Language dynamics within adaptive networks: an agent-based approach of nodes and links coevolution

Christos Charalambous*, David Sanchez and Raul Toral

Institute for Cross-Disciplinary Physics and Complex Systems IFISC (UIB-CSIC), Campus Universitat Illes Balears, Palma, Spain

Motivated by the dramatic disappearance of endangered languages observed in recent years, a great deal of attention has been given to the modeling of language competition in order to understand the factors that promote the disappearance of a language and its unfolding dynamics. With this in mind, we build on existing network models of language competition in bilingual societies. These models deal with the interplay between the usage of a language (link state) and the preference or attitude of the speakers towards the language (node state). In this work, we allow for the case where agents have the freedom to adapt their local interactions in accordance with their language preference. This is modeled by introducing a local rewiring mechanism triggered by the dissatisfaction of an agent with its usage of a given language. Our numerical simulations show that permitting this freedom to agents likely results in linguistically segregated communities for small network sizes. However, for networks of sufficiently large size, the extinction of one of the languages is the most probable scenario. Furthermore, we analyze how the fraction of minority speakers changes with the system size and we find that this fraction grows as the total population increases, which is consistent with existing data. Overall, the results of this work help us understand the impact of speakers' preferences and choices in the complex language landscape of bilingual societies.

KEYWORDS

language competition, agent-based modeling, adaptive networks, coevolving networks, language dynamics, coupled nodes and links, rewiring

1 Introduction

In recent years, ideas and tools from the complex systems framework have been extensively used to model various types of language dynamics, such as language evolution (or how the structure of language evolves) (Steels, 2011), language cognition (or the way in which the human brain processes linguistic knowledge) (Edelman and Waterfall, 2007), and language competition (or the dynamics of language use in multilingual communities) (Sole et al., 2010). Particular emphasis has been given in language competition, since a mass extinction of languages is anticipated by linguists (Krauss, 1992). The focus, in understanding the factors that govern the competition of languages, lies in social interactions. One of the main goals in the study of language competition is to determine the mechanisms and conditions that lead either to the coexistence of different languages or to the dominance of a single language. This is of paramount importance for the design of appropriate revitalization efforts and language planning policies.

Various approaches have been considered to model the dynamics of language competition depending on the depth of societal description under consideration. For example, at a macroscopic level, where only the population fractions speaking each language are considered, there are approaches that make use of evolution equations borrowed from ecological models (Mira and Paredes, 2005) whereas others consider reaction-diffusion equations (Kandler, 2009; Patriarca and Heinsalu, 2009). These theoretical frameworks, amenable to an analytical treatment, are able to describe the observed trends in language growth or decline when compared to empirical data. One of the pioneering works in this direction was put forward by Abrams and Strogatz (2003) who studied the dynamics of endangered languages. In this model, individuals in a binary-state society can speak either language A or B, allowing only for the possibility of societal bilingualism, i.e., the coexistence of two different monolingual groups (Appel and Muysken, 2006). Subsequent generalizations of the original model considered the existence of bilingual groups (Wang and Minett, 2005).

In general, the above approaches work well as long as individual behavior can be neglected and one focuses on aggregated patterns. If this is not the case, one should resort to a more microscopic description of the system, considering approaches within the framework of agent-based models (ABM) (Prochazka and Vogl, 2017). In this case, both the various individual characteristics that provide probabilities to switch to another language group and the detailed interactions between agents are taken into account, hence allowing for a description of the shifting mechanisms at an individual level. This permits a deeper understanding of how changes in different social factors can affect the dynamics of real-world systems (Williams, 2018). In their work, Minett and Wang (2008) focus on ABMs in language competition models, exploring the effect of language status and education policies on the maintenance of a minority language. Another example would be that of the work of Castello et al. (2013), which studies the dynamics of a community where two languages are used, each spoken by both monolingual and bilingual speakers. These works examine how language status and the individual likelihood to shift to another language impact language growth or decline. An alternative ABM framework is provided by Patriarca et al. (2012), which considers a game-theoretical approach to language competition modeling and discusses two different strategies encountered among minority-language speakers.

Here, we build on an ABM suggested in a previous work by Carro et al. (2016), which comprises a natural way of accounting for speakers who are potentially able to speak both languages. The main idea is that language is treated as a property of the interactions between individuals, rather than a property of the speaker. In this way, language serves as a means of communication, embodying distinct states for the links within the network. Unlike the work of Castello et al. (2013), bilingualism is not an intermediate state but the result of individuals using different languages in different interactions. Furthermore, the node of the state can now encode the preference towards one of the languages, hence allowing for various degrees of bilingualism in the system. This preference in fact can be the result of a number of factors, such as the level of competence in that language or the degree of cultural attachment and affinity with the respective speech community. Our main

contribution as compared to the work of Carro et al. (2016) is that in our model, as explained below, we allow for the possibility that speakers adapt the topology of their network of interactions if they are not satisfied with particular links (rewiring). Further, contrary to most rewiring mechanisms considered in the literature, we do not introduce a new global parameter for the propensity of the individuals to rewire, but rather this depends locally on the preference of each individual.

Models that address the dynamics of link states have received increasing attention from different research areas such as social balance theory (Antal et al., 2005), community detection (Traag and Bruggeman, 2009), network controllability (Nepusz and Vicsek, 2012), and opinion formation (Saeedian et al., 2019). Relevant to our work, Fernández-Gracia et al. (2012) implement a majority rule for link states, while Carro et al. (2014) develop a coevolution model that couples the aforementioned majority rule dynamics of link states with the evolution of the network topology. As previously discussed, in the context of language competition, Carro et al. (2016) analyze coupled node and link dynamics. Their main result is that, in contrast to most of the previously proposed models, the extinction of one of the languages is not an inevitable outcome of the dynamics. On the contrary, there is a wide range of possible asymptotic configurations, including not only extinction states but also frozen and dynamically trapped coexistence states, the probability of extinction being a decreasing function of the population size. Furthermore, recent studies consider the coupling of all three dynamics, i.e., nodes, links and network topology, as in the work of Saeedian et al. (2020). Here, the link state is considered to be either attractive or repulsive, and it is found that repulsive interactions are a major cause of social polarization.

In this work, we assume that node and link dynamics are coupled in the same way as in the work of Carro et al. (2016) but we introduce a rewiring mechanism whose activation depends on the speakers' preferences. The main result of our work is that rewiring itself can contribute to the survival of both languages in the long time limit, with both languages having a significant presence. Here, we define survival as having over 5% of the population speaking that language. We estimate the survival probability, the relative sizes of the two communities as well as that of the bilingual individuals (individuals that are using both languages), as a function of both the system size and the node update probability. We find that increasing the system size decreases the probability of language coexistence. The contrary holds true for the node update probability. However, by increasing either the system size or the node update probability, in those cases where the two languages end in a coexistence state in the long-time limit, the size of the two communities approaches each other, while the relative size of the community of bilinguals remains constant. This is a remarkable finding that is consistent with data taken from today's bilingual societies. We also define a quantity, termed satisfaction, that captures the alignment of agents' preferences with their spoken language, and we see that, independently of the rewiring mechanism, agents appear almost always to be completely satisfied, except the bilinguals.

The work is organized as follows. In Sec. 2 we present the model, explaining the node and link dynamics and how these are coupled. We also introduce the specifics of the rewiring mechanism and how these interfere with the topology of the network. The structural

constraints imposed by the definition of the model as well as the particularities of the networks used for the numerical simulations are also described in this section. In Sec. 3 we present the results of our studies in the long time limit. Finally, Sec. 4 summarizes the main results of the work.

2 The model

We consider a population of N speakers who are linguistically interacting among themselves. We represent this on a network, where nodes correspond to the speakers and links correspond to their interactions. Both nodes and links are characterized by state variables. The node state is a continuous variable $x_i \in [0, 1]$ that represents the preference of speaker i towards language A ($1 - x_i$ being their preference towards language B). Therefore, $x_i = 1$ indicates an absolute or extreme preference for language A whereas $x_i = 0$ an absolute or extreme preference for language B. The link state represents the language used in the interaction between agents and is described with a discrete binary variable $S_{ij} \in \{0, 1\}$ such that $S_{ij} = 1$ denotes the usage of language A between agents i and j, and $S_{ij} = 0$ corresponds to the usage of language B.

States of nodes and links evolve asynchronously (single node or link update at each step) according to stochastic rules that we now describe. First, with probability q a random node is chosen for an update, and with the complementary probability $1 - q$ a random link is chosen instead. It should be noted that in the node selection, we assume a link-update dynamics, as described in the work of Suchecki et al. (2004). This means that in order to select a random node, a link is selected first, and then one of the two nodes at the ends of the link is randomly selected. A fundamental difference of this procedure with traditional node-update schemes (Castelló et al., 2006; Patriarca et al., 2012) is that nodes with a high degree are chosen more often to be updated.

The probability q sets the relationship between the evolution time scale of the speakers' preferences and the time scale at which the language used in conversations changes. In the original version of the model without rewiring (Carro et al., 2014), it was shown that the parameter q simply affects the speed with which the asymptotic state is reached, but not its main features. As we later show, when rewiring is introduced the parameter q does play a significant role. The posterior evolution depends on whether we have chosen to update a node or a link.

As node and link updates are both present, there appears an ambiguity as to how to measure time in the simulations. Here we simply assume that the time unit, one Monte Carlo step, corresponds to N updating attempts, irrespective of nodes or links.

2.1 Network structure

Several studies have shown that real social networks are characterized by a great number of triangles, resulting in high values of the clustering coefficient (Dorogovtsev and Mendes, 2003; Newman and Park, 2003; Newman, 2010; Foster et al., 2011; de Simón et al., 2013). To fulfill this, we here consider networks with many triangles generated according to the algorithm proposed by Klimek and Thurner (2013). This is a

socially inspired algorithm and is based on the principle of triadic closure which states that individuals tend to make new acquaintances among friends of friends. There are three different mechanisms present in this algorithm: a) random link formation, b) triadic closure (link formation between nodes with a common neighbor) and c) node replacement (removal of a node with all its links and introduction of a new node with a certain number of links). With a suitable set of parameters, this model can reproduce data from a well-studied massive multiplayer online game (Szell et al., 2010; Szell and Thurner, 2010; 2012; Klimek et al., 2016). However, as in the work of Carro et al. (2016) we modify the aforementioned algorithm in order to guarantee that initially all nodes participate in at least one triangle. We also use the same parameter values found by Klimek and Thurner (2013) when calibrating their algorithm to the friendship network of the above-mentioned online game: a probability of triadic closure $c = 0.58$ ($1 - c$ being the probability of random link formation) and a probability of node replacement $r = 0.12$.

2.2 Evolution of node states

The update of one node state, i.e., the preference of an agent for either language A or B is determined by the languages spoken by its neighboring agents during their interactions. After randomly selecting two of its neighbors engaged in communication, the agent adjusts its preference toward the language being utilized by this specific pair of neighbors. Effectively, we are assuming that group relationships take place on triangular social structures [see Serrour et al. (2011) for a study on the relationship between communities and triangles]. More precisely, when a node i is chosen for updating, its state x_i evolves according to the following rules.

1. With probability $\frac{T_i^A}{T_i}$, set $x_i \rightarrow \min (x_i + \Delta x_i, 1)$.
2. With probability $1 - \frac{T_i^A}{T_i}$, set $x_i \rightarrow \max (x_i - \Delta x_i, 0)$.

Here, $\Delta x_i = 1/k_i$, with k_i the degree of node i (its total number of neighbors); T_i is the total number of links between the neighbors of node i and T_i^A is the number of those links in state 1, i.e., those in which language A is used. We should note that Δx equals $1/k_i$ because we use a link-update scheme and nodes with many links are updated more often. Thus, these nodes also increase or decrease their state in smaller steps. Another justification for the form of Δx is that nodes with fewer links tend to belong to fewer triangles and, therefore, each of them has a stronger influence on the node.

2.3 Evolution of link states

Two factors contribute to the update of the language used between two agents. First of all, we assume that the interaction between two given speakers tends to take place in the language most often used by both of them in their communications with other speakers. We postulate this as a way to encode the tendency of speakers to use as few languages as possible since the usage of more languages requires more cognitive effort (Jackson et al., 2001; Abutalebi and Green, 2007). For each link i–j we define the

majority pressure for language A, F_{ij}^A, as the fraction of the number of interactions in which language A is used,

$$F_{ij}^A = \frac{k_i^A + k_j^A - 2S_{ij}}{k_i + k_j - 2},\tag{1}$$

where k_i^A stands for the number of neighbors in which speaker i uses language A, and k_i is the total number of neighbors, the degree, of node i. A similar definition for language B, using k_i^B as the number of neighbors in which speaker i uses language B, leads to $F_{ij}^B = 1 - F_{ij}^A$.

The second important factor that should be taken into account when deciding the language used is the preference of speakers towards one language or the other. Combining the preferences of both participants in each interaction i–j, we define the link preference for language A as

$$P_{ij}^A = \begin{cases} \dfrac{x_i x_j}{D}, & \text{if } D \neq 0, \\ \dfrac{1}{2}, & \text{otherwise.} \end{cases}\tag{2}$$

with $D = x_i x_j + (1 - x_i)(1 - x_j)$. Following Eq. 2, when a link contains extreme preferences in both languages (either $x_i = 0$ and $x_j = 1$ or $x_i = 1$ and $x_j = 0$), the link preference is set to 1/2. In addition, P_{ij}^A yields 1 if both agents prefer language A ($x_i = x_j = 1$) and 0 if both agents prefer language B ($x_i = x_j = 0$). Furthermore, if one of the nodes is neutral with respect to language A, $x_i = 1/2$, then the link preference is set to match the preference of the other node, x_j. Finally, the link preferences satisfy $P_{ij}^A(x_i, x_j) = 1 - P_{ij}^A(1 - x_i, 1 - x_j)$, reflecting the symmetry of the definition with respect to both languages. The definition for language B obeys $P_{ij}^B = 1 - P_{ij}^A$.

When a link i–j is selected for updating, its new state is chosen according to the following rules.

1. If the majority pressure for language A is larger than the link preference for language B, then language A is chosen.
2. If the majority pressure for language A is smaller than the link preference for language B, then language B is chosen.
3. If there is a tie between both languages, then one of them is randomly chosen.

Leveraging the previously discussed symmetry between both languages, these rules can be mathematically expressed as

$$S_{ij} = \begin{cases} 1 \ (\text{language A}), & \text{if } F_{ij}^A > P_{ij}^B, \\ 0 \ (\text{language B}), & \text{if } F_{ij}^A < P_{ij}^B, \\ 0 \text{ or } 1 \text{ randomly}, & \text{if } F_{ij}^A = P_{ij}^B. \end{cases}\tag{3}$$

Note that if all the speakers' preferences are fixed as $x_i = 1/2$, $\forall i$, then all link preferences are also set to 1/2 and we recover the majority rule for link states analyzed by Fernández-Gracia et al. (2012): the state of a link is updated to the state of the majority of its neighboring links. With freely evolving preferences of the speakers, on the contrary, the threshold for a state to be considered a majority is not universal anymore and fixed at 1/2, but becomes local and dynamic. It is also noteworthy that, in a static network topology, speakers with extreme preferences ($x_i = 0$ or $x_i = 1$) impose their preferred language in all their conversations, except

when they have to interact with a speaker with an extreme but opposite preference, in which case the language for their interaction is randomly selected. Thus, we are implicitly assuming that all speakers are able to use both languages, which makes our model suitable for bilingual societies. When a dynamic topology is considered, the speakers with extreme preferences do not necessarily impose their preferred language, since dissatisfied agents can instead rewire and form a new pair to use their preferred language.

2.4 Rewiring dynamics

We thus far have described the model developed in the work of Carro et al. (2016). The main novelty of our work is the introduction of the possibility that agents can change the local structure of their interactions in case of dissatisfaction. For illustrative purposes, the dynamics which we now describe are depicted in Figure 1.

When, as a result of the application of the conditions in Eq. 3 the link is required to adapt its state, we allow for a link rewiring, maintaining the link's current state. The agent who retains the link is chosen based on being the one primarily dissatisfied with the expected change that results from the application of the conditions specified in Eq. 3. For instance, if the proposed change was to switch the link state from 1 (language A) to 0 (language B), then the agent with the largest value of preference x_i (i.e., with the largest preference towards language A) would be selected to rewire to another agent, say ℓ, keeping the link state $S_{i\ell} = 1$.

This rewiring mechanism is probabilistic. In other words, either the proposed change of the state of the link is performed, or rewiring takes place. The latter occurs with a probability proportional to the preference of the most dissatisfied agent, i.e., the one who retains the link. The rewiring probability depends on the values of the nodes' preferences according to the following rules.

1. If the proposed link update was S_{ij}: $0 \to 1$, then set rewiring probability $\to 1 - \min(x_i, x_j)$.
2. If the proposed link update was S_{ij}: $1 \to 0$, then set rewiring probability $\to \max(x_i, x_j)$.

The newly selected agent ℓ is randomly chosen from the set of neighbors of the neighbors of the agent exhibiting the highest level of dissatisfaction, say i, which are not themselves already neighbors of i. We also require that the new link i–ℓ satisfies the conditions given by Eq. 3, i.e., $P_{i\ell}^B \leq F_{i\ell}^A$ if $S_{i\ell} = 1$, or $P_{i\ell}^B \leq F_{i\ell}^A$ if $S_{i\ell} = 0$. By choosing ℓ from the set of neighbors of neighbors of i we ensure that, after rewiring, the number of triangles remains unaltered. However, if the proposed rewiring implies that the agent j losing the link does not belong anymore to any triangle, then the rewiring is discarded. In this way, we adhere to the original constraint that all nodes must belong to at least one triangle. Finally, we note that, unlike the original model of Carro et al. (2016), the possible preference values are not defined as multiples of the inverse degree $1/k_i$. Hence, when rewiring is considered, the node preference remains unchanged even though its degree k_i has changed.
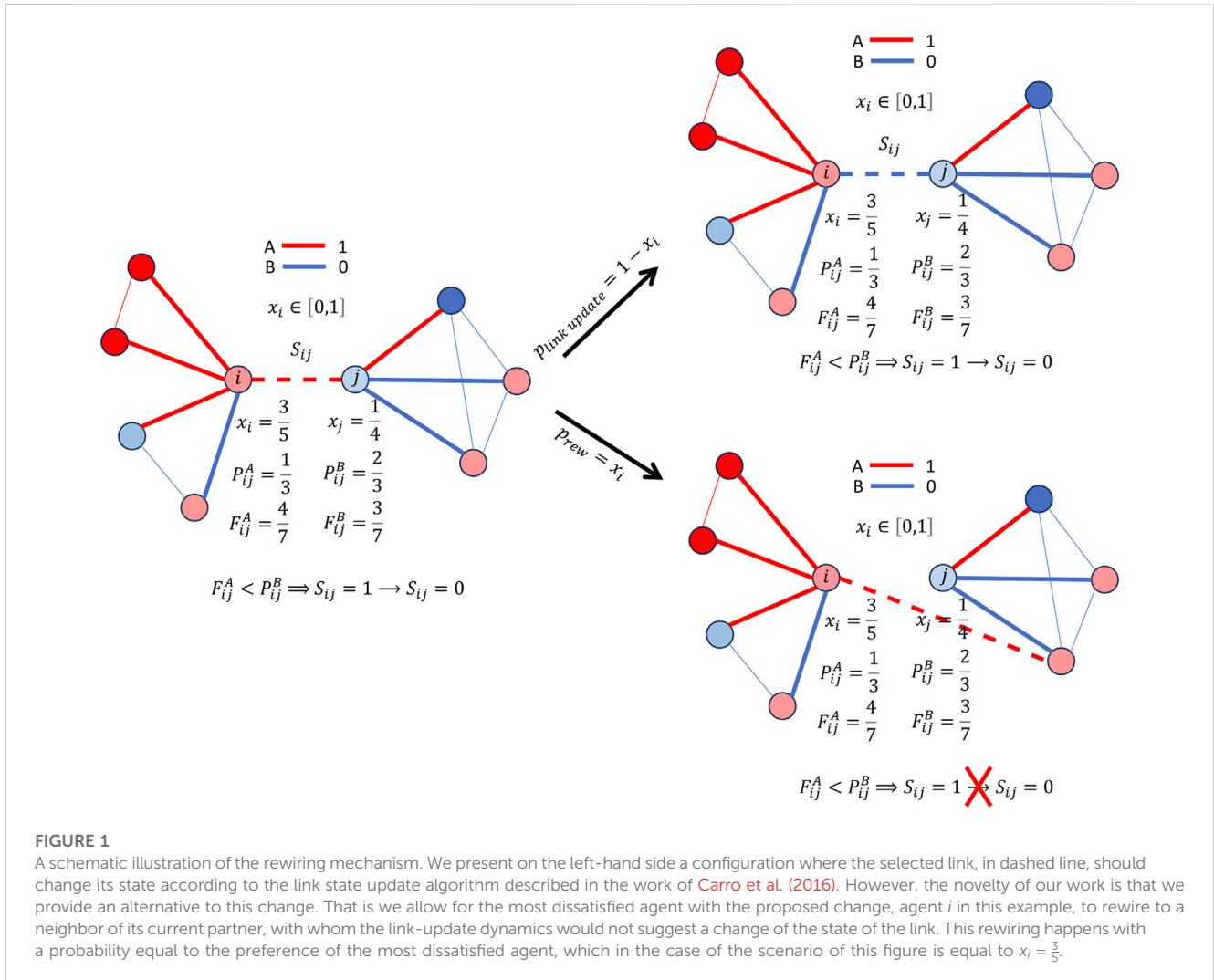
FIGURE 1
A schematic illustration of the rewiring mechanism. We present on the left-hand side a configuration where the selected link, in dashed line, should change its state according to the link state update algorithm described in the work of Carro et al. (2016). However, the novelty of our work is that we provide an alternative to this change. That is we allow for the most dissatisfied agent with the proposed change, agent i in this example, to rewire to a neighbor of its current partner, with whom the link-update dynamics would not suggest a change of the state of the link. This rewiring happens with a probability equal to the preference of the most dissatisfied agent, which in the case of the scenario of this figure is equal to $x_i = \frac{3}{5}$.

3 Results

Now we explore the convoluted dynamics of the system described above through numerical simulations. In particular, we are interested in the asymptotic configurations reached after a long time. All our results reach a final time $t_{\text{final}} = 10^5$ in units of Monte Carlo steps, although we have checked that larger times do not change significantly the main results for the system sizes analyzed here. When not stated explicitly, the network size considered is $N = 1000$, the node update probability is set to $q = 1/2$ and we average the results over a number of independent runs (typically 100).

Since a major goal is to examine the possibility of both languages surviving in the long term, we first consider the density of nodal interfaces (Fernández-Gracia et al., 2012; Carro et al., 2014), defined as the fraction of pairs of connected links that show different states,

$$\rho = \frac{\sum_{i=1}^{N} k_i^A k_i^B}{\sum_{i=1}^{N} k_i (k_i - 1)/2}. \tag{4}$$

The order parameter ρ is a measure of local order in the system, taking the value $\rho = 0$ when all connected links share the same state

and $\rho = 1/2$ for a random distribution of link states. Alternatively, ρ can also be seen as the usual density of active links (fraction of links connecting nodes with different states) when considering the line-graph of the original network (van Rooij and Wilf, 1965; Chartrand and Stewart, 1969; Mańka-Krasoń et al., 2010; Krawczyk et al., 2011; Fernández-Gracia et al., 2012; Carro et al., 2014).

We next introduce the fraction of links in the minority language as a non-local measure characterizing the system in terms of link states:

$$m = \frac{\min\left(\sum_i k_i^A, \sum_i k_i^B\right)}{\sum_i k_i}. \tag{5}$$

Minority language is obviously defined as the language that is used less in the network. In this way, even if a population majority uses a certain language in part of their interactions, we will still consider it to be the minority language if only a minority of the total number of communications actually takes place in that language.

Finally, we characterize the system in terms of node states by introducing the average preference of the speakers for the minority language,
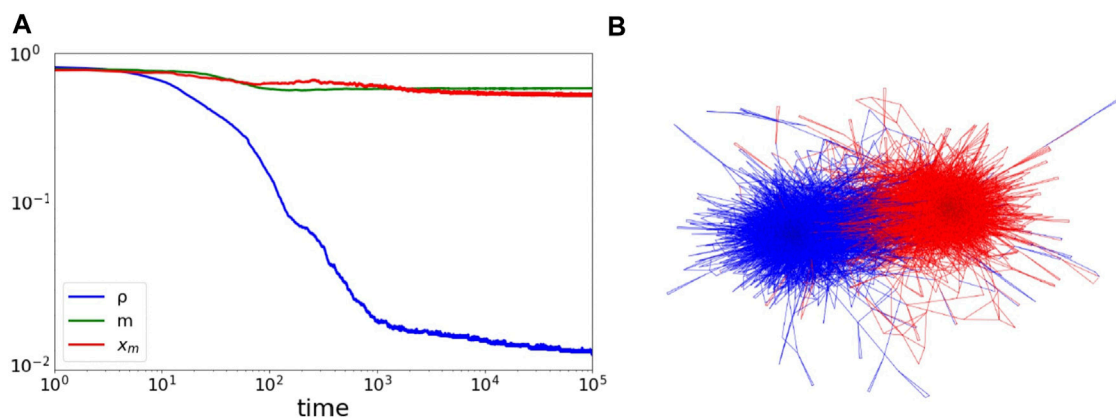
FIGURE 2
(A) Effect of rewiring on the order parameter ρ, the fraction of links in the minority language, m, the average preference of the speakers for the minority language, x_m, as well as on the final structure of the network. We see that the values of m and x_m remain high (at the scale presented here the two are approximately equal and they overlap), while the value of ρ decreases, indicating an increase of the number of nodes connected with links of the same language. At the same time, the fraction of links in the minority language is not reduced, indicating that the minority language comprises a significant fraction of the network. Furthermore, the average preference of the speakers for the minority language is not reduced. **(B)** Network with rewiring in the long time limit. We depict with red color the edges in the state $S_{ij} = 1$ (language A) and with blue color the edges in the state $S_{ij} = 0$ (language B). We observe a clear segregation between the two communities, facilitating the coexistence of both languages. Parameters: $N = 1000$, $q = 0.5$, $t_{\text{final}} = 10^5$. Results have been averaged over 100 runs.

$$x_m = \begin{cases} N^{-1}\sum_i x_i, & \text{if } \sum_i k_i^A \leq \sum_i k_i^B, \\ N^{-1}\sum_i (1 - x_i), & \text{otherwise.} \end{cases} \quad (6)$$

This is a measure of the level of attachment for speakers of the minority language with their preferred language.

The time evolution of these three measures is presented in Figure 2A. All realizations start from a random initial distribution of states for both nodes and links. Regarding the links' state, each link is assigned the value 0 or 1 with an equal probability, i.e., with probability 0.5. Regarding the agents' preferences, each agent is assigned a preference value between 0 and 1 selected uniformly at random. As a consequence, the three measures ρ, m and x_m start from the value 0.5. Then, the local order parameter ρ monotonically decreases toward zero, which is a sign that most nodes are only connected to links of the same state. We checked the behavior of ρ for a number of different values for the system size N and the node update probability q and it behaves approximately the same. The other two parameters, which are global in nature, for the specific simulation plotted here, remain constant. This is because the network splits into two approximately equal-sized communities, with each community speaking its own language. We confirm this by plotting the final state of the network in Figure 2B. This is in stark contrast with the results of Carro et al. (2016), where the authors find that in most cases a network of 1000 nodes either reaches the entire dominance of a single language or at most ghetto areas where one of the languages is spoken but as a minority language. We thus infer that the introduction of the rewiring mechanism significantly changes the final state of the system. In the next subsections, we examine this effect in more detail.

3.1 Rewiring effect on network structure

The main change in the network structure due to rewiring is shown in Figure 3. We observe that the number of nodes with degree 2 increases with respect to its initial value, but the rest of the degree distribution remains basically unaltered by the rewiring. We also find that rewiring flattens the average clustering coefficient distribution. It follows from these observations that the rewiring effect increases the number of triangles in the network.
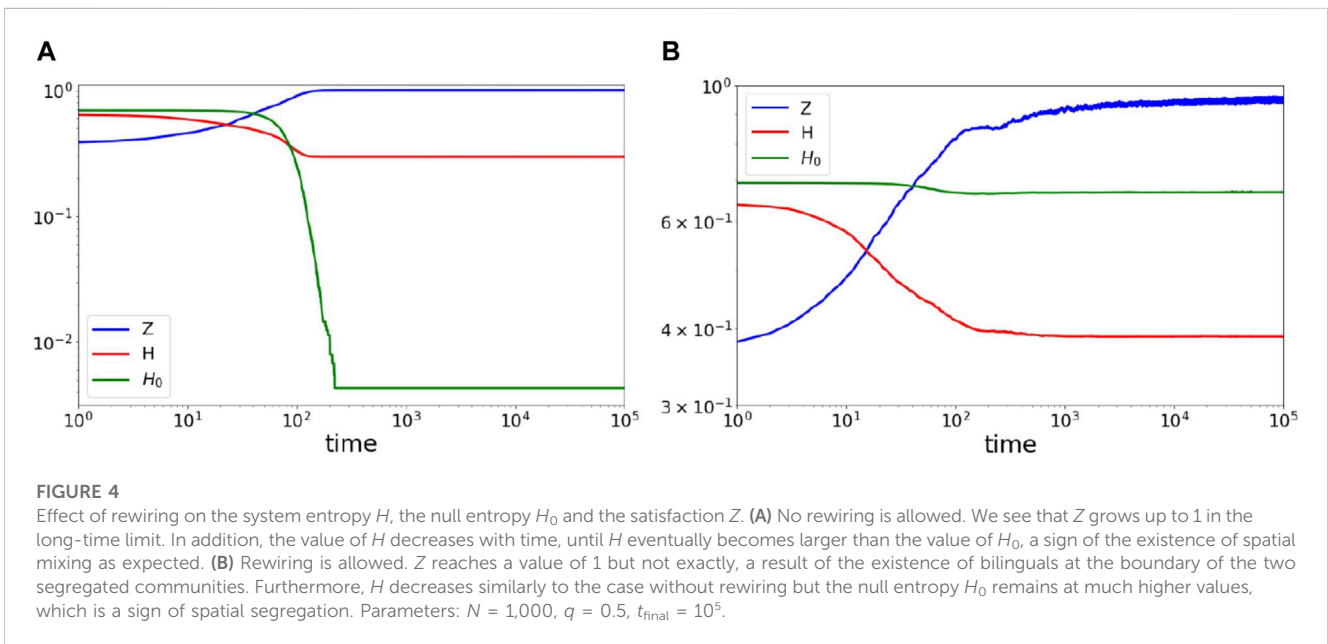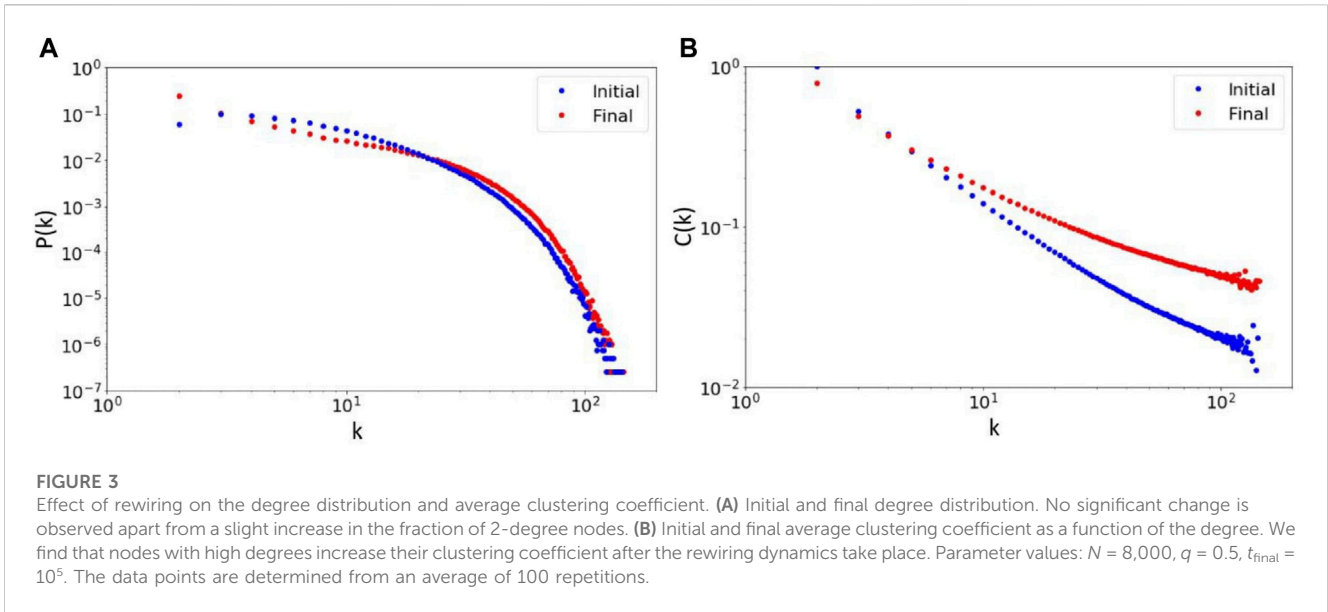
3.2 Entropy and satisfaction

To better understand the formation of linguistically polarized communities due to rewiring in convoluted node and link dynamics, we now introduce the entropy, which quantifies the spatial mixing or segregation in multigroup societies (Reardon and Firebaugh, 2002; Louf et al., 2021). Segregation can be extracted from the entropy of an agent's interactions with a given group on average, compared to the entropy of the interactions of the whole population. The latter constitutes the null model, in which interactions in different languages are assigned anywhere on the network randomly and independently of the rest of the interactions of each individual.

Let H_i be the entropy associated with the states of the links attached to the ith individual:

$$H_i = - \sum_{s\in\{A,B\}} p(s|i)\log p(s|i), \quad (7)$$

where

$$p(s|i) = \frac{k_i^s}{k_i}.$$

FIGURE 3
Effect of rewiring on the degree distribution and average clustering coefficient. **(A)** Initial and final degree distribution. No significant change is observed apart from a slight increase in the fraction of 2-degree nodes. **(B)** Initial and final average clustering coefficient as a function of the degree. We find that nodes with high degrees increase their clustering coefficient after the rewiring dynamics take place. Parameter values: $N = 8,000$, $q = 0.5$, $t_{final} = 10^5$. The data points are determined from an average of 100 repetitions.

FIGURE 4
Effect of rewiring on the system entropy H, the null entropy H_0 and the satisfaction Z. **(A)** No rewiring is allowed. We see that Z grows up to 1 in the long-time limit. In addition, the value of H decreases with time, until H eventually becomes larger than the value of H_0, a sign of the existence of spatial mixing as expected. **(B)** Rewiring is allowed. Z reaches a value of 1 but not exactly, a result of the existence of bilinguals at the boundary of the two segregated communities. Furthermore, H decreases similarly to the case without rewiring but the null entropy H_0 remains at much higher values, which is a sign of spatial segregation. Parameters: $N = 1,000$, $q = 0.5$, $t_{final} = 10^5$.

We average Eq. 7 over all individuals,

$$H = \sum_{i=1,...,N} p(i)H_i, \qquad (8)$$

where

$$p(i) := \frac{\sum_{s\in\{A,B\}}k_i^s}{\sum_{i,s}k_i^s} = \frac{k_i}{2L}, \qquad (9)$$

with L the total number of links. Eq. 8 is to be compared with the null model, which assumes a non-segregated society:

$$H_0 = -\sum_{s\in\{A,B\}} \frac{L_s}{L}\log\left(\frac{L_s}{L}\right), \qquad (10)$$

where L_s is the number of links at state s. This quantity tends to the value 0 when one of the languages dominates over the other, while it is equal to $-\log(0.5)$ when there is an equal number of links in language A and B in the network.

We plot both H and H_0 in Figure 4B and compare with the case of no rewiring in Figure 4A. We see that at the final state of the dynamics without rewiring the system entropy is higher than the value predicted with the null model. This is a sign of mixing in the network. In addition, the value of the null model tends to 0, which shows that one language dominates over the other in the network. On the other hand, in Figure 4B, the entropy with rewiring is much lower than that of the null model, which indicates heterogeneity in the network, as a result of the fact that the network indeed splits into
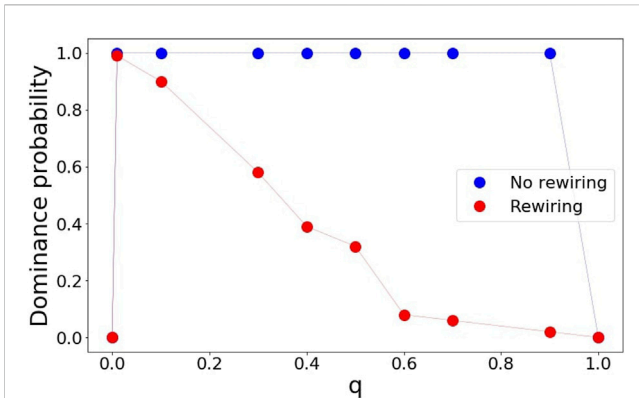
FIGURE 5
Fraction of simulations that end up in the dominance state, i.e., where 95% of the network agents only speak one language. Blue curve: no rewiring is allowed. The dominance scenario is fully present independently of $q \neq 0, 1$. Red curve: rewiring is allowed. Dominance probability decreases q increases, vanishing when the update probability is high. As expected, in both cases the dominance probability vanishes for either $q = 0$ or $q = 1$. The parameters used were: N =1,000, $t_{final} = 10^5$ and 100 simulations are considered to calculate the probability.

two groups. The fact that the null entropy's value is close to $-\log(0.5)$ shows that there is a significant presence of both types of links in the network.

The network can be further characterized by calculating the average satisfaction of the agents:

$$Z = \frac{1}{2L} \sum_{i,j} \left(x_i + S_{ij} - 1 \right)^2, \tag{11}$$

where $\sum_j \left(x_i + S_{ij} - 1 \right)^2$ is understood as a measure of the local satisfaction at node i since this quantifies the agreement between the local environment of the agent and its language preference.

Figure 4 also shows Z. We find that while the final satisfaction reaches its maximum value 1 when no rewiring is allowed

(see Figure 4A), the introduction of rewiring slightly reduces the maximum value as can be seen in Figure 4B. This is a consequence of the existence of bilingual agents in the boundaries of the two groups. As these agents lie at the intersection of the two groups, their preferences as well as their links' states can fluctuate. As a consequence, the local environment and their preferences do not exactly match.

3.3 Dependence on the node update probability

We recall that, despite the fact that all agents are able to speak both languages, an agent is considered bilingual only if it contains both A and B type links. The random initial conditions that we assumed above hence imply a high number of bilingual agents initially. In addition, as explained above, we define language dominance as the scenario where one of the languages, either A or B, is spoken by more than 95% of the agents that use a single language. With this in mind, in Figure 5 we plot the probability that a simulation ends in a dominant state as a function of the node update parameter q. As we can observe, for the case without rewiring (blue dots), this probability remains unchanged as q increases (except for the extreme cases $q = 0$ and $q = 1$. For the latter case, i.e., when the state of the links does not change, we see that we have a high number of bilingual agents, which is a consequence of the initial conditions we assume.). However, in the case where we do allow for rewiring (red dots), we obtain that this probability reduces as fewer updates of the link states occur.

In Figure 6, we study the final form of the system depending on the parameter q, for those networks in which language dominance is not reached. We now focus on the case with rewiring only, since the case without rewiring showed no significant changes in Figure 5. We find that for the extreme values $q = 0$ and $q = 1$, the network is mostly composed of bilingual agents, although for $q = 1$ the number of bilingual agents
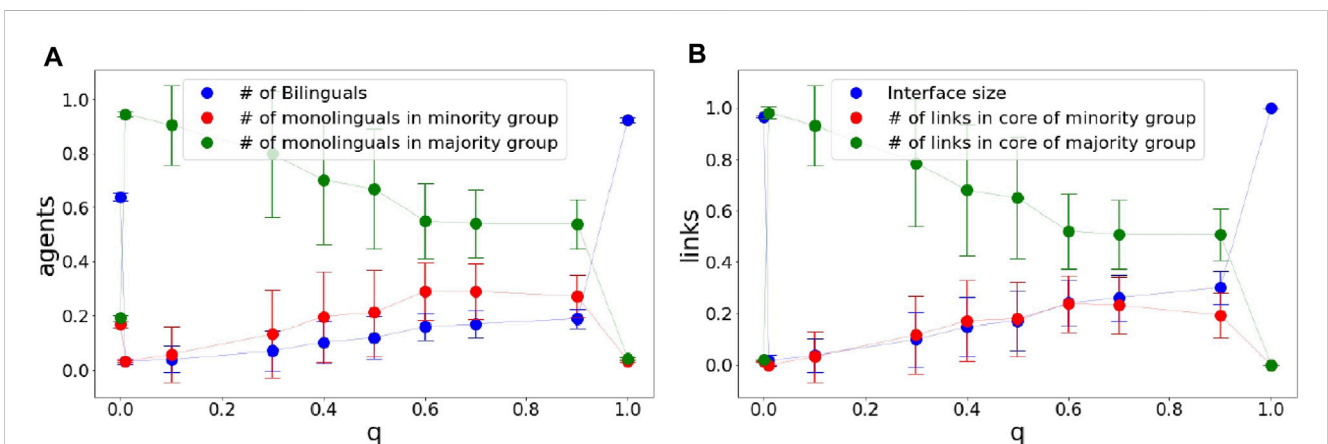

FIGURE 6
Language group populations as a function of the node update parameter q. For the extreme values of $q = 0$ and $q = 1$, i.e., when only links are updated or when only nodes are updated, the network is mixed. In other words, bilingual agents prevail. As soon as a mixture of updates is allowed ($0 < q < 1$), we see the appearance of a majority group. The network reaches a polarized state (segregation). Furthermore, we see that the number of bilingual agents **(A)** as well as that of interface links **(B)** increases as q increases. At the same time, the sizes of the minority and majority groups approach each other. The displayed error bars correspond to one standard deviation. Parameters: N =1,000, $t_{final} = 10^5$.
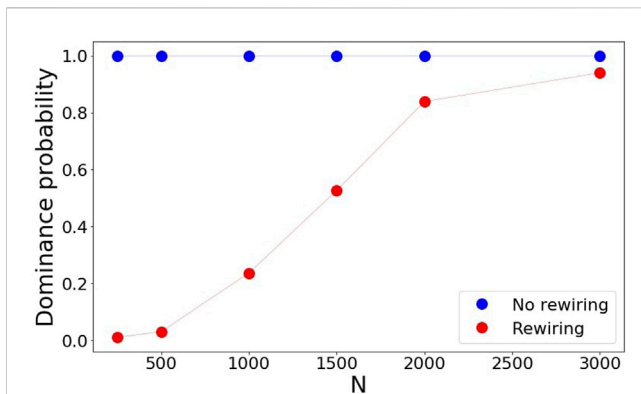
FIGURE 7
Effect of rewiring on the dependence of the dominance probability on system size. Blue curve: rewiring is not allowed. As the system size increases the probability of dominance is constant and always equal to 1. Red curve: rewiring is allowed. As the system size increases it becomes more and more likely that one language dominates in the long run. Parameters: $q = 0.5$, $t_{final} = 10^5$ and 100 realizations.

is significantly higher than for the $q = 0$ case. For all other values of q, we find that in the absence of rewiring the network is characterized by the dominance of one of the two languages, whereas when we allow for rewiring the network is polarized. Furthermore, increasing q causes a decrease in the difference between the majority and minority group sizes. We find that the ensemble average preference for all groups of agents (the bilinguals, the majority language, and the minority language groups) fluctuates around 0.5 as expected from the symmetry of our model with regard to both languages. We understand this dependence on q as follows. When q increases the link probability also increases but in turn this leads to more rewiring, which in the end favors coexistence. Then, we expect that in more dynamic societies (e.g., urban regions) the linguistic scenario will be more heterogeneous with the presence of both languages unlike static societies (e.g., rural regions), where one of the languages is more

likely to disappear. This is consistent with the usage of linguistic varieties recently observed in social media (Gonçalves and Sánchez, 2014). Finally, we should note that initial conditions with lower percentage of bilingual agents were also examined. For example, we considered the case of starting with an already segregated society in terms of links' state (which implies a low number of bilinguals), but maintaining the uniformly random distribution of preferences and the results remained largely unchanged. We also considered other scenarios of intermediate number of bilinguals and we verified the robustness of our results.

3.4 Dependence on system size

We now calculate the probability of the simulation reaching dominance as a function of the system size. Our results are depicted in Figure 7. We find that as the system size increases it is highly likely that the final state of the network reaches language dominance (red dots). To compare with, we also show the results when no rewiring is allowed (blue dots). In this case, the dominance probability is independent of the system size. Therefore, rewiring is key in determining the fate of the network. In particular, for small systems coexistence is the most likely scenario. For large systems, however, rewiring is less effective for the maintenance of a language.

In Figure 8, we study the dependence of the final state on the system size N. Our simulations show that as the system size increases the proportions of the majority and minority communities approach each other, while the magnitude of the interface constitutes roughly the same fraction of the entire network, independently of the system size. This is reflected in both individuals (see Figure 8A) and their connections (see Figure 8B). We remark that the most likely situation for N large is the extinction of one of the languages. Yet, in the scenario that both languages manage to coexist in a large region, the communities speaking these languages would approximately have the same sizes. Furthermore, the magnitude
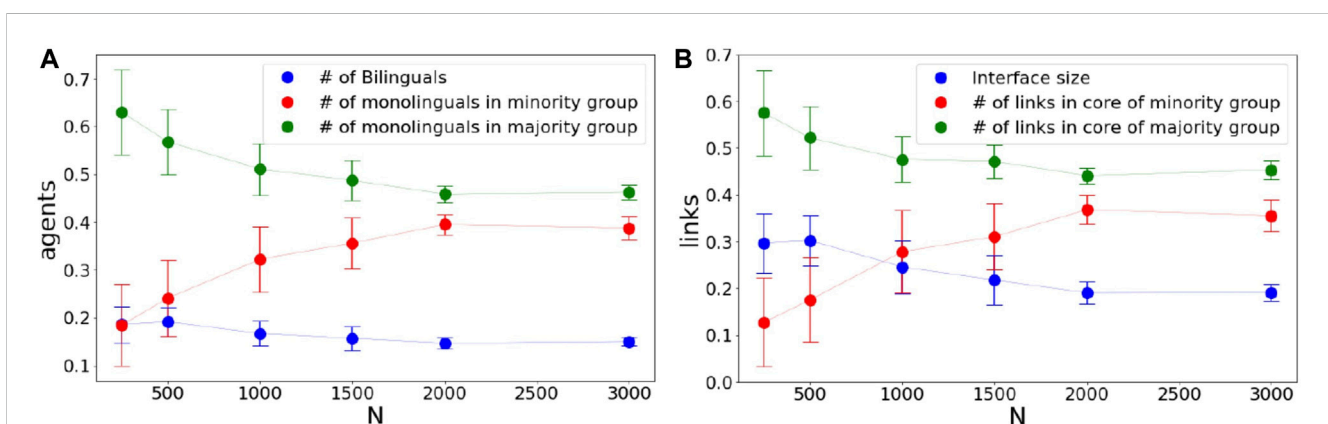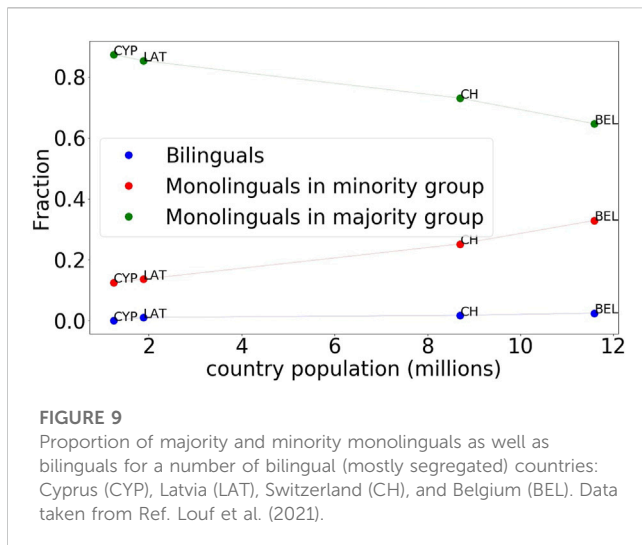

FIGURE 8
Effect of system size on the proportions of bilingual, minority and majority speakers when rewiring is permitted. **(A)** fraction of agents belonging to each group as a function of the system size. **(B)** fraction of links belonging to each group as a function of the system size. As we can see in both figures, as the system size increases the minority and majority groups become almost equal in size. The displayed error bars correspond to one standard deviation. Parameters: $q = 0.5$, $t_{final} = 10^5$.

FIGURE 9
Proportion of majority and minority monolinguals as well as bilinguals for a number of bilingual (mostly segregated) countries: Cyprus (CYP), Latvia (LAT), Switzerland (CH), and Belgium (BEL). Data taken from Ref. Louf et al. (2021).

of the bilingual group of the region would remain approximately constant. This is a remarkable finding that deserves further attention.

To examine whether this is indeed observed in real life, we consider a number of bilingual countries and regions showing a clear segregation and plot the fraction of speakers of both the majority and minority languages as well as the percentage of bilinguals. We obtain these data from geolocalized posts in Twitter with automatic language detection (Louf et al., 2021). We find that for a number of countries studied there is a qualitative agreement between the data and our findings, as can be seen in Figure 9. In particular, we observe that from the populations of the bilingual countries of Cyprus, Latvia, Switzerland (ignoring the Italian-speaking population) and Belgium, the sizes of the majority and minority groups approach each other for bigger countries. Moreover, we find that the percentage of bilingual individuals remains constant, independently of the country size, also in agreement with our numerical predictions discussed above.

4 Conclusion

We have studied an agent-based model for language dynamics valid for bilingual societies, enabling rewiring in the link dynamics. This models the agents' decision to continue speaking their preferred language while engaging with a different agent. As a consequence of this, we find that both languages can coexist in the long-term state of the network, resulting in a bipolarized network. As the system size increases, the probability that a single language dominates also increases. Surprisingly enough, for those simulations that show language coexistence, we find that the relative size of the bilingual agents remains constant with the system size, while the difference between the majority and the minority language groups decreases. We compare this with population data and language usage in bilingual countries and obtain good agreement. Another key quantity is the likelihood of individuals updating their language preference based on the languages spoken by their neighbors. Increasing this likelihood also strengthens the probability of reaching language coexistence, thereby reducing language shift.

Clearly, our work implies that understanding the dynamics of social network interactions has a profound significance for language planning (Kaplan and Baldauf, 1997) and the design of adequate policies that promote the revitalization of endangered languages (Sallabank, 2013). Our model assumes that language is a property of the interaction rather than a property of the speaker and that bilingualism (or multilingualism) is a result of language use. Of course, individuals have a language preference and this is modeled with a continuous parameter. But, importantly, agents can also adapt the topology of their interactions to align with their preferences.

Further extensions of the model should take into account not only linguistically segregated regions but also mixed societies, especially when the interlinguistic similarity is high (Mira et al., 2011). Another important issue is demographics. In this case, networks of varying sizes should be treated dynamically. It would be interesting to apply the formalism of multilayer networks (Domenico, 2023) to account for heterogeneous speech communities and how layer interdependence affects language coexistence, possibly incorporating social factors such as economic class (Louf et al., 2023) and status (Rosillo-Rodes et al., 2023). Furthermore, one could also consider the role of memory in the dynamics of both the nodes and the links as in the work of Scialla et al. (2023). In addition, one could consider the case where the dynamics, e.g., of the language usage decision, is defined taking into account group interactions rather than simply pairwise interactions (Perc et al., 2013). Finally, we could also consider probabilistic updates for the link choice as well, i.e., the introduction of noise. Currently, for language shift to take place it suffices that the majority pressure for the other language is larger than the link preference for the currently used language, but we could assume a probabilistic change that would depend on the difference between the two quantities. This would describe the scenario of a not completely rational choice of the agents when taking into account the two sociological mechanisms, the majority pressure and the link preference, considered in this work.

Data availability statement

Publicly available datasets were analyzed in this study. This data can be found here: https://figshare.com/articles/dataset/Spatial_distributions_of_languages_on_Twitter/14339321.

Author contributions

CC: Conceptualization, Data curation, Formal Analysis, Investigation, Methodology, Software, Validation, Visualization, Writing–original draft, Writing–review and editing. DS: Conceptualization, Funding acquisition, Investigation, Methodology, Project administration, Writing–review and editing. RT: Conceptualization, Funding acquisition, Investigation, Methodology, Project administration, Writing–review and editing.

Funding

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

Abrams, D. M., and Strogatz, S. H. (2003). Modelling the dynamics of language death. *Nature* 424, 900. doi:10.1038/424900a

Abutalebi, J., and Green, D. (2007). Bilingual language production: the neurocognition of language representation and control. *J. Neurolinguistics* 20, 242–275. doi:10.1016/j.jneuroling.2006.10.003

Antal, T., Krapivsky, P. L., and Redner, S. (2005). Dynamics of social balance on networks. *Phys. Rev. E* 72, 036121. doi:10.1103/PhysRevE.72.036121

Appel, R., and Muysken, P. (2006). *Language contact and bilingualism*. Cambridge, UK: Cambridge University Press.

Carro, A., Toral, R., and Miguel, M. S. (2016). Coupled dynamics of node and link states in complex networks: a model for language competition. *New J. Phys.* 18, 113056. doi:10.1088/1367-2630/18/11/113056

Carro, A., Vazquez, F., Toral, R., and San Miguel, M. (2014). Fragmentation transition in a coevolving network with link-state dynamics. *Phys. Rev. E* 89, 062802. doi:10.1103/PhysRevE.89.062802

Castelló, X., Eguíluz, V. M., and Miguel, M. S. (2006). Ordering dynamics with two non-excluding options: bilingualism in language competition. *New J. Phys.* 8, 308. doi:10.1088/1367-2630/8/12/308

Castello, X., Loureiro-Porto, L., and San Miguel, M. (2013). Agent-based models of language competition. *Int. J. Sociol. Lang.* 2013, 21–51. doi:10.1515/ijsl-2013-0022

Chartrand, G., and Stewart, M. J. (1969). The connectivity of line-graphs. *Math. Ann.* 182, 170–174. doi:10.1007/bf01350320

Colomer-de-Simón, P., Serrano, M. A., Beiró, M. G., Alvarez-Hamelin, J. I., and Boguñá, M. (2013). Deciphering the global organization of clustering in real complex networks. *Sci. Rep.* 3, 2517. doi:10.1038/srep02517

Domenico, M. D. (2023). More is different in real-world multilayer networks. *Nat. Phys.* 1–16 . doi:10.1038/s41567-023-02132-1

Dorogovtsev, S., and Mendes, J. F. F. (2003). *Evolution of networks: from biological nets to the Internet and WWW*. Oxford, UK: Oxford Academic.

Edelman, S., and Waterfall, H. (2007). Behavioral and computational aspects of language and its acquisition. *Phys. Life Rev.* 4, 253–277. doi:10.1016/j.plrev.2007.10.001

Fernández-Gracia, J., Castello, X., Eguiluz, V. M., and San Miguel, M. (2012). Dynamics of link states in complex networks: the case of a majority rule. *Phys. Rev. E* 86, 066113. doi:10.1103/PhysRevE.86.066113

Foster, D. V., Foster, J. G., Grassberger, P., and Paczuski, M. (2011). Clustering drives assortativity and community structure in ensembles of networks. *Phys. Rev. E* 84, 066117. doi:10.1103/PhysRevE.84.066117

Gonçalves, B., and Sánchez, D. (2014). Crowdsourcing dialect characterization through Twitter. *PLOS ONE* 9, e112074. doi:10.1371/journal.pone.0112074

Jackson, G. M., Swainson, R., Cunnington, R., and Jackson, S. R. (2001). Erp correlates of executive control during repeated language switching. *Biling. Lang. Cognition* 4, 169–178. doi:10.1017/S1366728901000268

Kandler, A. (2009). Demography and language competition. *Hum. Biol.* 81, 181–210. doi:10.3378/027.081.0305

Kaplan, R. B., and Baldauf, R. B. (1997). *Language planning from practice to theory*. Bristol, UK: Multilingual Matters.

Klimek, P., Diakonova, M., Eguiluz, V. M., San Miguel, M., and Thurner, S. (2016). Dynamical origins of the community structure of an online multi-layer society. *New J. Phys.* 18, 083045. doi:10.1088/1367-2630/18/8/083045

Klimek, P., and Thurner, S. (2013). Triadic closure dynamics drives scaling laws in social multiplex networks. *New J. Phys.* 15, 063008. doi:10.1088/1367-2630/15/6/063008

Krauss, M. (1992). The world's languages in crisis. *Language* 68, 4–10. doi:10.1353/lan.1992.0075

Krawczyk, M. J., Muchnik, L., Mańka-Krasoń, A., and Kułakowski, K. (2011). Line graphs as social networks. *Phys. A Stat. Mech. its Appl.* 390, 2611–2618. doi:10.1016/j.physa.2011.03.009

Louf, T., Ramasco, J. J., Sánchez, D., and Karsai, M. (2023). When dialects collide: how socioeconomic mixing affects language use. Available at: https://arxiv.org/pdf/2307.10016.pdf.

Louf, T., Sánchez, D., and Ramasco, J. (2021). Capturing the diversity of multilingual societies. *Phys. Rev. Res.* 3, 043146. doi:10.1103/PhysRevResearch.3.043146

Mańka-Krasoń, A., Mwijage, A., and Kułakowski, K. (2010). Clustering in random line graphs. *Comput. Phys. Commun.* 181, 118–121. doi:10.1016/j.cpc.2009.09.010

Minett, J. W., and Wang, W. S.-Y. (2008). Modelling endangered languages: the effects of bilingualism and social structure. *Lingua* 118, 19–45. doi:10.1016/j.lingua.2007.04.001

Mira, J., and Paredes, Á. (2005). Interlinguistic similarity and language death dynamics. *Europhys. Lett.* 69, 1031–1034. doi:10.1209/epl/i2004-10438-4

Mira, J., Seoane, L. F., and Nieto, J. J. (2011). The importance of interlinguistic similarity and stable bilingualism when two languages compete. *New J. Phys.* 13, 033007. doi:10.1088/1367-2630/13/3/033007

Nepusz, T., and Vicsek, T. (2012). Controlling edge dynamics in complex networks. *Nat. Phys.* 8, 568–573. doi:10.1038/nphys2327

Newman, M. E. J. (2010). *Networks: an introduction*. Oxford, UK: Oxford University Press.

Newman, M. E. J., and Park, J. (2003). Why social networks are different from other types of networks. *Phys. Rev. E* 68, 036122. doi:10.1103/PhysRevE.68.036122

Patriarca, M., Castello, X., Uriarte, J. R., Eguiluz, V. M., and San Miguel, M. (2012). Modelling two-language competition dynamics. *Adv. Complex Syst.* 15, 1250048. doi:10.1142/S0219525912500488

Patriarca, M., and Heinsalu, E. (2009). Influence of geography on language competition. *Phys. A Stat. Mech. its Appl.* 388, 174–186. doi:10.1016/j.physa.2008.09.034

Perc, M., Gómez-Gardeñes, J., Szolnoki, A., Floría, L. M., and Moreno, Y. (2013). Evolutionary dynamics of group interactions on structured populations: a review. *J. R. Soc. Interface* 10 (80), 20120997. doi:10.1098/rsif.2012.0997

Prochazka, K., and Vogl, G. (2017). Quantifying the driving factors for language shift in a bilingual region. *Proc. Natl. Acad. Sci.* 114, 4365–4369. doi:10.1073/pnas.1617252114

Reardon, S. F., and Firebaugh, G. (2002). 2. Measures of multigroup segregation. *Sociol. Methodol.* 32, 33–67. doi:10.1111/1467-9531.00110

Rooij, A. C. M., and Wilf, H. S. (1965). The interchange graph of a finite graph. *Acta Math. Acad. Sci. Hung.* 16, 263–269. doi:10.1007/bf01904834

Rosillo-Rodes, P., San Miguel, M., and Sánchez, D. (2023). Modelling language ideologies for the dynamics of languages in contact. *Chaos*, 33 (11) doi:10.1063/5.0166636

Saeedian, M., San Miguel, M., and Toral, R. (2019). Absorbing phase transition in the coupled dynamics of node and link states in random networks. *Sci. Rep.* 9, 9726. doi:10.1038/s41598-019-45937-y

Saeedian, M., San Miguel, M., and Toral, R. (2020). Absorbing-state transition in a coevolution model with node and link states in an adaptive network: network fragmentation transition at criticality. *New J. Phys.* 22, 113001. doi:10.1088/1367-2630/abbfd0

Sallabank, J. (2013). *Attitudes to endangered languages: identities and policies*. Cambridge, UK: Cambridge University Press.

Scialla, S., Liivand, J.-K., Patriarca, M., and Heinsalu, E. (2023). A three-state language competition model including language learning and attrition. *Front. Complex Syst.* 1, 1266733. doi:10.3389/fcpxs.2023.1266733

Serrour, B., Arenas, A., and Gomez, S. (2011). Detecting communities of triangles in complex networks using spectral optimization. *Comput. Commun.* 34, 629–634. doi:10.1016/j.comcom.2010.05.006

Sole, R. V., Corominas-Murtra, B., and Fortuny, J. (2010). Diversity, competition, extinction: the ecophysics of language change. *J. R. Soc. Interface*. 7, 1647–1664. doi:10.1098/rsif.2010.0110

Steels, L. (2011). Modeling the cultural evolution of language. *Phys. Life Rev.* 8, 339–356. doi:10.1016/j.plrev.2011.10.014

Suchecki, K., Eguíluz, V. M., and Miguel, M. S. (2005). Conservation laws for the voter model in complex networks. *Europhys. Lett.* 69, 228–234. doi:10.1209/epl/i2004-10329-8

Szell, M., Lambiotte, R., and Thurner, S. (2010). Multirelational organization of large-scale social networks in an online world. *Proc. Natl. Acad. Sci.* 107, 13636–13641. doi:10.1073/pnas.1004008107

Szell, M., and Thurner, S. (2010). Measuring social dynamics in a massive multiplayer online game. *Soc. Netw.* 32, 313–329. doi:10.1016/j.socnet.2010.06.001

Szell, M., and Thurner, S. (2012). Social dynamics in a large-scale online game. *Adv. Complex Syst.* 15, 1250064. doi:10.1142/s0219525912500646

Traag, V. A., and Bruggeman, J. (2009). Community detection in networks with positive and negative links. *Phys. Rev. E* 80, 036115. doi:10.1103/PhysRevE.80.036115

Wang, W. S. Y., and Minett, J. W. (2005). The invasion of language: emergence, change and death. *Trends Ecol. Evol.* 20, 263–269. doi:10.1016/j.tree.2005.03.001

Williams, R. A. (2018). Lessons learned on development and application of agent-based models of complex dynamical systems. *Simul. Model. Pract. Theory* 83, 201–212. doi:10.1016/j.simpat.2017.11.001