Bartoli's areal norms revisited: an agent-based modeling approach

Dalmo Buzato
Faculty of Letters
Universidade Federal de Minas Gerais
Belo Horizonte, Brazil
dalmobuzato@ufmg.br

Evandro L. T. P. Cunha
Faculty of Letters
Universidade Federal de Minas Gerais
Belo Horizonte, Brazil
cunhae@ufmg.br

Abstract

In this article, computational agent-based modeling and simulation is used to evaluate the linguist Matteo Bartoli's areal norms regarding the relationship between language change and spatial features. To achieve this, a simple modeling algorithm was developed to allow the transmission of innovative linguistic elements within a network based on European geographical data. The results obtained show differences between propagation patterns of items originating from and reaching more connected, more isolated and more peripheral regions. These outcomes support Bartoli's theory, including regarding West Iberian languages (such as Portuguese and Galician) and their sometimes archaizing tendencies.

1 Introduction

"*Le romanisme est le domaine qui se prête le mieux à illustrer les développements linguistiques, et celui où les méthodes qui conviennent à l'histoire des langues se laissent le mieux discuter*"[1] (Meillet, 1923, p. 80)

The understanding that languages change has been reported in linguistic tradition since the primary studies and first descriptions of languages. Nowadays, the perception that languages change in relation to time, space, and social classes is almost unanimous. This perception goes beyond the awareness of linguists and is shared by most speakers. However, the scope of language change in the various linguistic theories diverges significantly.

In the 19th century, the *neogrammarian school* was established by German linguists with the objective of proposing laws for explaining language change (especially sound change). In their vision, sound change would have regularity, following infallible laws. The importance of the neogrammarian school is indisputable for the consolidation of structural linguistics.[2] However, even at the peak of neogrammarian school development, not all linguists agreed with the hypothesis of universal laws for sound change, which disregard the impact of any other variables on language change. An example of a group of thinkers who opposed neogrammarian thought is the *neolinguistic school*, founded by the Italian linguist Matteo Bartoli (1873-1946).

The neolinguistic school, later also known as *spatial linguistics* (*linguistica spaziale*), claims that, contrary to the neogrammarian thought, language change does not follow universal, infallible, abstract rules inherent to the linguistic system. In Bartoli's and his colleagues' view, this phenomenon is closely tied to sociogeographical features and to concepts such as centrality, peripheralness and isolation.

1.1 Bartoli's areal norms

In his two main works (Bartoli, 1925, 1945), Bartoli proposes five norms regarding the relationship between language change and the geography of a specific area. For all five norms, Bartoli provides a series of examples and cases to illustrate the dissected norm, and at each stage he presents counterexamples to the proposed norms. This is a key point of Bartoli's theory that deserves emphasis: unlike the neogrammarian view, which advocates universal laws, Bartoli introduced norms that would apply most of the time and could be adopted for diachronic analyses – however, exceptions could exist due to the specificities of the social and geographical context.

[1] "The field of Romance studies is the one that best lends itself to illustrating linguistic developments, and the one where the methods appropriate to the history of languages are best discussed".

[2] For a detailed explanation of the importance of neogrammarian thinking for the constitution of the linguistic structuralism found in the works of Saussure and Bloomfield, for example, as well as the implications of these views for the study of language change, please refer to the first chapter of Weinreich et al. (1968).

The five norms proposed by Bartoli were briefly summarized in the following topics by Manczak (1988):

I The more isolated area usually preserves the earlier stage.

II If one of two linguistic stages is found in peripheral areas and the other in a central area, the stage occurring in the peripheral areas is usually the earlier one.

III The larger area usually preserves the earlier stage.

IV The earlier stage is usually preserved in the later area.

V If one of two linguistic stages disappears or becomes moribund and the other survives, the stage that disappears is usually the earlier one.

In Bartoli's proposal, there is extensive discussion about the differentiation of Iberian languages, including Portuguese, as opposed to other Romance languages. His five norms seem to explain, through spatial linguistics, the reasons why, in some cases, linguistic innovations never reached the Iberian Peninsula, especially the Lusophone area. As a consequence, the linguistic resources used by the speakers of these languages and dialects can be more archaic: a lexical example is the Portuguese verb *comer*, which comes from the Latin *comedere*, a more archaic linguistic form. In Latin itself, at a later time, another verbal form emerged for the same meaning: *manducare*. This innovative form led to verbs like *mangiare* in Italian and *manger* in French. Areas geographically closer to the center of this innovation[3] adopted the new verbal form, as seen in Italian, French and Catalan (*menjar*) – but also in Romanian (*mânca*), in the the easternmost limit of the Romance world. Meanwhile, the westernmost areas kept the archaic form, as in Portuguese, Galician (*comer*), and Spanish (*comer*), as represented through the map depicted in Figure 1.

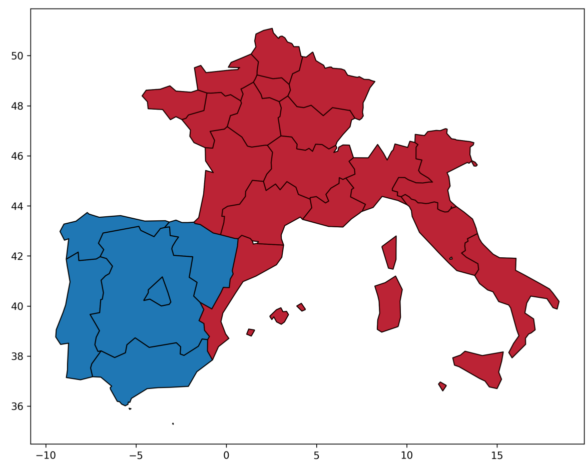

Figure 1: Part of present-day Romance language-speaking Europe that use forms for the verb 'to eat' derived from Latin *manducare* (red) and *comedere* (blue)

1.2 Agent-based modeling in linguistics

According to Šešelja (2023), agent-based models are "computational models that simulate the behavior of individual agents in order to study emergent phenomena at the level of the community". In agent-based modeling (ABM), each agent's decision-making is carried out autonomously and is implemented through simple computationally defined rules. This approach allows for decentralized local interactions among agents, as well as interactions between agents and the environment they are situated in. The advantage of this methodology is the ability to gradually and controllably observe the emergence of complex phenomena at the population level.

The use of this methodological approach in linguistics has become increasingly common, especially in areas related to language dynamics,[4] such as language change, language acquisition, areal linguistics, evolutionary linguistics, and language contact. In these fields, obtaining real linguistic data can be challenging, and in some cases, it may not provide all the necessary evidence to support hypotheses.

For instance, if a new word emerges in the city of Santiago de Compostela (approximately 100,000 inhabitants in 2021) and we want to study the spread of this lexical item among speakers over a short period of time, if we only have the option to use real/concrete data, we would probably need to conduct periodical interviews with a portion of the city's inhabitants and search for the new lexicon in the recordings. Apart from the immense methodological challenges of data collection, transcription, and storage, there are other issues: for

[3] In the case of Romance languages, in their genesis and initial development, central Italy can be considered an important center of innovations due to Rome being the capital of the Roman Empire, and a major population and prestige center.

[4] For a definition of *language dynamics*, see Wichmann (2008). For a general understanding of the view of language as a complex adaptive system, please refer to Steels (2000), Beckner et al. (2009) and de Oliveira (2018).

example, the analyzed word could be in the idiolect (the individual linguistic repertoire) of a speaker, but they might not have used it in that particular recording. Furthermore, it would be difficult, if not impossible, to make precise notes on whom a particular speaker received such an innovation from. All these problems are alleviated in ABM, and through it we can make such observations – albeit probabilistic, but allowing us, in many cases, to gain insight into real-world phenomena.

The use of agent-based modeling in linguistics began approximately in the 1990s, initially in the study of phonetic or morphosyntactic changes. However, in recent years, there has been a significant increase in the use of this methodology to explore broader aspects of language, such as the emergence of communication, languages, and grammatical systems related to evolutionary linguistics. This includes investigating aspects of compositionality and holophrase in early communication systems, as well as issues related to language competition and language contact. Additionally, it encompasses more traditional aspects of language change, like the spread of innovations and the influence of speakers' and listeners' prestige. Non-exhaustive examples of research that employ agent-based modeling in linguistics include: Harrison et al. (2002); Castelló et al. (2008); Troutman et al. (2008); Ke et al. (2008); Fagyal et al. (2010); Castelló et al. (2013); Chirkova and Gong (2014); Civico (2019); Dekker and De Boer (2020); Louf et al. (2021); Charalambous et al. (2023); Rosillo-Rodes et al. (2023).

1.3 Modeling Bartoli's norms

For Albrecht (1996), "the tenets of neolinguistics became well established in the historical and geographical approaches". Over the last century, since the publication of Bartoli's first work with his areal theory, several case studies questioning and validating his hypotheses have been presented. Nonetheless, despite the various examples and observations of Bartoli (all grounded in the Romance world, especially in the differentiation among Romance languages and dialects) and the numerous successors who tested his hypotheses in specific cases, Bartoli's norms still lack widespread empirical validation.

The aim of this article is to revisit the first three areal norms developed by Bartoli through computational modeling, given the potential for the computational implementation of geographical and linguistic concepts. To achieve this, we have developed an agent-based modeling algorithm, and to simulate the sociogeographical environment we have used a network based on the spatial description of the European continent. It is worth adding that modeling language change in Romance linguistics through computational simulation could be focused on observing and analyzing the peculiarities found in European varieties of majority Romance languages, as well as in minority Romance varieties (such as Mirandese and Astur-Leonese, for example) and Atlantic and Pacific varieties (such as Brazilian, African and Asian Portuguese). This article emphasizes European varieties to establish compatibility criteria with Bartoli's results in his works. However, ideally, Bartoli's norms could be observed in various geolinguistic contexts.

The results obtained through the simulations show differences between propagation patterns of items originating from and reaching more connected, more isolated and more peripheral regions. These outcomes support Bartoli's theory, including regarding West Iberian languages (such as Portuguese and Galician) and their sometimes archaizing tendencies, which are discussed in Section 3.

The remainder of this article is organized as follows: in Section 2, the methodology for developing the model and the fundamentals of the simulations carried out are explained; in Section 3, the results of the simulations are presented, and some issues relating to specific cases in the Iberian Peninsula are discussed; finally, Section 4 presents the conclusions of the article, raising possibilities for future work.

2 Methodology

To build a model of the transmission of linguistic items, we relied on a map of the current regions of the European Union.[5] With this goal, we used a `geojson` file (an open standard format designed to represent simple geographic features) of the NUTS 1[6] classification (First-level Classification of Territorial Units for Statistics), which is a geocode standard created by the European Union for referencing the administrative divisions of coun-

[5] Ideally, it would be better to have a computationally tractable file with geographic data of Medieval Europe, but we did not find such an option.

[6] https://ec.europa.eu/eurostat/web/gisco/geodata/reference-data/administrative-units-statistical-units/nuts

tries for statistical purposes.

All the procedures described in this study were performed using the Python programming language. Specifically, we utilized the `libpysal` (Rey and Anselin, 2007), `NetworkX` (Hagberg et al., 2008), and `GeoPandas` (Jordahl et al., 2020) packages for spatial network modeling, and `pandas` and `NumPy` for quantitative analyses resulting from the model. To transform the data from the `geojson` file into a network, we employed the open code available in the `NetworkX` documentation.[7] Essentially, it involves extracting centroids (the average of the coordinates that define the polygon's boundary) to connect the regions and subsequently constructing the graph based on the Queen model (where the graph considers two polygons as connected if they share a single point on their boundary).

Few modifications were made to the graph originally generated. In order to establish compatibility with Bartoli's theory, we excluded the Canary Islands and the Portuguese archipelagos, as well as Iceland. Additionally, the Italian and Greek islands, along with Scandinavia and Britannia, were originally not connected to the rest of the European continent, which would not allow the transmission and propagation of linguistic forms to and from these locations. Therefore, we introduced a link between a node in these disconnected regions and the nearest point on the European continent, ensuring that the entire graph was interconnected and enabling the potential transmission of items. The resulting graph used in this study can be seen in Figure 2.

Figure 2: Network used in the simulations, after the described modifications. Nodes represent different locations (regions and cities), while edges represent a connection between them

[7] `https://networkx.org/documentation/stable/auto_examples/geospatial/plot_polygons.html`

The algorithm developed to analyze the transmission of items can be described as follows. Initially, we take all pairs of nodes that are connected by an edge, and we will refer to these nodes as A and B. Since verbal communication always involves a two-way exchange, for each pair we consider two interaction possibilities: $A \to B$ and $B \to A$. To provide transmission, we randomly generate a probabilistic item ranging from 0 to 1. If the generated number is greater than 0.9,[8] and if the originating node has the innovative item while the destination node does not, then transmission occurs. If the originating node does not have the innovative item or the destination node already has it, nothing happens.

Initially, all nodes are set with the innovation variable equal to 0, meaning that none of them possesses the innovation.[9] At the start of each simulation run, one node is set with the innovation variable equal to 1, and from that point we can observe how this innovation propagates through the network. The origin of the innovation is one of the variables that will be analyzed in the following section, so being able to trace where the innovation originates is a valuable addition to this study.

The graph used in the modeling has 120 nodes and 254 edges. Each simulation consists of 100 rounds, and we ran the model 1,000 times for robust quantitative analyses. To analyze the centrality of each node in the network, we opted to use the *betweenness centrality* method. Currently, there are several different mathematical options for calculating node centrality. We chose betweenness centrality because it is a way of detecting the amount of influence that a node has over the flow of information in a graph. For more information on different centrality calculations in complex networks and a more detailed description of betweenness centrality, please refer to Golbeck (2015).

3 Results and discussion

Before presenting the quantitative data from the 1,000 simulation runs, we will first present the data from just one run. Although these results lack empirical strength, we believe that they serve as a good introduction to understanding the results that we will present next.

[8] I.e., with a probability of 0.1.

[9] It is important to note that, in this study, this linguistic form can be interpreted in various ways, such as a new phoneme, idiomatic or syntactic construction, or even a new lexical item.

The first question we aim to address in this article is whether innovations originating from different positions in the network (consequently, with different centralities) exhibit different patterns of propagation at equal times. By investigating the most and least central nodes in the network, we decided to run the model with the innovation originating from a node near the Greek islands, one of the most isolated locations on the graph (with only one connection). Figure 3 shows that, in 100 rounds, the number of nodes that received the innovation was low compared to the total, with only 37 receiving nodes. It is worth noting the significant period during which no node received the innovation, indicating that it stabilized in a few isolated locations, particularly between rounds 0 and 60. Therefore, we can infer that for the majority of the simulation, the innovation remained restricted to a few isolated locations and reacted with stability under these conditions.

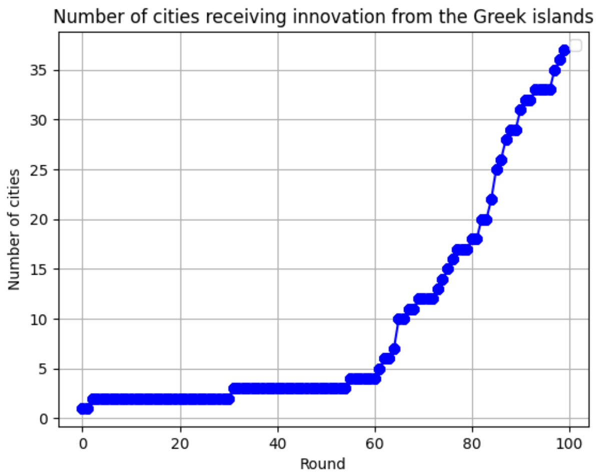

Figure 3: Number of nodes (representing regions and cities) that received an innovation from the Greek islands (one of the most isolated locations on the graph) after 100 rounds, in one simulation run

Bartoli's theory is based on the comparison between two or more locations to establish concepts such as center, isolation and periphery, for example. One location will almost always be more central in relation to another, but not necessarily the most central in the network, as applied in our context. Therefore, when we define a central, isolated, or peripheral node in our analyses, we are defining it within that specific context, meaning in comparison to the other node we want to observe.

When we observe innovation originating from Rome, or central Italy, a node more central compared to the Greek islands, we can notice a different propagation rate. The trajectory depicted in Figure 4 seems highly prototypical of the S-curve definition present in sociolinguistic tradition (Weinreich et al., 1968; Blythe and Croft, 2012). According to this postulate, a new linguistic form would initially be found in the idiolect of only a few speakers and could progressively grow within the population of a particular speech community. If it succeeded in the process of diffusion and competition with other forms, it could ultimately reach the final stability of being present in the idiolect of all or nearly all speakers in that speech community.

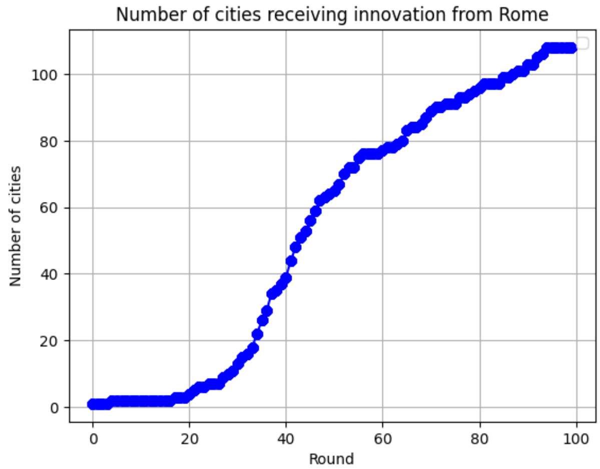

Figure 4: Number of nodes (representing regions and cities) that received an innovation from Rome (one of the most central locations on the graph) after 100 rounds, in one simulation run

In addition to the shape of the curve, another important aspect deserves emphasis in this introductory analysis: the number of nodes affected. Previously, we saw that only 37 nodes received the innovation from the isolated location at the end of the 100 rounds; now, on the other side, almost all nodes received the innovation from the central location.

Next, we will present the results of the analysis with the model being run 1,000 times, which means that we will be presenting the results of 100,000 rounds in total. As we can observe in Table 1, innovations originating from a more central node (in this analysis, Rome, with centrality = 0.06) have significantly greater and more consistent reach across the entire network when compared to innovations originating from a more isolated node (in this analysis, the Greek islands, with centrality < 0.01). In addition to the data recorded in the ta-

Region of origin	μ	sd
Greek islands	41.27	30.971
Rome	107.48	9.252

Table 1: Mean number of receiving nodes (and standard deviation) according to the innovation's node (representing regions and cities) of origin

ble, another interesting result found is that when innovation originates from the isolated location, in none of the 1,000 runs did the innovation reach all nodes of the graph. Instead, the maximum value reached was 115 receiving nodes, occurring only once. Conversely, there were multiple occurrences of all nodes becoming recipients of the new form when the innovation originated from the more central location.

Another question that we can answer through our data is in which round, on average, the innovation reaches central and isolated locations when they originate from other central and isolated locations.

To address this question, we selected five nodes with low centrality (i.e., isolated) and five with high centrality (i.e., central) and observed the results of when innovations reached them in all 1,000 runs of the model, depending on whether the innovation originated from a more central or a more isolated node.

As shown in Figure 5, when an innovation originates from a central location, other central locations (average round in which central nodes receive innovations from other central nodes: $\mu = 28.76$, $sd = 16.706$), disregarding graph distances, typically receive the innovation earlier compared to isolated regions ($\mu = 63.64$, $sd = 24.120$).

These data, along with the previous ones, seem to provide evidence for the correctness of some of Bartoli's norms. When innovations arise from central regions, if they manage to reach isolated regions, they take much longer to establish themselves in those areas, meaning that these regions become linguistically isolated. If we observe this state over time, we will find that more archaic forms are concentrated in isolated regions, while innovations are used in more central areas, as stated in Bartoli's first norm ("the more isolated area usually preserves the earlier stage").

Conversely, when we observe the spread of items originating from more isolated nodes, especially in terms of which round the spread reaches more central and more isolated nodes, we find that in-

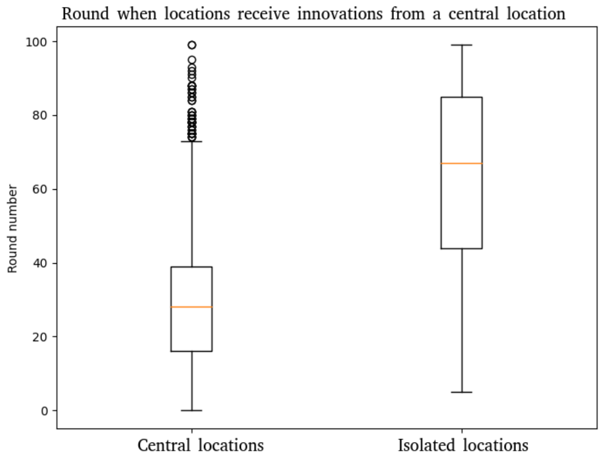

Figure 5: Round in which nodes (representing regions and cities) receive innovations from a more central node

novations from isolated nodes take more rounds to reach central nodes ($\mu = 76.39$, $sd = 15.918$) and also to reach other isolated nodes ($\mu = 80.19$, $sd = 14.562$), with no significant difference between them, as demonstrated in Figure 6. Innovations originating from isolated nodes face much more difficulty in establishing themselves in the entire graph, regardless of whether the recipients are more central or more isolated. The fact that innovations originating from isolated regions rarely achieve significant spread compared to innovations originating from central regions seems to give evidence for the demonstration of the third norm ("the larger area usually preserves the earlier stage"), as it suggests that innovations that emerge in isolated locations tend not to be able to spread across multiple locations, i.e., larger areas.

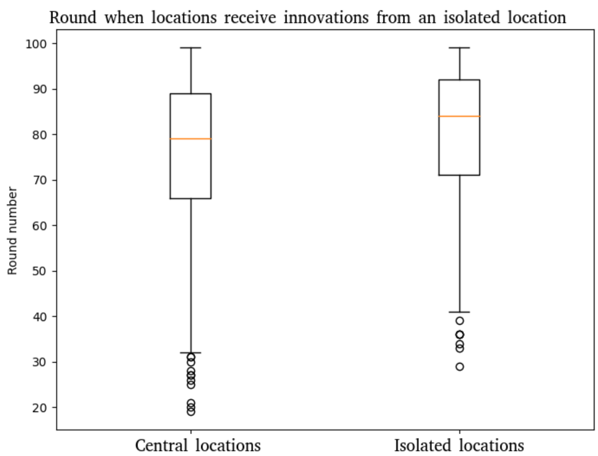

Figure 6: Round in which nodes (representing regions and cities) receive innovations from a more isolated node

What seems to happen is that innovations that arise from central areas have much greater ease of stabilization and propagation initially in equally central areas. However, after a certain amount of time, if the transmission is progressive, the form may also reach isolated zones, reaching a significant number of regions in comparison to the total across the rounds, potentially achieving the stage of full realization in all sociolects, as would be the ideal end for a new form on the S-curve. In contrast, forms that originate from isolated areas face long periods of stabilization in the sociolects of only a few regions and may never be transmitted progressively. These forms require much more time (measured through the number of rounds) to start consistent transmission and rarely reach a high percentage of regions (we remember that in this study, no form originating from the more isolated node reached all the nodes). As seen, the few regions that effectively received the innovative form received it without any apparent significant distinction between being more central and more isolated.

3.1 Peripherality and the case of West Iberia

Bartoli's second norm ("if one of two linguistic stages is found in peripheral areas and the other in a central area, the stage occurring in the peripheral areas is usually the earlier one") pertains, for instance, to areas like Portugal and Galicia, which in this graph are not necessarily isolated. For example, Portugal has three connections in our graph, ensuring a certain flow with the rest of the European continent. Nonetheless, it is geographically considered a peripheral area of the map, since the node representing Portugal is located in the westernmost point of the graph. According to Bartoli, the more archaic linguistic form would more likely be preserved in such a peripheral area. To explore this, we will determine in which round, on average, items from central and isolated regions reach the node corresponding to the territory of Portugal in our simulations.

Innovations originating from isolated regions took more time to reach the peripheral area ($\mu = 91.09$, $sd = 6.561$) – in this case, Portugal – than other isolated regions. In the case of innovations originating from central regions, we observed similar results to those of innovations coming from central regions reaching isolated ones ($\mu = 56.17$, $sd = 17.173$).

What our data seems to indicate is that the peripheral area indeed tends to keep the more archaic linguistic form, as Bartoli postulated. Innovations from central areas take over half of the simulation run, on average, to reach the peripheral area, while if the innovation comes from an isolated area, the item only reaches the peripheral area on average after the 90th round. Our data do not indicate a significant difference to designate a hierarchy between the peripheral area rule ($\mu = 56.17$) and the isolated area rule ($\mu = 63.64$). Certainly, if an area is both isolated and peripheral, the innovative form will face more difficulty in implementing itself in that location, regardless of whether its origin is central, isolated or peripheral.

Bartoli, and some of his readers over the years, initially proposed a hierarchy among the norms. It is not within the scope of this work to computationally model the hierarchy between the norms in detail. However, through this analysis, we can preliminarily suppose that there is no substantial empirical evidence in our model to delineate a hierarchy between the peripheral area norm and the isolated area norm. Although the average value of the isolated area has given a higher value, when we analyze the difference between them, along with the standard deviation values, we find that the differences are not significant enough to support the emergence of a hierarchy.

In any case, the peripheral area in our modeling – in the case examined here, Portugal – also preserved the more archaic linguistic form on average for a longer period compared to central areas and for a comparable time to purely isolated areas. These data support the explanation for the distinction between other Romance languages and the West Iberian languages, which maintain a number of earlier Latin traits in contrast to other Romance languages, retaining lexical, syntactic, and stylistic/orthographic forms, which might be more archaic due to geographical peripherality. The example of the verb *comer* was given before for illustrative purposes, but many other examples can be found in the literature on Romance linguistics (e.g. (Lausberg, 1956)).

3.2 An explanation for the archaism present in Mozarabic

Mozarabic was a set of Ibero-Romance varieties that developed in Al-Andalus, the part of the medieval Iberian Peninsula under Islamic control. This set of varieties likely became extinct by the end of the 14th century, being replaced by Andalusi Arabic as the main spoken language in the Muslimcontrolled south, in addition to the Romance varieties (especially Castilian) from the Christian kingdoms in the north that advanced southward during the Reconquista.

The speakers of Mozarabic referred to their language as 'ladino' due to the proximity of these linguistic varieties to Late Latin. Currently, the term 'ladino' is exclusively attributed to Judeo-Spanish, and the name 'Mozarabic' comes from the term 'Mozarab', which in Arabic means 'Arabized'. This term was used to refer to Christians in Al-Andalus. As evidenced by Wright (1982), indeed, in terms of phonology and morphology, Mozarabic is closer to Latin than other Romance varieties. This aspect even complicates its classification within this group since the language lacks many of the typical phonetic evolutions of Ibero-Romance languages – for example, the lenition of intervocalic consonants /p, t, k/, as in the Mozarabic words *lopa* (Port. *loba*, 'she-wolf'), *toto* (Port. *todo*, 'everything'), and *formica* (Port. *formiga*, 'ant'). In other peninsular Romance languages, changes occurred such as /p/ becoming /b/, /t/ becoming /d/, and /k/ becoming /g/, but not in Mozarabic.

Initially, we can correlate Mozarabic with Bartoli's fifth norm, which postulates that the variety that becomes extinct in favor of another is usually the oldest. This is indeed confirmed in the language competition between Mozarabic and the other Iberian Romance varieties. However, when we consider the medieval scenario of coexistence between Mozarabic, Andalusi Arabic, and the emerging Romance varieties to the north, we can leverage the results obtained in modeling to explain the archaism of Mozarabic in contrast to other Romance varieties, based on the concepts of isolation and periphery.

The locations under Arabic rule and consequently speaking Mozarabic can be considered peripheral due to their location on the Iberian Peninsula (increasingly restricted to the south due to the Reconquista) and isolated due to linguistic, cultural, and political barriers between them and the Christian kingdoms to the north. Linguistic innovations from other areas speaking Romance varieties reached the northern Christian kingdoms and their subsequent languages, including through cooperation during the Reconquista, but never reached Mozarabic. As seen in the simulations above, innovations from isolated locations are extremely costly to reach central and other isolated regions.

As discussed earlier, the establishment of the concepts of center, isolation and periphery in Bartoli's theory is always based on comparison. When we consider Rome and the Iberian locations, we see that Rome is a more central region, and Iberia is a peripheral region. In comparison to Rome, the regions speaking Mozarabic were also peripheral; however, they were possibly more peripheral than the regions under the dominion of the Christian kingdoms, fixed to the north of the peninsula. This is due to their proximity to regions in present-day France, speaking Romance varieties, and also more isolated due to the religious and political conflicts during the Reconquista, hindering linguistic and informational transmission between the conflicting locations. Thus, as seen in the simulations, the innovations present in the peripheral regions of the rest of the Iberian Peninsula did not reach the Mozarabic-speaking locations due to their extremely peripheral and isolated characteristics.

4 Conclusion

The objective of this study is to test the norms developed by the Italian linguist Matteo Bartoli regarding the relationship between language change and geographical space, for which we utilized a computational methodology called agent-based modeling. Specifically, we analyzed the spread of innovations among central, isolated and peripheral regions and sought to correlate the findings in our model with Bartoli's theoretical propositions. Despite the simplicity of the algorithm behind the model, we observe how, through simple rules of agent interaction, a phenomenon emerges that is compatible with the dynamics of languages in reality. As far as we are concerned, this is the first study to propose an agent-based modeling for the analysis of Bartoli's areal norms.

To perform the simulation, we constructed an algorithm based on a complex network of data from contemporary European sociodemographical space. The model developed appeared valid for testing three of Bartoli's five norms, which we were able to confirm in the simulations. Bartoli, in his works, provides examples of various cases in Romance linguistics; however, computational modeling can provide evidence for diverse cases with significant quantitative robustness, as discussed throughout this work. Bartoli believed in a hierarchy among the norms; however, in this study, we did not find significant empirical evidence to support such a hierarchy. We believe that future studies with specific modeling can be conducted to verify this hypothesis, requiring specific and more detailed investigations.

One aspect briefly touched upon in this work, which certainly plays a decisive role in Romance linguistics, is language contact. Certainly, areal and contact aspects are intertwined and mutually influence each other, as we can observe typological characteristics being transmitted areally through physical contact between speakers of different languages. Contact with Slavic languages certainly has its place in explaining the differentiation between Romanian and Moldavian compared to other Romance languages. The same can be said about Arabic influence on Portuguese, Celtic influence on French, but also on diachronic movements and not necessarily areal, such as the distinction between Galician and Portuguese (in addition to the influence of the Celtic substrate on the Galician and Astur-Leonese languages).

Future studies can be developed to verify the other two norms, using similar algorithms and networks, but enriched with data on population demographics and area size, for example. Furthermore, studies proposing an expansion of Bartoli's norms beyond Europe and the evolution of Romance languages can also be conducted, encompassing other language families and areal zones, as well as minority Romance varieties, or other historical contexts of linguistic diversity and contact – such as the Upper Rio Negro in the Brazilian Amazon and Oceania, for example, where areal factors certainly also play a fundamental role in linguistic distinction and change, but not exclusively (migration aspects also influence). Enriching the transmission algorithm can also be explored, taking into account more variables, such as prestige and population volatility, and not just purely geographical aspects of centrality and the number of connections for the transmission of information and linguistic items.

References

Jörn Albrecht. 1996. Neolinguistic school in Italy. In E. F. K. Koerner and R. E. Asher, editors, *Concise History of the Language Sciences: From the Sumerians to the Cognitivists*, pages 243–246. Pergamon.

Matteo Bartoli. 1925. *Introduzione alla Neolinguistica: Principi e Metodi*. Olschki, Geneva.

Matteo Bartoli. 1945. *Saggi di Linguistica Spaziale*. Vincenzo Bona, Turin.

Clay Beckner, Richard Blythe, Joan Bybee, Morten H Christiansen, William Croft, Nick C Ellis, John Holland, Jinyun Ke, Diane Larsen-Freeman, et al. 2009. Language is a complex adaptive system: Position paper. *Language Learning*, 59:1–26.

Richard A Blythe and William Croft. 2012. S-curves and the mechanisms of propagation in language change. *Language*, pages 269–304.

Xavier Castelló, Víctor M Eguíluz, Maxi San Miguel, Lucía Loureiro-Porto, Riitta Toivonen, Jari Saramäki, and Kimmo Kaski. 2008. Modelling language competition: bilingualism and complex social networks. In Andrew D M Smith, Kenny Smith, and Ramon Ferrer i Cancho, editors, *The evolution of language*, pages 59–66. World Scientific.

Xavier Castelló, Lucía Loureiro-Porto, and Maxi San Miguel. 2013. Agent-based models of language competition. *International Journal of the Sociology of Language*, 2013(221):21–51.

Christos Charalambous, David Sanchez, and Raul Toral. 2023. Language dynamics within adaptive networks: An agent-based approach of nodes and links coevolution. *arXiv preprint arXiv:2309.17359*.

Katia Chirkova and Tao Gong. 2014. Simulating vowel chain shift in xumi. *Lingua*, 152:65–80.

Marco Civico. 2019. The dynamics of language minorities: Evidence from an agent-based model of language contact. *Journal of Artificial Societies and Social Simulation*, 22(4).

Marco Antônio de Oliveira. 2018. Origem, propagação e resolução da variação linguística na perspectiva da linguagem como um sistema adaptativo complexo. *Caletroscópio*, 6:11–36.

Peter Dekker and Bart De Boer. 2020. Neural agent-based models to study language contact using linguistic data. In *4th NeurIPS Workshop on Emergent Communication: Talking to Strangers: Zero-Shot Emergent Communication*.

Zsuzsanna Fagyal, Samarth Swarup, Anna María Escobar, Les Gasser, and Kiran Lakkaraju. 2010. Centers and peripheries: Network roles in language change. *Lingua*, 120(8):2061–2079.

Jennifer Golbeck. 2015. Chapter 21 - analyzing networks. In Jennifer Golbeck, editor, *Introduction to Social Media Investigation*, pages 221–235. Syngress, Boston.

Aric Hagberg, Pieter Swart, and Daniel S Chult. 2008. Exploring network structure, dynamics, and function using networkx. Technical report, Los Alamos National Lab (LANL), Los Alamos, NM (United States).

K David Harrison, Mark Dras, and Berk Kapicioglu. 2002. Agent-based modeling of the evolution of vowel harmony. In *North East Linguistics Society*, volume 32, page 14.

Kelsey Jordahl, Joris Van den Bossche, Martin Fleischmann, Jacob Wasserman, James McBride, Jeffrey Gerard, Jeff Tratner, Matthew Perry, Adrian Garcia Badaracco, Carson Farmer, Geir Arne Hjelle, Alan D. Snow, Micah Cochran, Sean Gillies, Lucas Culbertson, Matt Bartos, Nick Eubank, maxalbert, Aleksey Bilogur, Sergio Rey, Christopher Ren, Dani Arribas-Bel, Leah Wasser, Levi John Wolf, Martin Journois, Joshua Wilson, Adam Greenhall, Chris Holdgraf, Filipe, and François Leblanc. 2020. geopandas/geopandas: v0.8.1.

Jinyun Ke, Tao Gong, William SY Wang, et al. 2008. Language change and social networks. *Communications in Computational Physics*, 3(4):935–949.

Heinrich Lausberg. 1956. *Romanische Sprachwissenschaft*, volume 1. W. de Gruyter.

Thomas Louf, David Sánchez, and José J Ramasco. 2021. Capturing the diversity of multilingual societies. *Physical Review Research*, 3(4):043146.

Witold Manczak. 1988. Bartoli's second "norm". In Jacek Fisiak, editor, *Historical Dialectology: Regional and Social*, Trends in Linguistics. Studies and Monographs, pages 349–355. Mouton de Gruyter, Berlin; New York; Amsterdam.

Antoine Meillet. 1923. Comptes rendus. *Bulletin de la Société de Linguistique de Paris*, 24.

Sergio J. Rey and Luc Anselin. 2007. PySAL: A Python Library of Spatial Analytical Methods. *The Review of Regional Studies*, 37(1):5–27.

Pablo Rosillo-Rodes, Maxi San Miguel, and David Sanchez. 2023. Modelling language ideologies for the dynamics of languages in contact. *arXiv preprint arXiv:2307.02845*.

Luc Steels. 2000. Language as a complex adaptive system. In *International Conference on Parallel Problem Solving from Nature*, pages 17–26. Springer.

Celina Troutman, Brady Clark, and Matthew Goldrick. 2008. Social networks and intraspeaker variation during periods of language change. *University of Pennsylvania Working Papers in Linguistics*, 14(1):25.

Uriel Weinreich, William Labov, and Marvin Herzog. 1968. *Empirical foundations for a theory of language change*. University of Texas Press, Austin.

Søren Wichmann. 2008. The emerging field of language dynamics. *Language and Linguistics Compass*, 2(3):442–455.

Roger Wright. 1982. *Late Latin and Early Romance in Spain and Carolingian France*. University of Liverpool (Francis Cairns, Robin Seager), Liverpool.

Dunja Šešelja. 2023. Agent-Based Modeling in the Philosophy of Science. In Edward N. Zalta and Uri Nodelman, editors, *The Stanford Encyclopedia of Philosophy*, Winter 2023 edition. Metaphysics Research Lab, Stanford University.