Regular Article

Are Dutch and French languages miscible?

Luís F. Seoane[1,2], **Jorge Mira**[3,a]

[1] Systems Biology Department, Centro Nacional de Biotecnología, Darwin 3, Madrid 28049, Spain
[2] Grupo Interdisciplinar de Sistemas Complejos (GISC), Madrid, Spain
[3] Departamento de Física Aplicada and iMATUS, Universidade de Santiago de Compostela, Santiago de Compostela 15782, Spain

Abstract French and Dutch are two languages of different origins (Germanic vs. Romance) that coexist within the nation-state of Belgium. While they are mostly segregated throughout the Belgian territory, in Brussels they reach an actual cohabitation with a relevant bilingual population. The dominant language in Brussels shifted from Dutch to French during the late XIX century in a process known as the Francization of Brussels. The fractions of speakers of each language and of bilinguals over that time were recorded periodically until political tensions ended the censuses in the country. This relevant linguistic shift has been the object of sociopolitical studies, but the available empirical data have never before been analyzed using a theoretical mathematical model that would allow us to quantify causal factors behind the observed dynamics. Here we carry out such study for the first time, measuring effective values of perceived interlinguistic similarity and language prestige, among others. This modeling and quantification allows us to speculate about possible trajectories of fractions of speakers over time—specifically, whether Dutch and French tend to be languages that can coexist in the long term. We find that there is an overall tendency of both tongues to grow segregated over time, suggesting, in physics terms, that Dutch and French are *not miscible*. The scenarios that would allow for language coexistence would often see a starkly dominating language. Notwithstanding, we also discuss the costs of attempting to sustain the cohabitation despite a natural tendency to the contrary.

1 Introduction

Languages are one of the human traits that better identify a society. This is clearly seen in some countries with more than one official tongue, where national identities are often linked to linguistic options. Belgium is a remarkable example of this. It is difficult to find a western-world country where a linguistic divide has shaped politics so intensely. Two main tongues dominate Belgium historically: Dutch, in the northern Flanders region; and French, in Wallonia, the southern region. An internal political and administrative boundary was established in 1932 delimiting French- and Dutch-speaking territories [1], and setting Brussels aside as a mixed region [2]. (A small German-speaking area also exists on the east side of the country.) This boundary, in practice, further consolidated the linguistic divide. The boundary was originally conceived as dynamic: It would be updated according to a linguistic census taken every 15 years. But in 1961 several Flemish municipalities refused to abide. The previous boundary (established in 1947) was adopted as definitive in 1962 by the Gilson's Act [3]. That also brought to a halt the periodic censuses and associated sociolinguistic studies – at least those accepted by both the Flemish and Walloon communities.

The identity divide can still be observed nowadays in the nomenclature of villages, in signaling across the country, or in geolocalized posts on social media [4, 5] (Fig. 1a). The Brussels-capital region (in French: Région de Bruxelles-Capitale; in Dutch: Brussels Hoofdstedelijk Gewest) stands out because both Dutch and French coexist more closely there – making it the only officially bilingual French–Dutch region in the country. This puts Brussels forward as a relevant case study for sociolinguistics. Its role as capital city of Belgium (located at Ville de Bruxelles / Stad Brussel) and main center of the European Union further highlights the importance of understanding the sociolinguistic dynamics in this area.

The clear linguistic boundary seen in the Belgian map contrasts starkly with other cases in which several languages coexist in a same country. Take for example the cohabitation of Catalan and Spanish in Catalonia [6]. Despite the acute political conflict (which also involves language and identity), mapping the usage of both tongues across the region reveals a much better mixture (Fig. 1b). Looking at the Catalan and Belgian cases through a physics lens, we might think, respectively, of *well-mixed* (as in dissolved salt) versus *segregated* phases (as in oil and water). How far away can we take this simile? Might mathematical tools help us model the dynamics of language shift in Belgium and thus gain some insight on whether both languages are *miscible*? Note that Dutch and French are more dissimilar than Catalan and Spanish: May interlinguistic similarity play a role in determining such miscibility?

---

[a] e-mail: jorge.mira@usc.es (corresponding author)

Fig. 1 Miscibility of languages captured by posts in social media. **a** This map of Belgium shows in blue Twitter posts written in Dutch and in yellow those in French. It is clear the separation of both languages (excepting the area of Brussels), following the border Flanders-Wallonia. **b** Map of Catalonia, where there are also two main cohabitating languages: Catalan and Castillian Spanish. But, in this case, there is a mixing of both (yellow: posts in Catalan; blue: posts in Spanish). Figure modified from [4]
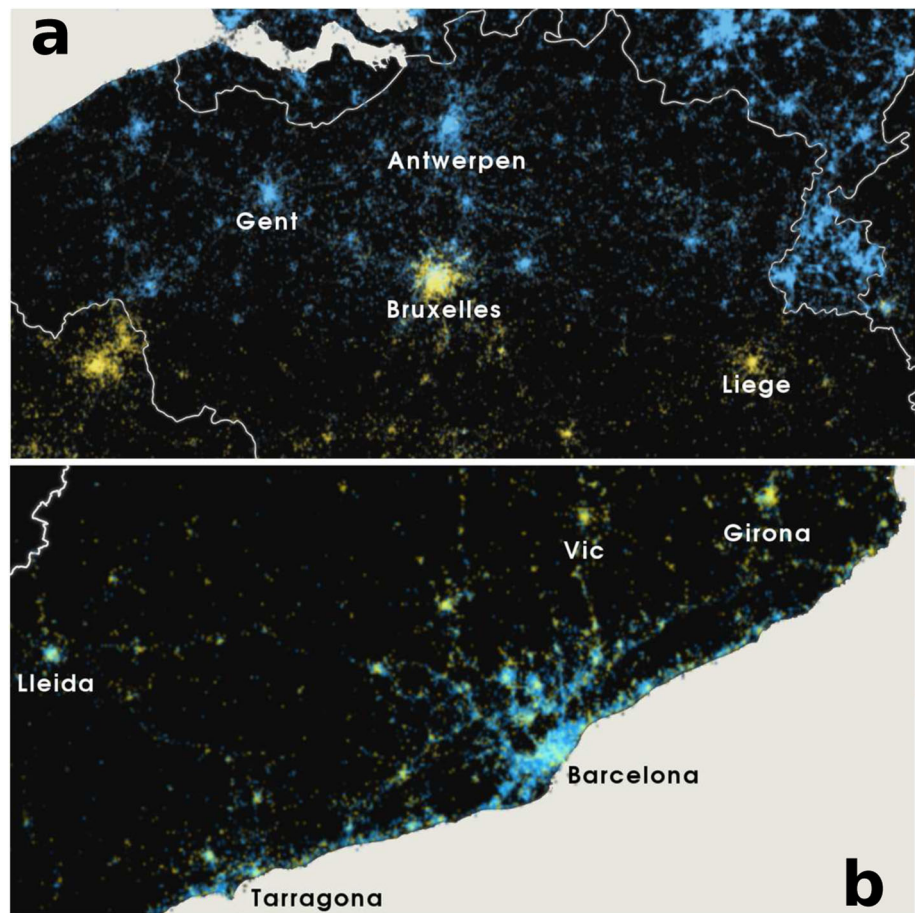
In this paper we use a system of ordinary differential equations to study historical data of speakers in different neighborhoods of the Brussels-capital region. By fitting model parameters to the observed sociolinguistic trajectories, we assess the stability of the system of cohabitating languages – i.e., whether both tongues might coexist in the long run, or whether one of them shall go extinct given the model and inferred parameters. We study a historical period that led to the Francization of Brussels [7], in which French displaced Dutch as the most spoken language in this area following a shift in the nationwide socioeconomic balance between southern and northern regions. The available data spans from late-XIX to mid-XX century [8]. Later socioeconomic processes (e.g., immigration, the enactment of the internal linguistic boundary, development of regulation to protect either tongue, etc.) have affected the Belgian linguistic scenario in ways that, very likely, supersede the dynamics expected from our fitted model alone. Due to the discontinuation of linguistic censuses, it does not exist comparable data to adapt our model to the newest scenario. However, our aim here is not to forecast how the number of speakers of each tongue might evolve over time, but rather to study the interaction between both languages during a limited period during which we have homogeneously-gathered data. We hope that insights about possible coexistence between both tongues can seed some light on the efforts necessary to keep the linguistic balance in the long term. From a more far-fetched perspective, given the political implications of identity issues, we hope that our results may inform the stability of Belgium as a political entity in a broader sense.

2 Methods

2.1 Data

In the Brussels-capital region, speakers of one language are exposed to the other in an effective way. This coexistence fosters sociolinguistic dynamics by which speakers switch or maintain tongues over time, which we attempt to study here. Historical data of such dynamics in the Brussels-capital region are publicly available for a period spanning from the second half of the XIX century to 1947 [8]. During this period, the so-called Francization of Brussels started, marking a decline of Dutch in favor of French in most neighborhoods of the Brussels-capital region.

Table 1 Numerical outputs of the fits, after processing the sets of values of volatility (a), dynamical rate (c), interlinguistic similarity (k) and Dutch prestige (s) for the 19 municipalities of the Brussels-capital region, as well as the 3 that were merged in the Brussels main municipality in 1921. Values $\hat{a}$, $\hat{c}$, $\hat{k}$ and $\hat{s}$ correspond to the best fit for each case (i.e., the fit with the lowest average fit error, χ^2). $\langle a \rangle$, $\langle c \rangle$, $\langle k \rangle$, $\langle s \rangle$ are the averages weighted across all the valid fits

Region	$\hat{a}$	$\hat{c}$	$\hat{k}$	$\hat{s}$	$\langle a \rangle$	$\langle c \rangle$	$\langle k \rangle$	$\langle s \rangle$
1 - Anderlecht	1.21	0.39	0.68	0.51	1.50	0.52	0.68	0.47
2 - Auderghem/Oudergem	1.44	0.37	0.67	0.42	1.56	0.66	0.69	0.42
3 - Berchem-Sainte-Agathe/Sint-Agatha-Berchem	1.28	0.75	0.68	0.49	1.45	0.55	0.67	0.40
4 - Bruxelles-ville / Stad Brussel	1.92	0.46	0.69	0.47	1.75	0.50	0.69	0.45
5 - Etterbeek	1.11	0.30	0.67	0.39	1.52	0.64	0.64	0.41
6 - Evere	1.46	0.13	0.74	0.23	1.49	0.43	0.66	0.40
7 - Forest/Vorst	1.36	0.44	0.66	0.43	1.56	0.58	0.66	0.41
8 - Ganshoren	1.25	0.42	0.59	0.45	1.41	0.58	0.62	0.41
9 - Haren	0.99	1.71	0.57	0.67	1.53	0.90	0.55	0.44
10 - Ixelles/Elsene	1.33	0.09	0.65	0.33	1.61	0.38	0.60	0.44
11 - Jette	1.40	0.31	0.65	0.47	1.54	0.46	0.67	0.43
12 - Koekelberg	1.08	0.07	0.75	0.14	1.50	0.55	0.70	0.44
13 - Laeken/Laken	1.30	0.27	0.69	0.52	1.60	0.55	0.69	0.48
14 - Molenbeek-Saint-Jean/Sint-Jans-Molenbeek	1.37	0.34	0.70	0.51	1.56	0.50	0.69	0.47
15 - Neder-Over-Heembeek	1.60	0.25	0.91	0.19	1.54	0.87	0.65	0.49
16 - Saint-Gilles/Sint-Gillis	1.12	0.26	0.64	0.39	1.55	0.47	0.67	0.42
17 - Saint-Josse-ten-Noode/Sint-Joost-ten-Node	2.15	0.68	0.72	0.48	1.83	0.52	0.70	0.46
18 - Schaerbeek/Schaarbeek	1.35	0.14	0.75	0.38	1.69	0.46	0.71	0.44
19 - Uccle/Ukkel	1.14	0.10	0.74	0.14	1.47	0.54	0.66	0.40
20 - Watermael-Boitsfort/Watermaal-Bosvoorde	1.11	0.30	0.66	0.42	1.54	0.55	0.66	0.40
21 - Woluwe-Saint-Lambert/Sint-Lambrechts-Woluwe	1.31	0.59	0.62	0.43	1.36	0.48	0.59	0.36
22 - Woluwe-Saint-Pierre/Sint-Pieters-Woluwe	1.21	1.01	0.58	0.46	1.27	0.54	0.61	0.39
23 - Brussels-capital region	1.58	0.31	0.62	0.46	1.65	0.48	0.64	0.44

The available censuses show the number of speakers of Dutch, French, and German over time in the 22 municipalities in which the Brussels-capital region was divided during the studied period. These neighborhoods are listed in Table 1. In the time series, we dismissed exclusive German speakers (which were always a minority in every municipality) and normalized the data to obtain fractions of monolingual Dutch and French as well as bilingual speakers. These time series are not homogeneous across municipalities: not all of them have data for the same number of years, as three regions (Haren, Laeken/Laken and Neder-Over-Heembeek) were merged in the Bruxelles-ville / Stad Brussel neighborhood in 1921 and there are no separated data available for them from that year on. We generated an additional time series for the merged Brussels-capital region by aggregating the speakers of each group across all municipalities for each year. We provide the resulting time series in an accessible format. Figure 2 shows this empirical data alongside fits to the model (see next section).

2.2 Model and language coexistence

The central issue in this article is to check the stability of the Dutch–French system in a region where both are simultaneously present. In order to study similar cases, in previous work we had modified a seminal model of language competition by Abrams
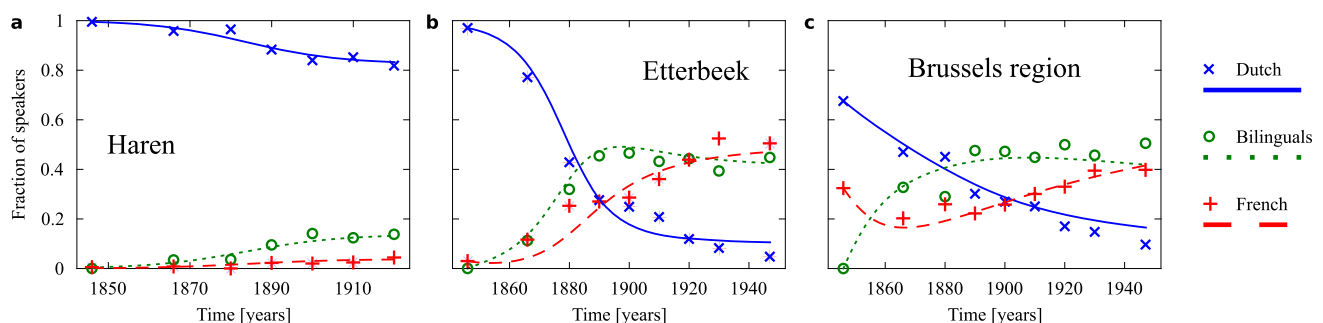
Fig. 2 Examples of fits of data to the model. The correspondence of lines (fits) and symbols (real data) is indicated at the right. **a** Representation of the best fit among all regions, which corresponds to the municipality of Haren. **b** Representation of the worst fit, which corresponds to the municipality of Etterbeek. **c** Fit to the aggregated data for the whole region of Brussels

and Strogatz [9] to include scenarios that allow bilingualism [10, 11]. In the past, we had applied this framework to real data from the Spanish Autonomous Communities of Galicia [10–12] (where Galician and Castillian Spanish are coofical languages) and Catalonia [6] (where Catalan and Castillian Spanish are also cooficial). Furthermore, in previous work we have worked out the stability conditions of the model, both from the computational [10, 11] and analytical points of view [13–15]. From a mathematical and computational perspectives, in the present paper we do not alter the model in any way, nor do we introduce any new terminology or phenomenology—i.e., we use the existing model of language coexistence with bilingualism as originally developed in [10] and later analyzed in [11, 13–15] and as used on other couples of competing languages in [6, 10–12].

The model consists of a system of coupled differential equations that contemplates two monolingual groups (X for Dutch and Y for French in this paper) and a bilingual one, B. These groups make up fractions x, y, and b of speakers, respectively, within a normalized population ($x + b + y = 1$). The model assumes that the probability that monolingual speakers acquire the opposite language is proportional to a prestige (s_X or s_Y) associated with the target tongue. We take $s_X, s_Y \in [0, 1]$ and $s_X + s_Y = 1$; so we can focus on $s \equiv s_X$ (i.e., Dutch's prestige).

We understand this prestige as in [9] – i.e. a parameter that reflects, in average, the social or economic opportunities that speakers attribute to the use of each language. More importantly, we take this (and other) parameters as effective properties that emerge from the data, thus constraining effective models that describe each sociolinguistic situation. In other words, the unfolding of the sociolinguistic dynamics is a way to measure the prestige of these tongues and other model parameters.

Of all speakers acquiring a new tongue, a fraction k of them retains the old one (hence becoming bilinguals) while $1 - k$ of them switch and forget. Since k captures how easy it is to retain both languages, it has been interpreted as a measure of similarity between both tongues (as perceived by their users) [10, 16] and has been termed *interlinguistic similarity*.

From these assumptions, the probabilities of leaving or entering each group (X, Y, or B) result in a set of differential equations that tells us how the fractions (x, y, and b) of speakers with each linguistic choice evolve over time:

$$
\begin{aligned}
\frac{dx}{dt} &= c\Big[(b + y)(1 - k)s(1 - y)^a \\
&\quad - x\big((1 - k)(1 - s)(1 - x)^a + k(1 - s)(1 - x)^a\big)\Big], \\
\frac{dy}{dt} &= c\Big[(b + x)(1 - k)(1 - s)(1 - x)^a \\
&\quad - y\big((1 - k)s(1 - y)^a + ks(1 - y)^a\big)\Big].
\end{aligned}
\tag{1}
$$

Since $b = 1 - x - y$, a third differential equation is not needed.

The interpretation of the different parameters has been discussed in the literature, more profusely so in the review [16], which also recollects the history of mathematical models that study the dynamics of linguistic choices. Let us summarize that history briefly. The earliest models of language choice dynamics date back to the early 90's with the work of Baggs and Freedman [17, 18]. These authors studied the stability of a quite general set of language choice dynamics, but unfortunately their work remained mostly ignored or undiscovered. A turning point in the mathematical modeling of sociolinguistic dynamics was the paper by Abrams and Strogatz [9], who proposed a very simple and concrete set of equations (in contrast to the more general and abstract ones by Baggs and Freedman) that allowed them to fit the model to empirical datasets with the evolution of language choices over time. Importantly for us, these languages were distant enough such that bilingualism could be safely ignored in their analysis—the model itself did not include the bilingual options.

Many alternative models have been developed over the years to study sociolinguistic contact [17–31], most of them inspired by the original work by Abrams and Strogatz. These works include agent-based approaches, mean-field equations drawn from ecological dynamics, or diffusion equations in which the physical distribution of speakers matters. Each model can be suited to different situations and data sets (e.g., reaction–diffusion-like models can be insightful if data is abundantly distributed over space, but sampling is scarce in time [30]). Most of the models brought about by the wave initiated by Abrams and Strogatz do not include bilingualism as an option (with a few notable exceptions). As we stated above, our model has been inspired by the work of Abrams and Strogatz as well. Several reviews discuss and interpret the existing models and can help navigate the possibilities [16, 32–34]. Our data suggest using a mean-field model, of which many in the literature account for bilingualism as well [17, 18, 27, 28]. Our choice of equations allows for stable coexistence solutions (which several models lack) and its stability has been thoroughly studied.

3 Results

Following the procedure introduced in [6], we fitted the model described by Eq. 1 to each of our time series separately (including the one with aggregated data from all regions). This method is a fast, stochastic least-square minimization that, for each time series, yields a set $\{a, c, k, s\}$ of parameters that informs us about the sociolinguistic dynamics. To circumvent the stochasticity of the process and the possibility that it might reach local optima, we ran the procedure 1000 times for each data set. From the resulting collections of plausible parameters we only retained those with meaningful values: $\{a, c, k, s\} > 0$ and $0 < \{k, s\} < 1$. Outside these
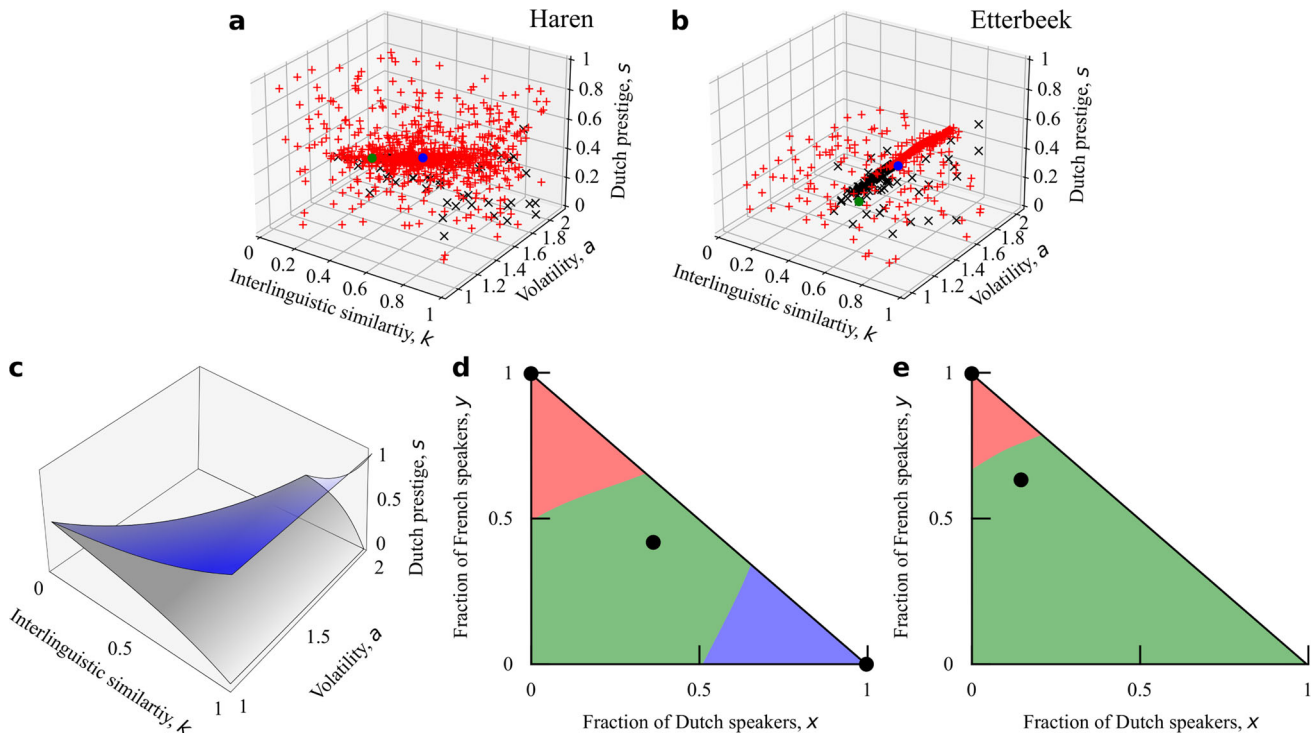
Fig. 3 Empirical data constrains model parameters and possible trajectories in the long term. **a** Representation of the $\Pi \equiv \{\pi_i\} = \{k_i, a_i, s_i\}$ obtained from the valid fits of the model to the data of Haren (best fit). Red pluses indicate (k_i, a_i, s_i) values that lead to the extinction of one of the languages. Black crosses are compatible with the coexistence of both tongues (depending on initial conditions). The point corresponding to the best fit is marked in green. The weighted average of all valid fits is marked in blue. While the data does not constrain interlinguistic similarity or prestige much, it appears to constrain the expected long-term evolution since most fits lead to the extinction of one of the tongues. **b** Same, for the municipality of Etterbeek (worst fit). The ranges of interlinguistic similarity and prestige are more narrow, but the data seem more ambiguous regarding language extinction. **c** The shadowed surfaces (calculated analytically in [14]) define the volume of the $k - a - s$ space for which the model allows long-term coexistence – points outside this region lead to the extinction of one of the languages. Within the volume of coexistence, the eventual fate might depend on the initial conditions. **d-e** Illustration of basins of attraction dependent on initial conditions (as fractions of Dutch monolinguals, x; French monolinguals, y; and bilinguals, $b = 1 - x - y$). Qualitatively reconstructed from [11]. **d** Three attractors (black dots) split up the $x - y$ space depending on whether a unique language would survive (Dutch in the blue basin of attraction, French in the red one), or whether coexistence follows (green). This coexistence attractor observes a symmetric presence of both tongues, with no dominant language. **e** Our model also contains basins of attraction with asymmetric coexistence (green) in which a language (here Dutch) would retain a minority presence with respect to the other

constraints we find unrealistic dynamics (e.g., time running backwards or negative fractions of speakers). The number of discarded fits was always negligible (below 5% for any time series), thus in all cases we retained a number $N_F \lesssim 1000$ of successful fits.

As a result we obtain a collection of sets of parameters for each empirical time series: $\Pi \equiv \{\pi_i, \geq i = 1, \ldots, N_F\}$, such that each $\pi_i \equiv \{a_i, c_i, k_i, s_i\}$ is the result of a single, valid fit. Each π_i renders a time evolution of the fraction of speakers $(x(t; \pi_i), y(t; \pi_i)$; Fig. 2), which we compare to the empirical time series to measure the goodness of each set of parameters:

$$\chi^2(\pi_i) = \frac{1}{N} \sum_{j=1}^{N} \Big[\big(x_j - x(t_j; \pi_i)\big)^2 + \big(y_j - y(t_j; \pi_i)\big)^2 \Big]^{1/2}. \tag{2}$$

N is the number of data in the current series. The x_j and y_j are the empirical values labeled by $j = 1, \ldots, N$. The t_j are the corresponding times (years) at which each empirical value was measured. The $x(t_j; \pi_i)$ and $y(t_j; \pi_i)$ result from evaluating the model with the corresponding parameters π_i at the sampling times. Figure 2 shows the fits resulting for the π_i with lowest $\chi^2(\pi_i)$ for regions Haren (which is the best fitted region), Etterbeek (worst fitted), and aggregated data for all of Brussels.

Our approach takes advantage of the speed and stochasticity of the fitting procedure to sample the *landscape* of model parameters compatible with the observed data. This allows us to visualize how much it is possible to constrain our model dynamics and its parameters given each empirical data series. Figures 3a-b show how the data constrains the parameter landscape for the best (Haren) and worst (Etterbeek) fitted regions (landscapes for other regions are provided in Supporting Material, but they mostly resemble one of these two extreme cases). Of all combinations of parameters resulting from fits to each data series, we observe that these appear more constrained for Etterbeek (most π_i fall within confined values of k and s, but not of a), while they are more disperse for Haren.

In Eq. 1, some combinations of parameters are compatible with the long-term coexistence of both tongues, while others imply the extinction of one language. Note, first, that this does not depend on the parameter c, which is a common factor in Eq. 1 telling us how fast or slow the dynamics unfold. Hence, the parameter landscapes in Figs. 3a-b (which show a, k, and s, but not c) capture all the information that we can extract from the model regarding long-term language coexistence. Computational [11] and analytical [13–15] studies of the model helped delimit a phase space of expected extinction or cohabitation. The latter is only possible within the surfaces outlined in Fig. 3c. Outside this volume, Eq. 1 have only $(x, y) = (1, 0)$ or $(0, 1)$ as attractors, meaning that one of the two languages would eventually go extinct. Within the coexistence volume, these two attractors might be present together with one of the kind $(x, y) = (x^*, y^*)$ with $0 < x^*, y^* < 1$. In these, $x^* + y^* < 1$, thus $b^* > 0$, and bilingualism would always be present alongside monolinguals of either language. Which attractor is reached for π_i within the coexistence volume depends on the initial fractions of speakers (i.e., the initial conditions, Fig. 3d-e).

In the landscapes of Fig. 3a-b, we plot in red all π_i outside the coexistence volume and in black all π_i within it. We see that, while the parameters appear more constrained for Etterbeek, they are compatible with diverging outcomes. On the other hand, while the parameter landscape appears less constrained for Haren, a vast majority of π_i in this region are not compatible with the long-term coexistence of both tongues.

We summarize the many fits for each data series through the fittest set of parameters:

$$\hat{\pi} \equiv \{\hat{a}, \hat{c}, \hat{k}, \hat{s}\} = \arg \min_{\pi_i} \left\{ \chi^2(\pi_i) \right\}. \tag{3}$$

It is also useful to compute an average over valid fits, for which we weight each set of parameters using:

$$\omega_i \equiv \exp(-\chi^2(\pi_i)). \tag{4}$$

Thus:

$$\langle a \rangle = \frac{1}{Z} \sum_{i=1}^{N_F} \omega_i a_i, \quad \text{with} \quad Z \equiv \sum_{i=1}^{N_F} \omega_i. \tag{5}$$

Similar weighted averages are computed for $\langle c \rangle$, $\langle k \rangle$, and $\langle s \rangle$. Both the best and average parameters for each region are reported in Table 1, and average parameters are plotted as a map over regions in Fig. 4a-d. (Similar maps showing standard deviation of each parameter are included in the Supporting Material, along with a map of mean χ^2.) We observe that the weighted average prestige of Dutch, $\langle s \rangle$, is always lower than that of French (Fig. 4d), as well as most prestige in the best fits, $\hat{s}$. Interlinguistic similarity is moderately large. (Average across all regions is 0.656 ± 0.039.) Note that this does not necessarily capture a similarity of grammar, syntax, or vocabulary alone. Rather, it captures the *ease* with which speakers keep both tongues. Strictly linguistic features might contribute, but this parameter can also be dominated by social factors (e.g., the socioeconomic advantage of communicating across different communities).

To summarize the overall trend to extinction or cohabitation of both languages, we introduce a *coexistence index*. This quantifies how often fits for a region are compatible with the long-term cohabitation of both languages. We assign $g_i = +1$ if the model with π_i evolves a region's initial conditions toward a coexistence attractor, $(x, y) = (x^*, y^*)$; and $g_i = -1$ if the evolution tends to an extinction attractor, $(x, y) = (1, 0)$ or $(0, 1)$. From [14] we have an analytic expression for the coexistence volume. Points outside this volume are assigned $g_i = -1$ trivially. We do not have an analytic expression for the boundaries of attractors as in Fig. 3d-e. We evaluated the convergence of these cases numerically (by checking whether they would fall within a radius of 10^{-5} around an extinction attractor in the far future).

We weight and average the g_i to obtain a coexistence index for each region:

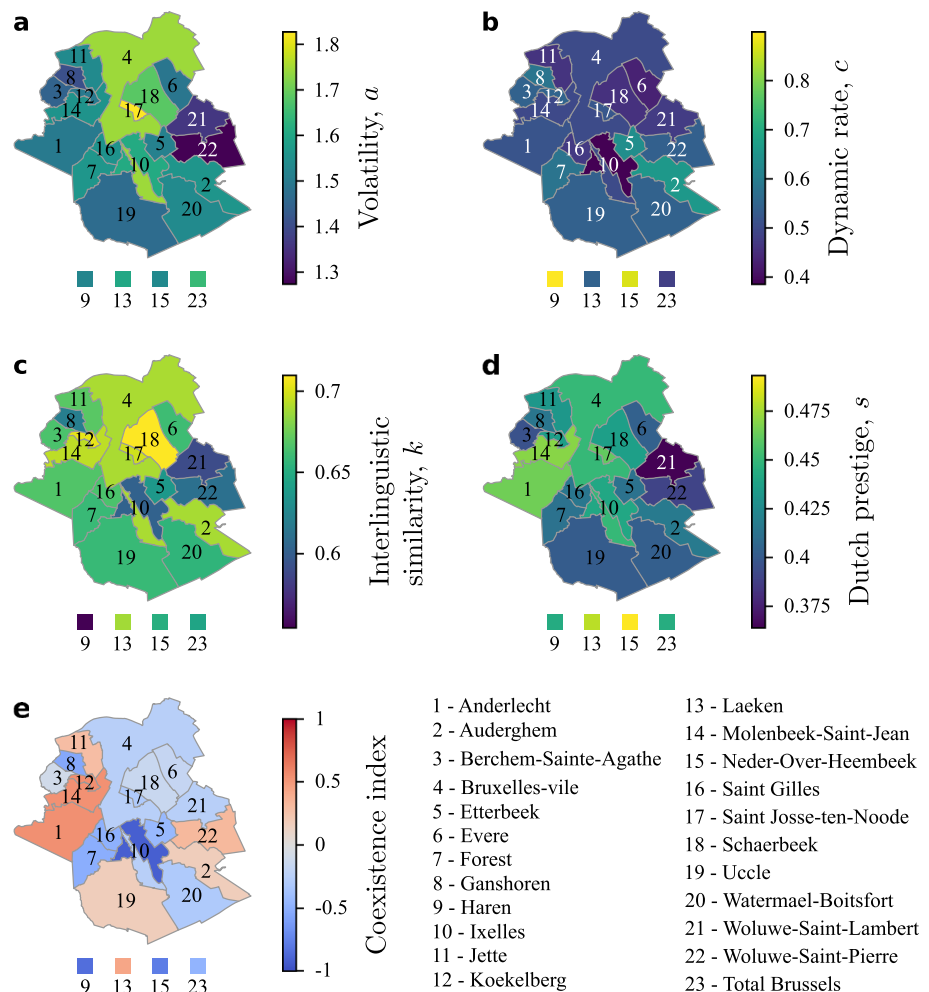$$\langle g \rangle = \frac{1}{Z} \sum_{i=1}^{N_F} \omega_i g_i, \tag{6}$$

as plotted in Fig. 4e. Taking the mean of the $\langle g \rangle$ across all regions, we find $g_{all} = -0.15 \pm 0.44$: There is an overall tendency toward non-coexistence. The large standard deviation (greater than the mean value) results from averaging over regions with clearly distinct tendencies. Taking separately those that tend to a survival of both languages we get $g_{coex} = 0.38 \pm 0.15$. For regions in which the extinction of one tongue would be more often expected we find $g_{ext} = -0.43 \pm 0.24$.

Among coexistence attractors $(x, y) = (x^*, y^*)$, we might find stark asymmetric scenarios in which the presence of one language is greater than the other ($x^* < y^*$ or $y^* < x^*$; Fig. 3e). To account for this effect, we computed a modified coexistence index:

$$\tilde{g}_i = 2 \left[1 - \sqrt{(x^* - y^*)^2} \right] - 1. \tag{7}$$

Where we take $(x^*, y^*) \equiv (0, 1)$ or $(1, 0)$ for extinction attractors. This number is still -1 if coexistence is not possible. If long-term cohabitation is possible, g_i is a graded number between -1 and 1 (the closer to 1, the more symmetric the coexistence). As before, we take, first, the corresponding weighted averages, and then we compute the mean across all regions. We find $\tilde{g}_{all} = -0.34 \pm 0.42$, thus the tendency toward extinction of one of the languages appears strengthened when accounting for asymmetric survival. Separating

Fig. 4 Geographic distribution of average model parameters and coexistence index. Maps show the 19 municipalities with complete time series. Regions 9, 13, and 15 were merged with Bruxelles-ville / Stad Brussel in 1921. Region 23 stands for the aggregated data for all of Brussels. **a** Average volatility, $\langle a \rangle$. **b** Average dynamic rate, $\langle c \rangle$. **c** Average interlinguistic similarity, $\langle k \rangle$. **d** Average Dutch prestige, $\langle s \rangle$. Note that for every municipality it is always lower than that of French. **e** Coexistence index, $\langle g \rangle$, for each municipality

1 - Anderlecht
2 - Auderghem
3 - Berchem-Sainte-Agathe
4 - Bruxelles-vile
5 - Etterbeek
6 - Evere
7 - Forest
8 - Ganshoren
9 - Haren
10 - Ixelles
11 - Jette
12 - Koekelberg
13 - Laeken
14 - Molenbeek-Saint-Jean
15 - Neder-Over-Heembeek
16 - Saint Gilles
17 - Saint Josse-ten-Noode
18 - Schaerbeek
19 - Uccle
20 - Watermael-Boitsfort
21 - Woluwe-Saint-Lambert
22 - Woluwe-Saint-Pierre
23 - Total Brussels

as before: $\tilde{g}_{coex} = 0.34 \pm 0.12$ and $\tilde{g}_{ext} = -0.52 \pm 0.25$. The fact that this index tilts further toward non-coexistence suggests that most cohabitation scenarios would have a very asymmetric configuration in which a language largely dominates the other.

4 Discussion

In this paper we have studied a period that led to the Francization of Brussels [7], in which French penetrated the formerly Dutch-dominated region of Brussels, eventually substituting the vernacular in most scopes. This shift was fostered by socioeconomic factors that included migration and the international prestige of French, which was also the exclusive administrative language at the time. In consistence with this, our analysis shows a higher average prestige for French across all studied regions. The process (that started in the XIX century) might have been slowed down in the XX century due to the enactment of the linguistic boundary. The official status attained by Dutch or a new shift of economic power to the northern (Dutch-speaking) half of the country might also have contributed to slow down the Francization of Brussels. However, some of the factors favoring French (such as its international prestige and wider usage, probably reinforced by the status of Brussels as center of the European Union [35]) extend their influence to the present day.

Our approach is to treat the contact between Dutch and French as a dynamical system limited to a space and time for which we have data series obtained with a consistent methodology, and to observe how these data series constrain a model of sociolinguistic evolution that allows multiple trajectories over time. Some of these trajectories would lead to the coexistence of both languages, while others would imply the extinction of one of the tongues. This would be so if the model would capture all relevant factors that matter in the Dutch–French cohabitation. This last condition is likely not met in the longer run, as effects outside the scope of our study (e.g., large-scale migration as seen in the XX century, further shifts in the socioeconomic balance of power, etc.) come into action. However, we hope that characterizing the dynamics over a limited period of time, with consistently gathered data, we might be able to capture attitudes of speakers toward each-other's tongues. These attitudes would dictate, e.g., perceived prestige or the effort worth making toward sustaining both Dutch and French – which are the main effects captured by our model. Based on

such attitudes, we attempt to infer whether both tongues would likely coexist in the long term. Our insights in the Dutch–French coexistence should be refined in the future if more detailed data, over longer periods of time is gathered, or as complementary models are used to assess, e.g., the influence of the geographic distribution of speakers and how this affects neighboring areas [5, 24].

Our results are tilted toward non-coexistence (as indicated by the overall index $g_{all} = -0.15 \pm 0.44$). Where cohabitation appears possible, our results further suggest an asymmetric scenario with a rather minoritary language (as captured by the modified coexistence index $\tilde{g}_{all} = -0.34 \pm 0.42$). Note that the large deviations in both cases arise from tallying together regions with differing outcomes. These results join the observed segregation of both languages in space (Fig. 1a) [4, 5] to suggest, in physics terms, that both languages tend to be *not miscible*. Note that we do not imply that a speaker would not be able to learn both tongues – as it is notably not the case. Rather, our results suggest that the overall situation of French and Dutch (together with the attitudes, risks, and efforts that speakers of each language appear ready to assume regarding the other tongue) are such that natural sociolinguistic dynamics would tend to segregate them. Coexistence is nevertheless *not* ruled out, but our results (together with the fact that Brussels is the only officially bilingual region in Belgium) suggest that the effort needed to enact the cohabitation of Dutch and French is higher than it would be for other coexisting tongues.

We take the moderately large interlinguistic similarity (mean value 0.656 ± 0.039 across all regions) as a possible indication that these efforts to keep both tongues were being made during the studied period. Note that we are characterizing the contact of a Latin and a Germanic languages, which are much more different than other couples of coexisting tongues (e.g., Catalan and Spanish [6], Fig. 1b). This would suggest a smaller interlinguistic similarity for the Dutch–French system (in purely linguistic terms), which should lead to less bilingualism. But the data shows an abundance of this group; hence, socioeconomic factors seem to compensate purely linguistic ones, thus making the use of both tongues more preferable than expected. The stability analysis of our model suggests that, if a balanced cohabitation of both languages is to be obtained, such socioeconomic advantage of bilingualism should be strengthened to move the sociolinguistic system further into the coexistence region. It would pertain to the people of Belgium (as day-to-day users of one or two tongues) and their structures of government to decide whether those efforts are worth making, or whether a tendency to a segregated society is preferred.

5 Contribution and concluding remarks

The main contributions of this manuscript is to study, for the first time with an analytical mathematical model, the very relevant sociolinguistic dynamics of language shift known as the Francization of Brussels. This process has been discussed with other methodologies both data-driven and based on historical and political discourse. These are relevant approaches that inspire our work. Complementary to these efforts, our reliance on a mathematical model fitted to empirical data allows us to quantify, for the first time, several aspects that might have had a causal effect on the sociolinguistic process (e.g., perceived interlinguistic similarity or perceived Dutch and French prestige).

Beyond our current contribution, we hope that the future study of sociolinguistic dynamics brings further together mathematical models, empirical data, and the discursive understanding that the humanities have reached about these relevant social processes. Joining models and data is important to discard sets of equations that cannot describe dynamics observed in the real world. Currently, our equations act as an effective model. This means that we do not know whether the actual sociolinguistic dynamics have the precise causal structure that our equations capture, but that, in effect, our model describes the empirical dynamics to a satisfactory degree. To rule out invalid models we would likely need better sources of empirical data. The data set that we explore in this paper is unusually rich, in that it has been consistently gathered with a homogeneous methodology over an extended period of time [8]. Nowadays, very abundant data of language use can be obtained, e.g., from twitter [5]. But this data is rather instantaneous and does not extend over the time scales that language shift dynamics need to unfold (i.e., decades, likely). It is within these limitations that we need to figure out how to falsify theoretical models.

Code availability All instructions needed to reproduce this work are included in the text.

Declarations

Conflict of interest We declare no competing interests.

Consent to participate Not applicable.

Consent for publication All authors have read and agree to the publication of this manuscript.

Ethics approval Not applicable

References

1. R. S. La frontière linguistique, 1878-1963. *Courrier hebdomadaire du CRISP* 2010/24-25 (no. 2069-2070) https://www.cairn.info/revue-courrier-hebdomadaire-du-crisp-2010-24-page-6.htm
2. Belgian Constitution, article 4. https://www.senate.be/doc/const_fr.html#t1
3. https://www.axl.cefan.ulaval.ca/europe/belgiqueetat-loi1962.htm
4. D. Mocanu, A. Baronchelli, N. Perra, Gonçalves B, Zhang Q, Vespignani A., The twitter of babel: Mapping world languages through microblogging platforms. PloS one **8**(4), 61981 (2013)
5. T. Louf, D. Sanchez, J.J. Ramasco, Capturing the diversity of multilingual societies. Phys. Rev. Res. **3**, 043146 (2021)
6. L.F. Seoane, X. Loredo, H. Monteagudo, J. Mira, Is the coexistence of Catalan and Spanish possible in Catalonia? Palgrave Commun. **5**(1), 1–9 (2019)
7. https://en.wikipedia.org/wiki/Francization_of_Brussels
8. Data extracted from the Moniteur belge / Belgisch Staatsblad are publicly available at the Wikipedia site: https://fr.wikipedia.org/wiki/R
9. D.M. Abrams, Strogatz SH modelling the dynamics of language death. Nature **424**, 900 (2003)
10. J. Mira, Á. Paredes, Interlinguistic similarity and language death dynamics. Europhys. Lett. **69**, 1031–1034 (2005)
11. J. Mira, L.F. Seoane, J.J. Nieto, The importance of interlinguistic similarity and stable bilingualism when two languages compete. New J. Phys. **13**, 033007 (2011)
12. M. Musa-Juane, L.F. Seoane, A. Muñuzuri, J. Mira, Urbanity and the dynamics of language shift in Galicia. Nat. Commun. **10**, 1680 (2019)
13. M.V. Otero-Espinar, L.F. Seoane, J.J. Nieto, J. Mira, An analytic solution of a model of language competition with bilingualism and interlinguistic similarity. Phys. D. **264**, 17–26 (2013)
14. R. Colucci, J. Mira, J.J. Nieto, M. Otero-Espinar, Coexistence in exotic scenarios of a modified Abrams-Strogatz model. Complexity **21**(4), 86–93 (2016)
15. R. Colucci, J. Mira, J.J. Nieto, M.V. Otero-Espinar, Non trivial coexistence conditions for a model of language competition obtained by bifurcation theory. Acta Appl. Math. **146**(1), 187–203 (2016)
16. L.F. Seoane, J. Mira, Modeling the life and death of competing languages from a physical and mathematical perspective, in *Bilingualism and Minority Languages in Europe: current trends and development*. ed. by F. Lauchlan, M.C. Parafita-Couto (Cambridge Scholars Press, Cambridge, 2017), pp. 70–93
17. I. Baggs, H.I. Freedman, A mathematical model for the dynamics of interactions between a unilingual and a bilingual population: Persistence versus extinction. J. Math. Sociol. **16**(1), 51–75 (1990)
18. I. Baggs, H.I. Freedman, Can the speakers of a dominated language survive as unilinguals?: a mathematical model of bilingualism. Math. Comput. Model. **18**(6), 9–18 (1993)
19. J.P. Pinasco, L. Romanelli, Coexistence of languages is possible. Physica A **361**(1), 355–360 (2006)
20. X. Castelló, V.M. Eguíluz, Miguel M. San, Ordering dynamics with two non-excluding options: bilingualism in language competition. New J. Phys. **8**(12), 308 (2006)
21. J.W. Minett, W.S. Wang, Modelling endangered languages: The effects of bilingualism and social structure. Lingua **118**(1), 19–45 (2008)
22. A. Kandler, J. Steele, Ecological models of language competition. Biol. Theory **3**(2), 164–173 (2008)
23. A. Kandler, Demography and language competition. Hum. Biol. **81**(3), 181–210 (2009)
24. M. Patriarca, E. Heinsalu, Influence of geography on language competition. Physica A **388**(2–3), 174–186 (2009)
25. A. Kandler, R. Unger, J. Steele, Language shift, bilingualism and the future of Britain's Celtic languages. Philos. T. R. Soc. B **365**(1559), 3855–3864 (2010)
26. X. Castelló, L. Loureiro-Porto, Miguel M. San, Agent-based models of language competition. Int. J. Sociol. Lang. **2013**(221), 21–51 (2013)
27. M. Zhang, T. Gong, Principles of parametric estimation in modeling language competition. Proc. Nat. Acad. Sci. **110**(24), 9698–9703 (2013)
28. E. Heinsalu, M. Patriarca, J.L. Léonard, The role of bilinguals in language competition. Adv. Complex Syst. **17**(01), 1450003 (2014)
29. N. Isern, J. Fort, Language extinction and linguistic fronts. J. R. Soc. Interface **11**(94), 20140028 (2014)
30. K. Prochazka, G. Vogl, Quantifying the driving factors for language shift in a bilingual region. Proc. Nat. Acad. Sci. **114**(17), 4365–4369 (2017)
31. S. Mimar, M. Mussa Juane, J. Mira, J. Park, A.P. Muñuzuri, G. Ghoshal, Linguistic evolution driven by network heterogeneity and the turing mechanism. Phys. Rev. Res. **3**, 023241 (2021)
32. C. Castellano, S. Fortunato, V. Loreto, Statistical physics of social dynamics. Rev. Mod. Phys. **81**(2), 591 (2009)

33. R. Solé, B. Corominas-Murtra, J. Fortuny, Diversity, competition, extinction: the ecophysics of language change. J. R. Soc. Interface **7**(53), 1647–1664 (2010)
34. M. Boissonneault, P. Vogt, A systematic and interdisciplinary review of mathematical models of language competition. Humanit. Soc. Sci. Commun. **8**(1), 1–12 (2021)
35. R. Janssens, Language use in Brussels and the position of Dutch. Some recent findings. Bruss. Stud. **13**, (2008). https://journals.openedition.org/brussels/520. https://doi.org/10.4000/brussels.520