted from RoBoCoP, we determine how many of them occur and with what frequency in the context of the languages in use, based on three online parallel corpora that contain all five Romance languages: *Wikipedia*, *Europarl* – focusing on proceedings of the European Parliament and *RomCro2.0* – containing literary texts in different languages, translated in Romance languages and Croatian.

1 Introduction and Related Work

Although they share a common Latin origin, the Romance languages have evolved in distinct ways, due to factors such as geographical position, historical circumstances, peoples and languages with which they have come in contact over time. The emergence of the early Romance vernaculars, currently associated with the standard Romance languages (Portuguese, Spanish, French, Italian, and Romanian), can be traced back to at least the ninth century (Malkiel, 1977, 1978; Wright, 1983; Ledge-way, 2012). Since then, these languages have undergone continuous evolution, gradually developing into their present-day configuration.

Based on their origin, Romance languages share common features at various levels, displaying similar behaviors in terms of phonetics and phonology (e.g., the basic five vowels from Latin /a/, /e/, /i/, /o/, /u/), morphology (e.g., verb inflection that produces changes in terms of tense, mood, aspect,

person and number), syntax (e.g., the Subject-Verb-Object (SVO) word order in affirmative sentences), semantics and lexicology (e.g., words of Latin origin that refer to the human body, family members or numbers).

Due to their distinct historical circumstances, social and political contexts, which led to the contact of Latin with different languages from various families, they also present specific characteristics. At the phonological level, French and Portuguese, for example, frequently use nasal vowels, while the other languages do not. At the morphological level, in Romanian the definite article is postponed, while in the other Romance idioms it precedes the noun. In terms of syntactic differences, French stands out as not being a pro-drop language, like the others, the subject being mandatory, since the verb inflection is not explicit in the spoken language. As for the semantic differences, these can be visible through polysemy which can lead to shifts in meaning and false friends (although the words in the Romance languages may be cognates or borrowings) – e.g., the Romanian word *aprinde* means ‘to ignite’, while the words with the same etymon (La. *apprehendere*) in the other Romance languages mean ‘to learn’ (Fr. *apprendre*, It. *apprendere*, Es./Pt. *aprender*).

In addition, at the lexical level, each Romance language inherited a varying proportion of words from Latin that constitute their fundamental lexical core, i.e. Ro. 30%, Fr. 36%, Es. 40%, It. 44%, Pt. 45% (Sala, 1988). Along with these, each Western Romance idiom exhibits borrowings from Latin in its basic lexicon, i.e., Pt. 25%, Fr. 27%, Es. 27%, It. 28%. Romanian, a standard Eastern Romance language, stands out as an exception. It comprises only slightly more than 1%, to which another 8% can be added, as Latin words loaned from French and Italian, that are part of the “cultural superstrate” (Reinheimer Ripeanu, 2004). Based on the num-

bers exposed, the total proportion of Latin words in the Romance languages reaches the following percentages: Ro. 32% (or 39% if we consider the extra 8%), Fr. 63%, Es. 67%, Pt. 70%, It. 72% (Cristea et al., 2021).

In spite of the fairly high percentages of words of Latin origin, that doesn't necessarily mean there is perfect correspondence at the lexical semantic level between the five Romance languages based on their common etymology, nor that the respective words are used in the same contexts or registers and with the same frequency in all idioms. Words appear in societies, and these can change in time in different ways, affecting not only the members of the respective communities, but also the languages, and thus leading to linguistic changes. In fact, linguistic change is a process that occurs in the evolution of a language used to communicate within a certain society and it can affect any linguistic level (phonological, morphological, syntactic, semantic) (Castro, 1991).

When it comes to the meaning of the words, linguistic change can be attributed to various factors. On the one hand, words can change their meaning or add new meanings as a way to adapt to changing/different realities – e.g., La. *piscina* used to refer to a 'small pond with fish', but in time its meaning switched to 'pool', as it can be observed in today's Romance languages, Pt./Es. *piscina*, Fr. *piscine*, Ro. *piscină*. On the other hand, there are words that are created to designate new concepts – terms like En. *hardware* or La. *corpus* appear in all Romance languages, referring to concepts related to the specialized domains of computer science and linguistics – and others become obsolete either because there is more than one term that refers to the same concept – e.g., Pt. *botica*, from Gr. *apothíki* via La. *apothēca*, has fallen out of use, the current preferred form to designate 'a shop that sells (and can also prepare) medicines' being Pt. *farmácia* from Gr. *pharmakeía*, via late La. *pharmacīa* – or because certain realities cease to exist – e.g., Ro. *agă*, designating an 'officer in the Ottoman army'.

In the act of translation - whether done by humans or by machine - there is often the danger of using as the equivalent of a word in language A its cognate in language B, even though it does not have the same frequency and would not be used in the context (e.g. It. *sapienza* 'wisdom' - Es. *sapiencia* [with the same meaning, but very rarely used - here a Spaniard would use *sabiduría*]). Moreover, when

it comes to borrowings from language A into language B, the temptation is to use the word pair as perfect equivalents, regardless of the fact that they do not cover the intended communicative situation, one of them being a marginal or highly specialized word (e.g., Es. *valija* comes from It. *valigia* 'suitcase', but the word preferred by native Spanish speakers is *maleta*).

In the light of the above, we propose a map of related words in Romance languages, having as a starting point the pairs of cognates and borrowings between any two of these idioms from RoBoCoP (the Romance Borrowings Cognates Package), a cognate database built upon electronic dictionaries containing etymological information for Portuguese, Spanish, French, Italian and Romanian Dinu et al. (2023). While cognates are words that share the same Latin origin, borrowings are words that are borrowed from one Romance language to another.

Having in mind that words are used and evolve in language communities over time – some maintaining their original meaning, others adding new ones or falling into oblivion, others (re)appearing to describe new concepts –, on the basis of the pairs extracted from RoBoCoP, we intend to determine how many of them occur and with what frequency in the context of the languages in use, based on three online parallel corpora that contain all five Romance languages: 1) *Wikipedia*, 2) *Europarl* – focusing on proceedings of the European Parliament (Koehn, 2005) and 3) *RomCro2.0* – containing literary texts in different languages, translated in Romance languages and Croatian (Mikelenić et al., 2025).

2 Data

In Dinu et al. (2023) is introduced RoBoCoP, the most comprehensive database of Romance cognates and borrowings based on the etymological information provided by the dictionaries (the largest known database of this kind, in our best knowledge). The pairs of cognates and borrowings were extracted between any two Romance languages by parsing electronic dictionaries of Romanian, Italian, Spanish, Portuguese and French.

RoBoCoP contains cognates and etymons in five Romance languages: Italian, Spanish, Portuguese, Romanian, and French. The total number of cognate and borrowing pairs is given in Table 1. In Table 2, the number of borrowings is broken down

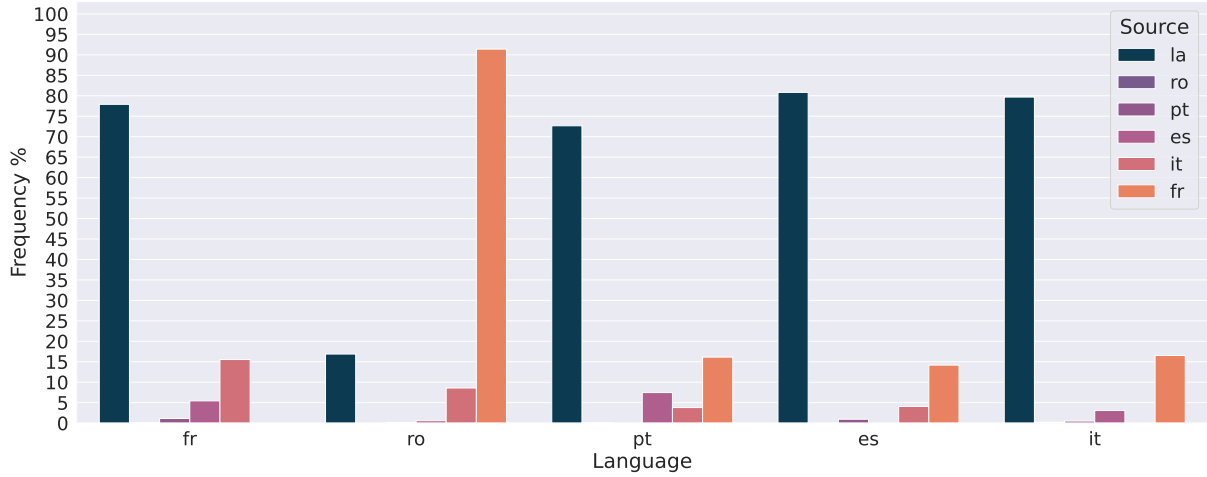


Figure 1: Etymology distribution for all words (cognates and borrowings) in the database

	Ro	It	Es	Pt	Fr
Ro		4,999	7,588	5,855	7,360
It	3,139		7,863	12,198	7,105
Es	209	770		9,533	10,220
Pt	103	620	1,201		7,783
Fr	33,311	2,896	1,690	2,450	

Table 1: Number of cognate pairs (above the main diagonal) and borrowing pairs (below the main diagonal) for each Romance language pairs in RoBoCoP dataset.

by the direction of the borrowing. The database comprises a total of 125,598 words across all languages and 90,853 cognate pairs. It has already been used with good results in various important historical linguistic tasks such as cognate identification (Dinu et al., 2023), cognate-borrowings discrimination (Dinu et al., 2024c), determining the borrowing direction (Dinu et al., 2024a), or protoword reconstruction (Dinu et al., 2024b).

3 Quantitative Results

First, we extracted from RoBoCoP the Latin and Romance etymology distribution for all related words and each analyzed language, illustrated in Figure 1. While in all the Western Romance languages the Latin etymons show the highest percentages (more than 70%), in Romanian most words (90%) have French origin. Also, some of Romanian words display multiple etymology – e.g., the word *natural* is presented in the Romanian dictionaries as coming from La., It., Fr., De.. The other Romance languages also contain words coming from French, but display much lower percentages than Romanian, around 15% of the total number of

terms.

We have also investigated the distribution of the main parts of speech (adjective, adverb, noun, and verb) for all related words in the RoBoCoP database, for all analyzed Romance languages. The parts of speech considered were only those directly inherited or borrowed, without considering the further derivational processes undergone by each language. The results are shown in Figure 2. In all languages, the most used parts of speech are nouns, with percentages varying between 55% and 65% of the total number of words. Adjectives and verbs show lower percentages, varying from 10% to 25%, while adverbs show percentages of less than 5%, with the exception of Portuguese, which shows a percentage of almost 10%.

Second, we investigated the coverage of related words present in the RoBoCoP database against three Romance languages corpora: *Wikipedia*, *Europarl* and *RomCro2.0*.

We compared the list of words from the RoBoCoP database with the lists obtained from the three corpora. Table 3 summarizes the coverage of RoBoCoP cognates and borrowings w.r.t. the three corpora, per each language. The total amount of words from the RoBoCoP database that were also found in the three corpora is of 11,124 Spanish words, 14,850 Portuguese words, 15,179 Italian words, 36,291 Romanian words, and 37,115 French words. The percentages of the RoBoCoP words used in all the corpora vary as following: 77% for Romanian, 79.10% for Italian, 81.74% for French, 87.76% for Portuguese and 92.93% for Spanish. Spanish and Portuguese show the highest percentages of words used in the three corpora. This seems to be a conse-

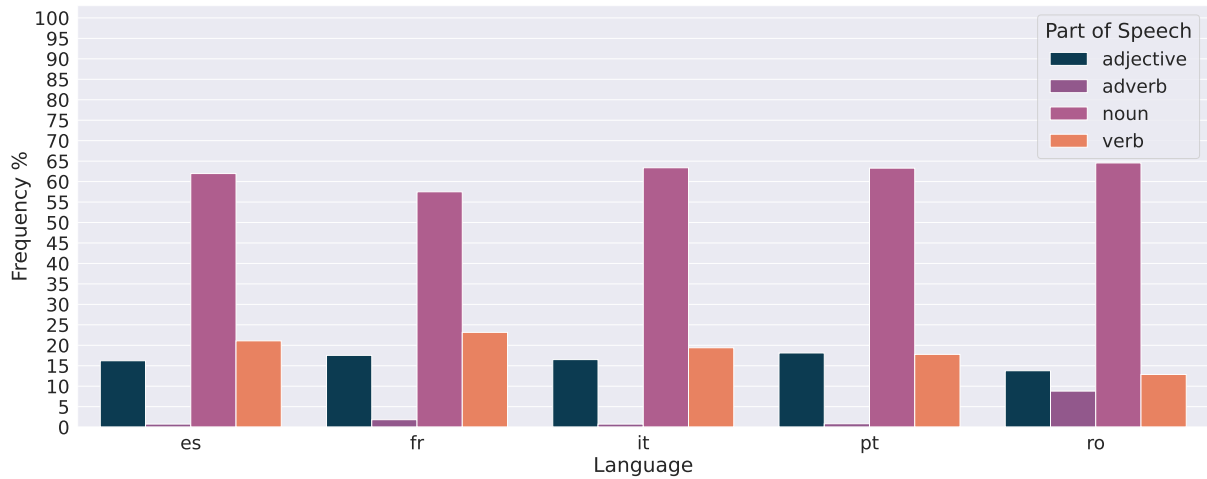


Figure 2: Part of speech distribution for all words (cognates and borrowings) in the database

quence of the less extensive sets of terms available in RoBoCoP for these languages. Among the three corpora, Wikipedia presents the highest frequency of words used (between 72.2% for Romanian and 83.02% for Spanish), while *Europarl* and *RomCro2.0* show visibly lower numbers. The *Europarl* and *RomCro* corpora displayed lower percentual coverage of the cognates and borrowings present in RoBoCoP database, most probably because of their more specialized content, while Wikipedia presents a more heterogeneous content, which enables the usage of a more diverse vocabulary.

The percentage and the total number of cognates and borrowings between any two Romance languages from RoBoCoP dataset also found in the three investigated corpora are shown in Table 4.

	Ro	It	Es	Pt	Fr
Ro		3,135	209	102	33,311
It	4		376	62	1,981
Es	0	394		104	1,366
Pt	1	558	1,097		2,369
Fr	1	915	324	81	

Table 2: Number of borrowing pairs in the RoBoCoP database per pair of languages, in both directions. The number in a cell represents the number of borrowings from the language on the same column to the language on the same line.

Finally, for a more *language in use* insight, we computed the frequencies with which the related words in the RoBoCoP database occur in the three corpora. The motivation was that dictionaries aim to contain the entire vocabulary of a language, but do not reflect the language use distribution. In

Language	Count	Frequency %	Corpus
fr	27,539	74.20	wikipedia
fr	13,633	36.73	europarl
fr	12,976	34.96	romcro
fr	30,337	81.74	all
fr	37,115	100.00	total
ro	26,210	72.22	wikipedia
ro	9,124	25.14	europarl
ro	11,722	32.30	romcro
ro	27,945	77.00	all
ro	36,291	100.00	total
pt	11,907	80.18	wikipedia
pt	7,783	52.41	europarl
pt	8,016	53.98	romcro
pt	13,033	87.76	all
pt	14,850	100.00	total
es	9,235	83.02	wikipedia
es	7,250	65.17	europarl
es	7,520	67.60	romcro
es	10,337	92.93	all
es	11,124	100.00	total
it	11,999	79.05	wikipedia
it	8,331	54.89	europarl
it	8,584	56.55	romcro
it	12,007	79.10	all
it	15,179	100.00	total

Table 3: RoBoCoP word distribution grouped by language and corpus

Table 5 we give the total frequency of words from RoBoCoP used in the three corpora.

4 Results Analysis

The results obtained are significant for understanding the distribution of the lexicon of each language and in relation to one another, as well as the actual usage of words. A hasty glance at dictionaries will almost always be misleading, since, regardless of the usage indications they may contain (e.g. literary term, colloquial term, etc.), there are no data

Language Pair	Wikipedia		Europarl		RomCro	
	Cognates	Borrowings	Cognates	Borrowings	Cognates	Borrowings
es-it	2,414 (65.85%)	571 (75.03%)	1,850 (50.46%)	287 (37.71%)	1,920 (52.37%)	324 (42.58%)
fr-it	1,563 (71.96%)	1,882 (65.05%)	1,203 (55.39%)	788 (27.24%)	1,254 (57.73%)	874 (30.21%)
it-pt	6,791 (65.17%)	425 (68.77%)	4,251 (40.79%)	185 (29.94%)	4,378 (42.01%)	229 (37.06%)
it-ro	1,800 (73.62%)	1,945 (61.75%)	1,157 (47.32%)	854 (27.11%)	1,392 (56.93%)	1,084 (34.41%)
es-fr	3,047 (74.48%)	1,199 (71.16%)	2,400 (58.67%)	631 (37.45%)	2,428 (59.35%)	707 (41.96%)
es-pt	2,728 (67.89%)	906 (76.33%)	2,078 (51.72%)	313 (26.37%)	2,234 (55.60%)	400 (33.70%)
es-ro	4,356 (74.54%)	56 (26.79%)	2,797 (47.86%)	14 (6.70%)	3,278 (56.09%)	30 (14.35%)
fr-pt	1,649 (73.88%)	1,670 (68.58%)	1,268 (56.81%)	810 (33.26%)	1,274 (57.08%)	830 (34.09%)
fr-ro	2,804 (82.08%)	18,409 (55.27%)	1,703 (49.85%)	5,282 (15.86%)	2,066 (60.48%)	6,622 (19.88%)
pt-ro	1,887 (74.15%)	28 (84.85%)	1,225 (48.13%)	7 (21.21%)	1,407 (55.28%)	14 (42.42%)

Table 4: Counts and frequencies (%) of cognates and borrowings by language pair across corpora.

there for an accurate picture of the usage status of each word in the native speaker’s lexicon.

The quantitative results obtained by counting how many of the related words in RoBoCoP database are also present in the three considered corpora show some surprising results. If we take a look at the numbers of cognates/borrowings for each language pair in Table 4, we can observe that the lexical correspondence between certain language pairs is in fact higher than expected (e.g. It-Pt), while between others it is insignificant (Ro-Pt). In accordance with our general knowledge of Romance languages and the relationships between them, we would have expected, for example, that Spanish and Portuguese would have the largest shared lexicon, since geographical proximity seems decisive. From a chronological perspective, looking at the stages of diversification of Latin and the birth of Romance languages, we would have expected Italian to be closer to French, since they had a longer period of coexistence (they are thought to have evolved together until ca. 7th-8th century AD, while Romanian, for instance, separated from the compact Roman territory before the 4th century AD). In the same vein, we have been taught since (Bartoli, 1925) that the lateral areas (i.e. Romanian at the extreme East and, at the other extreme, Portuguese, alongside Spanish) retain many common lexical elements, lost in the rest of the Romance territory (e.g. Ro. *masă* - Es./Pt. *mesa* “table” < La. *mensa* “id.”, while It. and Fr. use *tavola* and *table* < Lat. *tabula*, which was an innovation over the archaic *mensa*). However, the present analysis shows a much greater affinity between the center - Italian - and the extreme West - Portuguese.

Leaving aside the scientific perspective and focusing on the pure intuition of the speaker, here again we are struck by divergences between our presuppositions and what the data actually show.

	cognates			borrowings		
	Wiki	EuroParl	RomCro	Wiki	EuroParl	RomCro
RO	52.83%	63.58%	51.49%	39.01%	46.21%	26.52%
IT	32.33%	41.09%	38.10%	3.67%	3.39%	3.31%
ES	44.59%	54.70%	52.43%	2.98%	3.60%	2.33%
PT	40.70%	49.33%	50.67%	4.29%	4.08%	4.31%
FR	37.13%	45.02%	45.16%	0.72%	0.59%	0.68%

Table 5: Frequency of related words computed across each of the three corpora. We display how many of the tokens in a given corpus are matched to a related word (cognate with another language, or borrowed from another language).

Any native speaker of Romanian can easily notice the similarity between Romanian and Italian, both at phonetic and morpho-syntactic level, and - even more pronounced - at lexical level: it is therefore very easy for a Romanian to learn Italian precisely because of the large number of cognates that have the same meaning and usage. Therefore, we would have expected the cognate pairs to be much more numerous, and yet they do not even reach 2,000 words, which is a sub-average number compared to Italian-Portuguese, whose cognate pairs reach almost 7,000 (see Table 4).

Table 4 displays how many of the cognate/borrowing pairs are covered by at least one occurrence of each of their constituents in one of the corpora. In other words, for a given language pair A-B we consider a cognate/borrowing word pair to be covered if both the word in A and the word in B occur at least once in the given corpus.

Apart from this static analysis, we are also concerned with the dynamic token frequencies of the cognate/borrowed words in the studied corpora. Thus, we gathered for each language the set of words that are either cognates with other languages or borrowed from other languages, and computed their frequency across the text corpus (i.e. how many of the tokens in a given corpus correspond to a cognate/borrowing). We present in Figure 3 and Table 5 these frequencies for each corpus and

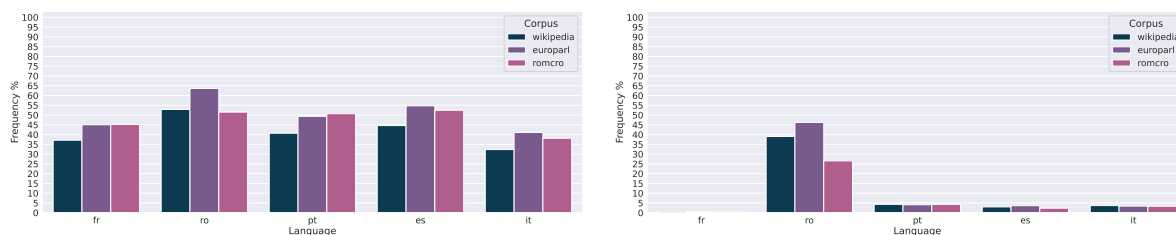


Figure 3: Frequency of cognate words (left) and borrowed words (right) computed across each of the three corpora.

each language, summed over all possible paired languages (i.e. the languages where the words are cognates in or borrowed from, with respect to the target language that is plotted). We notice that, due to the asymmetry of words borrowed from one language to another (e.g. Romanian borrowed from all other languages more than any of them borrowed from Romanian, see Table 2) there is a consistent difference regarding the presence of cognate and borrowed words in each of the corpora.

This situation must be determined by the share of Latin-derived words in semantic fields with high frequency in speech. For example, among languages where the basic semantic fields are mostly made up of words of Latin origin, the frequency of cognate pairs in corpora is higher; on the other hand, languages that have undergone major influences from different, non-Romantic languages, do not have such a high lexical correspondence, therefore the frequency of pairs will not reach the same level as the others either. For example, almost 10% of the Romanian lexicon consists of borrowings from Slavic languages, and they cover essential concepts – for instance, ‘to love’ (*a iubi*), ‘friend’ (*prieteni*), ‘dear’ (*drag*) -, while in the other languages these concepts are verbalized by words of Latin origin. In Spanish and Portuguese, numerous words of Arabic origin cover the semantic field of food, constructions, etc. They will not correspond to Romanian, where these concepts are expressed mostly by words either of Latin or of French origin. Also, words of Turkish, Greek and Slavic origin may occur in Romanian in such semantic fields.

As far as Inter-Romance borrowings are concerned, the analysis is also revealing. If borrowings from Latin in most Romance languages cover between 25 and 30% of the vocabulary, verbalizing essential concepts in most lexical fields, borrowings from one Romance language to another are extremely few, with a frequency of use that can be considered insignificant. The only valuable lexical contribution, and indeed essential for the develop-

ment of a literary language, was that provided by French to Romanian: the pairs consisting of Romanian borrowings and their French source words number more than 20,000, of which more than 18,000 are present in the wikipedia corpus, and less than 5,000 in the other corpora. The number is not at all negligible if we bear in mind that the next most represented borrowings in order of quantity are the Italian loans in Romanian, which nonetheless do not reach 2500, followed by French loans in Portuguese, around 1800. In both cases, the representation in the Europarl and Romcro corpora does not exceed 1500. As for word exchanges between, for example, Italian and Portuguese, or Portuguese and Romanian, they are close to 0.

5 Conclusion and Future Work

In this paper, we were interested about the usage of related words in the main Romance languages (Portuguese, Spanish, French, Italian and Romanian). Specifically, we have investigated the behavior of cognates and borrowing for any pair of Romance languages on three different corpora, and we compared with their behavior at the dictionary level. The analysis was performed starting with the RoBoCoP database, the largest database of Romance cognates and borrowings based on the etymological information provided by the dictionaries. In order to establish a map of related words in the Romance languages, we compared the list of words from the RoBoCoP database with the lists obtained from the three corpora – *Wikipedia*, *Europarl* and *RomCro2.0* – having in mind, on the one hand, that dictionaries aim to contain the entire vocabulary of a language, but may not be updated in real time and, on the other hand, languages may be affected by changes that are visible in language use, which, in its turn, depends on the context. The obtained results confirm the general linguistics intuitions and theories, but also bring new perspectives on the usage of the Romance lexicon.

In future work we intend to extend the analysis

by including also the Catalan language. We also intend to make a diachronic analysis and to include polysemy as a fundament of our research.

Acknowledgments

This work was supported by Ministry of Research, Innovation and Digitization, CNCS-UEFISCDI, project SIROLA, number PN-IV-P1- PCE-2023-1701, within PNCDI IV.

References

- Matteo Bartoli. 1925. *Introduzione alla neolinguistica:(principi, scopi, metodi)*. LS Olschki.
- Ivo Castro. 1991. Curso de história da língua portuguesa.
- Alina Maria Cristea, Anca Dinu, Liviu P. Dinu, Simona Georgescu, Ana Sabina Uban, and Laurentiu Zoicas. 2021. [Towards an etymological map of Romanian](#). In *Proceedings of the International Conference on Recent Advances in Natural Language Processing (RANLP 2021)*, pages 315–323, Held Online. IN-COMA Ltd.
- Liviu Dinu, Ana Uban, Anca Dinu, Ioan-Bogdan Iordache, Simona Georgescu, and Laurentiu Zoicas. 2024a. [It takes two to borrow: a donor and a recipient. who's who?](#) In *Findings of the Association for Computational Linguistics: ACL 2024*, pages 6023–6035, Bangkok, Thailand. Association for Computational Linguistics.
- Liviu P. Dinu, Ana Sabina Uban, Alina Maria Cristea, Anca Dinu, Ioan-Bogdan Iordache, Simona Georgescu, and Laurentiu Zoicas. 2023. Robocop: A comprehensive romance borrowing cognate package and benchmark for multilingual cognate identification. In *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing, EMNLP 2023, Singapore, December 6-10, 2023*, pages 7610–7629. Association for Computational Linguistics.
- Liviu P Dinu, Ana Sabina Uban, Alina Maria Cristea, Ioan-Bogdan Iordache, Teodor-George Marchitan, Simona Georgescu, and Laurentiu Zoicas. 2024b. [Verba volant, scripta volant? don't worry! there are computational solutions for protoword reconstruction](#). In *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 6314–6326, Miami, Florida, USA. Association for Computational Linguistics.
- Liviu P. Dinu, Ana Sabina Uban, Ioan-Bogdan Iordache, Alina Maria Cristea, Simona Georgescu, and Laurentiu Zoicas. 2024c. [Pater incertus? there is a solution: Automatic discrimination between cognates and borrowings for Romance languages](#). In *Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024)*, pages 12657–12667, Torino, Italia. ELRA and ICCL.
- Philipp Koehn. 2005. Europarl: A parallel corpus for statistical machine translation. In *Proceedings of machine translation summit x: papers*, pages 79–86.
- Adam Ledgeway. 2012. *From Latin to Romance: Morphosyntactic typology and change*, volume 1. Oxford Studies in Diachronic a.
- Yakov Malkiel. 1977. Factors in the unity of romania. *Romance Notes*, 18(2):263–271.
- Yakov Malkiel. 1978. The classification of romance languages. *Romance Philology*, 31(3):467–500.
- Bojana Mikelenić, Gorana Bikić-Carić, Metka Bezljaj, Antoni Oliver, and Marko Tadić. 2025. [RomCro v.2.0 - parallel corpus of romance languages and croatian](#). HR-CLARIN.
- Sanda Reinheimer Ripeanu. 2004. *Les emprunts latins dans les langues romanes*. Editura Universității din București.
- Marius Sala. 1988. *Vocabularul reprezentativ al limbilor romanice*. Ed. Științifică și Enciclopedică.
- Roger Wright. 1983. Unity and diversity among the romance languages. *Transactions of the Philological Society*, 81(1):1–22.