REVIEW ARTICLE

Check for updates

# A systematic and interdisciplinary review of mathematical models of language competition

Michael Boissonneault[1]✉ & Paul Vogt[2]

During the last three decades, scientists in formal and natural sciences have been proposing models of language competition. Such models could prove instrumental in informing efforts made towards preserving the world's linguistic diversity but have yet to gain significant interest among linguists. This situation could be due to a lack of overlap between the concepts and methods used in those models and those used by linguists. In an effort towards promoting interdisciplinary dialogue on the topic of language competition, this study describes the concepts and methods used in mathematical models of language competition and assesses whether these concepts and methods are becoming more similar over time to those used by linguists. To this end, studies that proposed mathematical models of language competition were systematically retrieved and analysed. Change over time in those models was first assessed concerning the way they are specified, including the parameters they contain. Next, it was checked whether models were increasingly fitted to empirical data. Finally, change in the disciplines covered by the journals where those models were published was evaluated. Results show that overall, models have been including few sociolinguistic parameters, have been relying little on empirical data, and have been mostly published in journals covering the fields of mathematics and physics. However, the last years have seen an important turnaround along each of these three axes. A common language seems to be emerging between fields regarding mathematical models of language competition, which should prove instrumental in informing efforts made towards preserving the world's linguistic diversity.

## Introduction

Since at least the 1950s, linguists have been playing a leading role on the language preservation scene. They were the first to sound the alarm about the perplexing pace at which languages are becoming extinct worldwide (Dorian, 1981; Hale et al., 1992; Krauss, 1992). They have been at the forefront of efforts made towards preserving the world's linguistic diversity, notably within international organisations such as the United Nations Educational, Scientific and Cultural Organization (UNESCO) (Grenoble and Whaley, 2006; Moseley, 2010). Finally, linguists have continued to save moribund languages from oblivion through extensive fieldwork and documentation (Crystal, 2000; Seifart et al., 2018).

---

[1] Netherlands Interdisciplinary Demographic Institute (NIDI-KNAW), University of Groningen, The Hague, Netherlands. [2] Department of Cognitive Science and Artificial Intelligence, Tilburg School for Humanities and Digital Sciences, Tilburg University, Tilburg, Netherlands. ✉email: boissonneault@nidi.nl

Recently, scientists in formal and natural sciences have also started to show interest in the problem of language death, albeit from a different perspective. Starting in the early 1990s, they have been proposing mathematical models that aim to establish the mechanisms through which languages go extinct (Seoane and Mira, 2017). Knowledge gained from such models could be tapped into to help to slow down the rate of language extinction worldwide but have yet to gain significant interest among linguists and other researchers in the humanities and social sciences. This situation could be due to a lack of overlap between the concepts and methods used in those models and those used by linguists. In fact, researchers interested in the question already suggested that it is "because models have not sought to engage with the intellectual framework used by linguists" that "the real influence of mathematical modelling has been severely limited in the field of language revitalization" (Fernando et al., 2010, p. 49). In an effort towards promoting interdisciplinary dialogue on the topic of language competition, this study aims to describe the concepts and methods used in mathematical models of language competition, assess how models have changed over time, and determine whether there is any common language that is emerging between the different fields of science in modelling how languages become extinct.

To achieve this aim, a systematic review of the literature on mathematical models of language competition was performed. Information was retrieved from relevant studies and organised along three axes. The first axis concerns the methods used for modelling language competition. This point is important because mathematical models of language competition were initially developed based on methods that already existed in formal and natural sciences, but not necessarily in linguistics. However, research on models of language competition could provide practitioners involved in language preservation more practical tools if it relied more on methods that are common to the different disciplines involved. Below, the methods on which models were built are characterised concerning the level of analysis (macroscopic, mesoscopic, microscopic), the form of the model (equation vs. simulation-based), the linguistic composition of the population being modelled, and the parameters included.

The second axis concerns the use of data for validating models. Mathematical models of language competition are often purely theoretical and do not take empirical data as input. Conversely, linguistic analyses of language endangerment are typically data-driven. In the second part of the results section, studies on language competition are analysed concerning whether they validate their models against empirical data and whether there is any trend towards greater use of such data. Also, for each study that fitted its model to empirical data, a detailed account of the languages considered is provided, alongside the regions (or countries) covered.

The third axis concerns the way that research on mathematical models of language competition is communicated among different disciplines. Poor communication has long been recognised as an impediment for successful collaboration between disciplines (Bracken and Oughton, 2006). If publications on language competition appear predominantly in journals covering formal and natural sciences, linguists and social scientists are less likely to take note of them. On the other hand, mathematical models of language competition might more easily reach linguists and social scientists if they are published in multidisciplinary journals or journals covering the fields of linguistics or social sciences. Furthermore, an increase in the number of publications on mathematical models of language competition in journals covering the fields of linguistics and social sciences could be an indication of a certain appropriation of such models by linguists and social scientists. The third part of the results section therefore provides an analysis of the disciplines covered by the journals in which models of language competition were published.

Other reviews have already covered some of the studies that are reviewed here (Gong et al., 2014; Schulze and Stauffer (2006a, 2006b); Seoane and Mira, 2017; Solé et al., 2010; Vogt, 2009; Wang and Minett, 2005). This study differs from those on at least three points. First, the focus of most of the previous reviews was broader than the focus of the present review. Those reviews discussed studies that used mathematical models to solve different problems allying languages and population dynamics, including—but not limited to—language competition. In contrast, our review concentrates specifically on language competition, which allows a more in-depth discussion of the relevant literature. Second, none of the previous reviews employed a systematic approach at retrieving and analysing articles. As a result, they could not provide a completely balanced and objective picture of the literature, which this study aims at doing. Third, the goal of the previous reviews was primarily to describe the main findings of the studies they covered. The present review mainly focuses on the methodological innovations that have taken place in the field over the years, which we hope will help us achieve our aim formulated above.

## Methods

**Search strategy**. Methods follow the Prisma statement for reporting systematic reviews (Moher et al., 2009). Relevant articles were retrieved using the search engines of Arxiv, Scopus and Web of Science. Articles must have included in their title the word "language" in combination with at least one of the following words: death, competition, extinction, endangerment, shift, disappearance, invasion, revitalization, coexistence, survival, or preservation. To avoid finding too many articles from psycholinguistics—which are not relevant for our purposes—articles must not have included in their title the following words: competence, teaching, learning, processing, disorder, acquisition and comprehension. We further specified that the word "model" should appear in either the title, abstract or keyword list. Truncation was used to allow for different words that share the same root and meaning. To avoid overseeing any important work, articles that cited two seminal studies in the field were searched using the Scopus search engine. These two studies are those of Baggs and Freedman (1990) and Abrams and Strogatz (2003). The reference lists of selected articles were also checked, and experts in the field were consulted for missing articles. Eligible studies must have been research articles, thus excluding book chapters, and must have been written in English. Searches were performed on the 25th of May 2020 and refreshed on the 27th of July 2020. The exact word strings used in searches are provided in Table S1 of the supplementary materials.

**Article selection and data extraction**. Figure 1 shows how studies were selected for review. The different search streams provided a total of 821 records, of which 624 remained after duplicates were removed. The first round of screening was performed based on titles and abstracts. A total of 134 articles with mathematical models of language competition as a topic was selected for a complete assessment. At this stage, more articles were excluded if they assessed the sensitivity or qualitative properties of existing models without proposing a new one (model analyses), if they were not published in the form of a research article, or if they did not have sufficient information about the type of model they presented. This resulted in 56 studies left for data extraction, to which six were added from the reference lists of the selected studies or based on expert knowledge, resulting in a total of 62 studies selected for analyses.
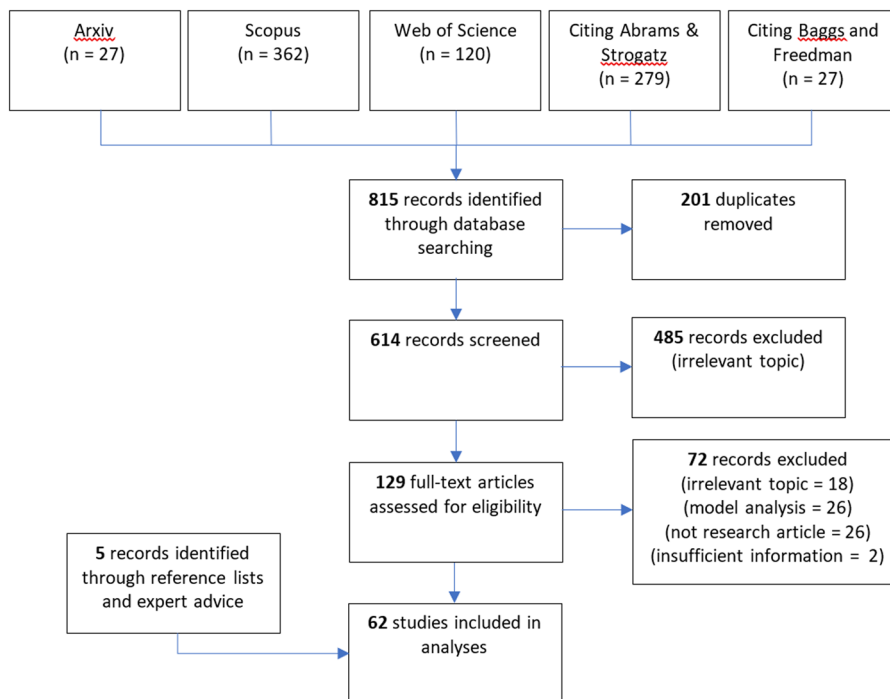
**Fig. 1 Article selection process.**

We extracted from the 62 selected studies information about authors' names, year and journal of publication, the level of analysis (macroscopic, mesoscopic, microscopic), the composition of the population concerning its linguistic groups, the parameters included in the models, and the languages and region or country of interest covered by the data (in case empirical data were used). Model parameters were considered as being the same if they denoted similar concepts despite being called differently. For example, the parameter "Status", which is sometimes used to refer to the socioeconomic position of the speakers of a given language, was considered as equivalent to the parameter "Prestige". A full account of the parameters included in the analyses and what they were called in the different studies is given in Table S2 of the supplementary materials.

### Analytical framework

**Evolution of methods**. The way that methods evolved over the years is accounted for using four characteristics. The first one refers to the level of analysis. We distinguish between the macroscopic, mesoscopic and microscopic levels. Studies on models of language competition, and mathematical models of social processes in general, commonly distinguish between the macroscopic and microscopic approaches (Castellano et al., 2009; Stauffer and Schulze, 2005). Macroscopic models describe population processes from an aggregate level. Results translate mean outcomes and do not account for the variability among individuals. An example of an early macroscopic model is provided by Abrams and Strogatz (2003), who use a differential equation to describe change over time in the proportion of speakers among two competing languages and apply their model to different situations of language competition around the world. Microscopic models, on the other hand, describe population processes considering each individual separately. This approach confers certain advantages compared to the macroscopic approach. More specifically, microscopic models allow to explore the full range of an outcome instead of just its mean. These models further allow for stochasticity, which can play an important role among smaller populations. Finally, they allow to

explicitly consider interactions between individuals. An example of an early microscopic model is provided by Castelló et al. (2006), who define a population of agents speaking either language A or B, or both (bilingual agents). Agents interact with each other and acquire a second language under the influence of repeated contact with the speakers of another language, or abandon an already known language due to a lack of interaction with the speakers of that language. Studies less often consider mesoscopic models as a separate approach. In practice, mesoscopic models are similar to microscopic models in that both consider each individual separately and allow for stochasticity and interactions. However, in mesoscopic models, the unit of analysis is not the speaker like in microscopic models, but rather groups of speakers, or even entire languages. An example of an early mesoscopic model is provided by Schulze and Stauffer (2005), who define a set of bit-strings where each different string represents a different language. The strings—or languages—can duplicate themselves, which leads to an increase in the number of speakers of that language, they can undergo mutations, which represents the phenomenon of language birth, or they can disappear, which represents language death. This kind of model is usually used to study the world's current languages distribution concerning their numbers of speakers, i.e., to explain why a few languages are spoken by a large share of the world's population while most languages are spoken by only a fraction of it.

The second characteristic used to account for the way that models have evolved refers to the type of model or the form of the equation used. Six types were identified. We first consider three types of models that are based on ordinary differential equations: ordinary differential equations models (ODE), Lotka–Volterra models (LV), and reaction-diffusion model (RD). ODEs refer to the simpler form of ordinary differential equations models. These models estimate the change in the proportion of speakers belonging to a language as a function of time and some parameters and can be fitted to data that include only limited information. In these models, the total population is normalised, meaning that these models assume that it is not the absolute size of a linguistic group that matters, but its respective proportion in

the whole population. It follows that in an ODE model with two linguistic groups, growth in one group means a decline in the other one. LV models are also based on ordinary differential equations, but additionally include information about group size, allowing the size of a linguistic group to change independently from its proportion in the whole population. Diffusion terms can be added to ordinary differential equations (including to LV models) to allow speakers to spread in space, in which case the model will be referred to as an RD model.

We further distinguish models that were developed using a system dynamics (SD) approach. These models offer the same possibilities as an ordinary differential equations model but are expressed in terms of stocks and flows rather than in terms of equations, which may increase their accessibility to non-mathematicians (Wyburn and Hayward, 2009). Furthermore, such models often make explicit some parameters that otherwise remain implicit in ordinary differential equations models.

Studies that rely on ODE, LV, RD and SD models usually model language competition from a macroscopic level. Among those studies that took a mesoscopic or microscopic approach, most relied on agent-based (AB) models or a conversation game (CG) framework, or a combination of both. AB models simulate interactions between speakers according to a set of pre-established rules. Due to the complex nature of these interactions, these models often provide insights that could not be obtained using equation-based models. CG, on the other hand, is a form of game theory that concentrates on decisions made by speakers during conversations. Speakers' information about the world is imperfect and leads them to make unexpected choices in multilingual settings. This approach was used on its own or in combination with agent-based models, in which case language shift depends on a series of one-on-one interactions. One study could not be assigned to any of the model types, presenting results from a simulation based on a series of equations. It will be referred to as "Other simulations" (OS).

The third characteristic refers to the composition of the population concerning its linguistic groups, i.e., their number and whether bilingualism is considered. Among the studies reviewed here, competition is either considered among two, three or four languages, or among hundreds or even thousands of them. While the former refers to competition in a specific country or a region, the latter refers to competition in the whole world. Bilingualism refers to the explicit modelling of speakers that are fluent in two or more languages.

The fourth and final characteristic refers to the parameters included in each model. All models consider at least one parameter that drives growth in the number of speakers of one language to the expense of the number of speakers of another language. The universe of such parameters is presented in Table 1. To facilitate analyses, parameters were assigned to three broad categories. The first category includes parameters that do not directly predict the intensity of the shift between languages but instead specify the form in which it occurs, i.e., whether it occurs through a horizontal, vertical or exogenous transmission. Horizontal transmission occurs from adults to adults, vertical transmission (also called intergenerational transmission) occurs from parents to their children, and exogenous transmission occurs via institutions such as schools. The second category refers to geodemographic parameters. These parameters describe how speakers occupy the space, reproduce, migrate and die. The third category refers to sociolinguistic parameters. These include the prestige conferred to speakers of a given language, the language to which speakers identify, and language use itself, e.g., the level of fluency. Some parameters directly influence shift. For example, speakers are more likely to shift towards languages that are spoken by larger numbers of speakers. Other parameters have an indirect influence on shift. For example, groups of speakers with higher birth rates will have higher numbers of speakers not because more births induce more speakers, but because more births allow for more vertical transmission.

**Data usage**. Information on data usage includes languages and countries (or regions) covered by the data, for each study. We consider that a study made use of data if it explicitly mentions the languages and country (or regions) covered by the data, and if data minimally reflects the numbers or proportions of speakers of two or more competing languages at two or more points in time, or transitions intensities between languages at different points in time. Conversely, we consider that a study did not make use of data in cases where the model is strictly theoretical, or in cases where data only reflects proportions of speakers at one point in time, which is often the case in studies that aim at reproducing the world's languages distribution according to their number of speakers. A distinction is made between studies that used data on a specific population for the first time or for a second time or more. These will be referred to as first and second analysis, respectively. Populations are considered different when they either speak a different combination of languages or speak a similar combination of languages in different countries (or regions). For example, several studies considered competition between Welsh and English in Wales. The study that first used data on the combination Welsh/English in Wales is considered as a first analysis. All subsequent studies that used data on the combination Welsh/English in Wales are considered as second analyses.

**Discipline coverage**. Scopus' *Subject area classifications* were used to assign a scientific discipline to each journal in which studies were published (Scopus, 2020). These "classifications" include a total of 26 disciplines covering the whole of the health, life, physical and social sciences. Scopus may assign more than one discipline to each journal. We only consider the first of those disciplines, which is also the main one. Journals in which the studies analysed here were published cover eleven of Scopus' 26 disciplines, namely: Agricultural and biological sciences; Arts and humanities; Biochemistry, genetics and molecular biology; Chemistry; Decision sciences; Economics, econometrics and finance; Engineering; Mathematics; Multidisciplinary; Physics and astronomy; and Social sciences. To facilitate analyses, these eleven categories were merged into six: Biology; Economics; Linguistics and other humanities; Multidisciplinary; Physics and Mathematics; and Social Sciences. The exact correspondences are presented in Table S3 of the supplementary materials.

## Results

**Number of publications by year**. According to our selection, the first mathematical model of language competition was proposed in 1990. It is however during the mid-2000s that this topic started to gain in popularity, mostly as a result of Abrams and Strogatz' influential study published in 2003. Studies on models of language competition continue to be more or less regularly published to this day, with between zero and eight publications per year since 2005. Information on the number of publications per year can be derived from Figs. 2 and 3 presented below.

**Evolution of methods**. Table 2 presents a comprehensive overview of how methods for modelling language competition evolved over the last thirty years. Studies are ordered chronologically. References in column 1 are colour-coded to represent the level of analysis. Column three shows the model employed, while column four shows the composition of the population according to its

**Table 1 List of parameters identified in the literature review and short description.**

| Parameter name | May also include | Description |
|---|---|---|
| **A. Shift parameters** | | |
| Horizontal transmission | Non-infant language acquisition, conversion rate, language switch | Models how speakers shift to a new language as a result of the influence of people others than their parents. |
| Vertical transmission | Infant language acquisition, Intergenerational transmission, Inheritance rate | Models how speakers shift to a new language as a result of the influence of the parents. |
| Exogenous influences | Education, media presence | Models how speakers shift to a new language as a result of schooling or media usage. |
| **B. Geodemographic parameters** | | |
| No. of speakers | Percentage of speakers, fraction of speakers, speaker density, frequency of language use | Captures the influence that the number of speakers of the own or the other language exercises on the likelihood of shifting to the other language. |
| Births and deaths | Growth rate, reproduction, population change | Captures the influence of births and deaths on the number of speakers of a language. |
| Migration | Immigration, emigration, movement | Captures the influence of migration on the number of speakers of a language. |
| Space | Lattice, proximity, distance | Captures the fact that greater distance between speakers of different languages reduces the likelihood of shift between those languages. |
| Movement | Spread, dispersal | Captures the fact that movement may increase or decrease the distance between speakers of different languages. |
| Border | Barrier, zones, island, refuge | Captures any natural or political obstacle that reduces the influence that speakers may have on each other. |
| Fitness | | Captures the fact that certain speakers have a greater capacity for reproduction, thus increasing their opportunities for vertical transmission. |
| Ageing | | Captures the fact that shifts may take different forms depending on the age of a speaker. |
| **C. Sociolinguistic parameters** | | |
| Prestige | Attractiveness, status, utility, attraction | Captures the socio-economic position of speakers, where groups with higher prestige are more likely to attract new speakers. |
| Loyalty | Volatility, resistance to change, identity, attachment, ideology, attitude | Captures the resistance of speakers to learning a new language or their attachment towards their language. Both reduce the propensity of speakers to shift to another language. |
| Network topology | | Captures the fact that speakers in a population have different likelihoods of interacting with each other and thus different influences on each other regarding shift. |
| Homophily | Family formation, mating | Captures the fact that speakers who share a common language are more likely to have children with each other than speakers who do not share a common language. |
| Environment | Urban/rural, media | Captures the possible influences of the environment on language shift, such as the fact of living in a city or a rural area, or the fact of being exposed to media in a given language. |
| Similarity | Distance | Captures the fact that speakers are more likely to learn each other's languages if the languages that they speak are similar to each other, effectively leading to more bilingualism. |
| Attrition | Memory, forget | Captures the fact that speakers may forget a language that they once knew upon ageing, effectively leading to less bilingualism. |
| Change | Mutations | Captures changes in the structure and vocabulary of a language over time. May lead to language birth. |
| Accommodation | Language use, personality type, diglossia | Captures the efforts made by speakers to adapt to an interlocutor or a specific situation, thus influencing language choice among bilinguals. |
| Competency | Fluency, proficiency, skills | Captures the variation in fluency among speakers of a second language, which may influence language choice. |

linguistic groups. Column five contains tiles which show the parameters used in each model. Information is also included about data usage, where darker tiles indicate the use of data, and about the journal of publication's main discipline, in column two. These last two points will the focus of the next two subsections and will not be commented on further here.

Early models of language competition took a macroscopic approach, and this approach remained popular throughout the observation. Most of these models are based on a system of ordinary differential equations, though many studies published between 2005 and 2014 used models of the types Lokta-Volterra, reaction-diffusion or system dynamics. Models that took a microscopic approach started to appear from 2005 onwards. This approach continued to be regularly utilised over time, relying either on agent-based or conversation games models or on both. The mesoscopic approach, on the other hand, was mostly used between 2005 and 2007 in the form of agent-based models. These focused on large numbers of languages as they were mostly used to study the world's linguistic diversity.

Turning to the parameters included in each model, a vast majority of models considered horizontal shift between speakers of different languages, while about half as many considered vertical shift. A few studies considered exogenous influences on language shift, while a few did not include any shift parameter but
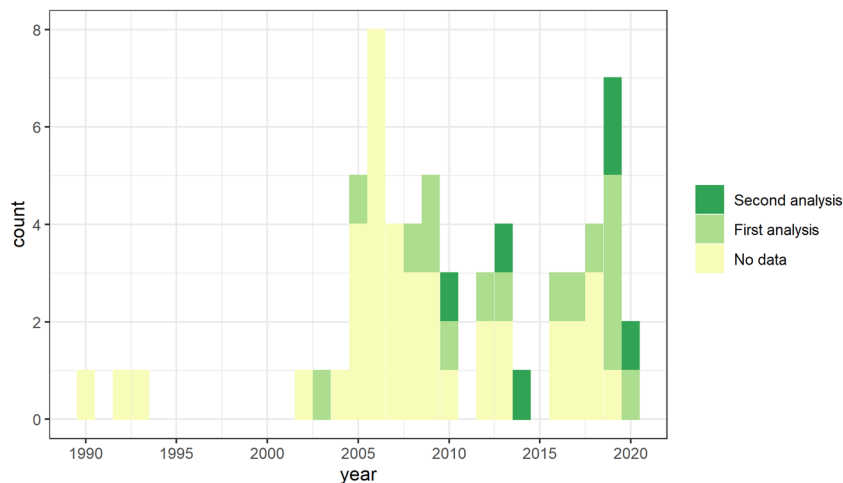
**Fig. 2 Studies breakdown by year and data usage.** Colours indicate whether a study used data or not, and if so, whether it used data on a specific population for the first time (First analysis) or for a second time or more (Second analysis). Populations are defined according to the languages they speak and the region or country they inhabit.
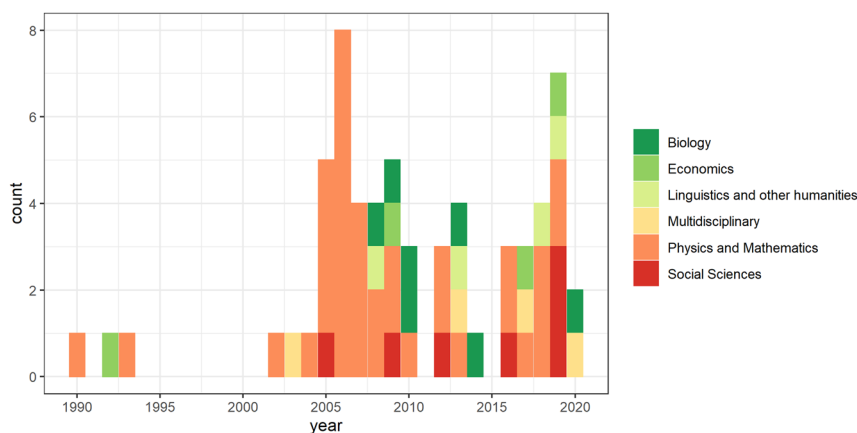


**Fig. 3 Study breakdown by year and main discipline of journal of publication.** Disciplines are based on Scopus' Subject area classification. Disciplines were merged into a reduced number of categories to facilitate interpretation (see Table S3 of the supplementary materials).

rather concentrated on the choice of different languages in different contexts using a conversation game framework. The choice of language shift parameters does not appear to have significantly varied over time or to systematically change according to the type of model used.

Regarding geodemographic parameters, the parameter "number of speakers" was included in an overwhelming number of studies. That is, most studies let shift between linguistic groups depend at least in part on the size of those groups: large groups thus tend to become larger and smaller groups tend to become smaller. Many studies additionally considered the births and deaths of speakers as a factor influencing the size of linguistic groups. Many studies considered the role of space, acknowledging the fact that speakers who are separated by a greater physical distance have less influence on each other than speakers who are closer to each other. Among the studies that considered space, more than half considered movement, a feature often modelled through means of reaction-diffusion models. Few studies considered the role of migration or borders. A small number of studies, all based on the mesoscopic approach, assumed that some speakers have a higher propensity to reproduce than others, i.e., they have higher fitness. Five studies distinguished different phases in the lives of speakers by including ageing in their

models. The use of features such as borders, fitness and ageing went from rare to almost inexistent towards the end of the period of observation.

The most commonly included sociolinguistic parameter is prestige, followed closely by loyalty. These two parameters can be considered as two opposing forces influencing language shift: speakers shift to a new language due to its prestige but remain loyal to the already known one for convenience or cultural reasons. These two parameters are to be found in most models that took a macroscopic perspective, as well as in some microscopic models, but are absent from mesoscopic models. The parameter loyalty tends to appear more in models that were proposed from 2012 onwards. The parameter language change, on the other hand, appeared in many mesoscopic models published in the mid-2000s but was almost completely abandoned afterwards. In fact, language change was mainly operationalized in the form of random mutations affecting artificially created languages (e.g., bit-strings) in the context of mesoscopic models. Other sociolinguistic parameters have become more common over time. This is the case of the parameters network topology, homophily, environment, accommodation and competency. These were often included in microscopic models, or in models that took a system dynamics approach. To finish,

**Table 2 Chronological list of studies broken down by type of approach, model employed, composition of the population by linguistic groups, use of data and parameters.**

Column groups: Shift parameters (Horizontal, Vertical, Exogenous); Geodemographic parameters (No. of speakers, Births & Deaths, Migration, Space, Movement, Borders, Fitness, Ageing); Socio-linguistic parameters (Prestige, Loyalty, Networks, Homophily, Environment, Similarity, Attrition, Change, Accommodation, Competency); TOTAL.

| Study | Journal discipline | Model | Composition | TOTAL |
|---|---|---|---|---|
| Baggs & Freedman 1990 | Phys. & Math. | LV | A,AB | 5 |
| Grin 1992 | Economics | ODE | A,AB | 4 |
| Baggs & Freedman 1993 | Phys. & Math. | RD | A,B,AB | 4 |
| El-Owaidy & Ismail 2002 | Phys. & Math. | ODE | A,B,C,AB | 4 |
| Abrams & Strogatz 2003 | Multidisc. | ODE | A,B | 4 |
| Patriarca & Leppänen 2004 | Phys. & Math. | RD | A,B | 6 |
| Kosmidis et al. 2005 | Phys. & Math. | AB/CG | A,B | 7 |
| Mira & Paredes 2005 | Phys. & Math. | ODE | A,B | 5 |
| Schwämmle 2005 | Phys. & Math. | AB/CG | A,B,AB | 9 |
| Schulze & Stauffer 2005 | Phys. & Math. | AB | A,B,C,... | 5 |
| Wickström 2005 | Social Sci. | ODE | A,B,AB | 5 |
| Castelló et al. 2006 | Phys. & Math. | AB | A,B,AB | 3 |
| De Oliveira et al. 2006a | Phys. & Math. | AB | A,B,C,... | 5 |
| De Oliveira et al. 2006b | Phys. & Math. | AB | A,B,C,... | 5 |
| Pinasco & Romanelli 2006 | Phys. & Math. | LV | A,B | 4 |
| Schulze & Stauffer 2006 | Phys. & Math. | AB | A,B,C,... | 4 |
| Stauffer et al. 2006 | Phys. & Math. | AB | A,B,C,... | 6 |
| Tesileanu & Meyer-Ort. 2006 | Phys. & Math. | AB | A,B,C,... | 6 |
| Tesileanu & Meyer-Ort. 2006 | Phys. & Math. | AB | A,B,C,... | 6 |
| Schwämmle 2006 | Phys. & Math. | AB/CG | A,B,AB | 6 |
| De Oliveira et al. 2007 | Phys. & Math. | AB | A,B,C,... | 5 |
| Schulze & Stauffer 2007 | Phys. & Math. | AB | A,B,C,... | 5 |
| Stauffer et al. 2007 | Phys. & Math. | AB | A,B | 4 |
| Tuncay 2007 | Phys. & Math. | AB | A,B,C,... | 4 |
| Hadzibeganovic et al. 2008 | Phys. & Math. | AB | A,B,C | 4 |
| Minett & Wang 2008 | Ling. & Hum. | ODE | A,B,AB | 5 |
| Minett & Wang 2008 | Ling. & Hum. | AB | A,B,AB | 6 |
| Wyburn & Hayward 2008 | Phys. & Math. | LV | A,AB | 5 |
| Kandler & Steele 2008 | Biology | RD | A,B | 5 |
| Beltran et al. 2009 | Social Sci. | AB/CG | A,AB | 5 |
| Kandler 2009 | Biology | RD | A,B,AB | 6 |
| Patriarca & Heinsalu 2009 | Phys. & Math. | RD | A,B | 7 |
| Schwämmle et al. 2009 | Phys. & Math. | OS | A,B,C,... | 4 |
| Wyburn & Hayward 2009 | Economics | SD | A,B,AB | 6 |
| Fernando et al. 2010 | Biology | SD | A,B,AB | 7 |
| Kandler et al. 2010 | Biology | RD | A,B,AB | 6 |
| Wyburn & Hayward 2010 | Phys. & Math. | SD | A,B,AB | 4 |
| An et al. 2012 | Phys. & Math. | AB/CG | A,B,AB | 5 |
| Bakalis & Galani 2012 | Phys. & Math. | ODE | A,AB | 5 |
| Iriberri & Uriarte 2012 | Social Sci. | CG | A,AB | 3 |
| Caridi et al. 2013 | Phys. & Math. | AB | A,B | 7 |
| Castelló et al. 2013 | Ling. & Hum. | AB | A,B,AB | 5 |
| Fort & Pérez-Losada 2013 | Biology | RD | A,AB | 5 |
| Zhang & Gong 2013 | Multidisc. | RD | A,B | 7 |
| Isern & Fort 2014 | Biology | RD | A,B | 7 |
| Parshad et al. 2016 | Phys. & Math. | RD | A,B,C | 8 |
| Templin et al. 2016 | Social Sci. | ODE | A,B,AB | 6 |
| Carro et al. 2016 | Phys. & Math. | AB/CG | AB | 3 |
| Lipowska & Lipowski 2017 | Phys. & Math. | AB/CG | A,B,AB | 4 |
| Prochazka & Vogl 2017 | Multidisc. | RD | A,B | 6 |
| Clingingsmith 2017 | Economics | AB | A,B,C,... | 4 |
| Barreira da Silva Rocha 2018 | Phys. & Math. | LV | A,B,AB | 7 |
| Karjus & Ehala 2018 | Ling. & Hum. | AB/CG | A,B,C | 6 |
| Tanaka et al. 2018 | Phys. & Math. | ODE | A,B | 6 |
| Wyburn 2018 | Phys. & Math. | LV | A,AB | 5 |
| Civico 2019 | Social Sci. | AB/CG | A,AB | 10 |
| Laruelle et al. 2019 | Social Sci. | CG | A,AB | 2 |
| Mussa et al. 2019 | Phys. & Math. | ODE | A,B,AB | 7 |
| Seoane et al. 2019 | Ling. & Hum. | ODE | A,B,AB | 5 |
| Templin 2019 | Social Sci. | ODE | A,B,AB | 12 |
| Wyburn 2019 | Economics | SD | A,AB | 5 |
| Wyburn & Hayward 2019 | Phys. & Math. | AB/CG | AB | 5 |
| Barrett et al. 2020 | Biology | ODE | A,B | 5 |
| Zhou et al. 2020 | Multidisc. | ODE | A,B,C | 4 |
| TOTAL | | | | |

TOTAL row parameter counts: No. of speakers: 17; Births & Deaths: 7; Migration: 3; Space: 5; (Prestige): 7; Loyalty: 5; Networks: 4; Homophily: 5; Environment: 8; Similarity: 13; Attrition: 8; Change: 5 (with several columns marked #).

**Approach**
Macroscopic
Mesoscopic
Microscopic

**Model**
*ODE* ordinary differential equations, *RDR* reaction-diffusion, *LV* Lotka-Volterra, *SD* system dynamics, *AB* agent-based, *CG* conversation game, *OS* other simulations

**Data usage**
■ : Model fitted to data
■ : Model not fitted to data

**Composition**
A,B: Monolinguals in A or B, no bilinguals; A,AB: Monolinguals in A, bilinguals in A and B (no monolinguals in B); A,B,AB: Monolinguals in A or B, bilinguals in A and B; AB: Bilinguals in A and B only; A,B,C: Monolinguals in A, B or C, no bilinguals; A,B,C,AB: Monolinguals in A, B or C, bilinguals in at least one combination of A, B and C; A,B,C,...: Monolinguals in one of more than three languages (up to thousands)

the parameter similarity was sporadically used throughout the observation, in the context of ordinary differential equations or mesoscopic agent-based models.

**Data usage**. Figure 2 displays for each year of interest counts of studies according to whether the models they presented were fitted to data, and if so, whether they used data on a specific population for the first time (first analysis) or for a second time or more (second analysis). A total of 22 studies fitted their models to empirical data. Of these, five did so before 2010, while seventeen did so in 2010 or later. Strikingly, eight of the nine studies published in 2019 and 2020 were fitted to empirical data.

Table 3 lists the languages and regions or countries for which data was used. Panel A lists the languages and the corresponding countries (or regions) that were covered in studies that modelled competition between two languages, while Panel B lists the languages and the corresponding countries (or regions) that were covered in studies that modelled competition between more than two languages. Studies modelled competition among fifteen pairs of languages, five combinations of three languages, and one combination of four languages. Twenty-seven countries or regions were covered. Different studies sometimes modelled competition between similar languages, either in a similar or a different country (or region). The same languages also sometimes appeared in different models in combination with different languages. Notably, eighteen studies considered competition involving the English language. Most studies considered language competition in economically developed countries, including European and North American countries, but also Hong Kong and Singapore.

**Discipline coverage**. Figure 3 breaks down the selected studies according to the disciplines covered by the journals in which they were published. Journals covering the fields of physics and mathematics have by far been the most popular outlets for studies on models of language competition, covering 36 publications out of a total of 62. Journals covering the field of biology and journals covering the field of social sciences come second, each with seven publications. Journals covering the field of economics, those covering the field of linguistics and other humanities, and multidisciplinary journals each published four studies on mathematical models of language competition. The dominance of journals covering the fields of physics and mathematics seems to be slowly diminishing: their share went from 75% before 2010 to 40% from that year onwards. Initially, publications in journals covering the field of biology tended to replace journals covering the fields of physics and mathematics. However, the picture is more diversified concerning the later period, with a higher number of publications in journals covering the fields of economics, linguistics and other humanities, and social sciences. We note that one particular journal was relied on heavily as an outlet for publishing models of language competition. Namely, *Physica A: Statistical Mechanics and its applications* published sixteen of the 62 publications reviewed here. In comparison, the second most common outlet is the *International Journal of Modern Physics C*, with five publications, while most other journals are represented once or twice (Table S3, supplementary materials).

## Discussion

In which direction should mathematical models of language competition evolve so that they engage better with the intellectual framework used by linguists, and thus better inform the current efforts made towards preserving the world's linguistic diversity? As seen above, mathematical models of language competition are appearing more and more in journals covering disciplines outside formal and natural sciences. This section aims at providing an answer to the question raised above by looking into how those studies published in the fields of humanities and social sciences differ from those published in the formal and natural sciences. Then, a discussion is provided as to whether models, in general, are becoming more similar over time to those recently developed in the fields of humanities and social sciences.

**Comparison of models across disciplines**. Results presented above suggest that models published in journals covering the fields of linguistics and other humanities exhibit properties that are different from models published in journals covering other disciplines. Table 4 shows how models differ between those two groups of disciplines which we, respectively, refer to as *Humanities and social sciences* and *Formal and natural sciences*. Those differences can be summarised the following way: compared to models published in journals covering formal and natural sciences, models published in journals covering humanities and social sciences rely more on a microscopic rather than a macroscopic approach, consider more often bilingualism as a separate state, privilege the use of sociolinguistic parameters over geodemographic ones, and are slightly more often based on data.

First, the more common use of the microscopic approach among linguists and social scientists might be explained by the fact this approach allows to model interactions between speakers more explicitly and accommodates a larger number of parameters. Some of the models of this class recently proposed by linguists and social scientists indeed include a large number of parameters, allowing for a high degree of precision (Civico, 2019; Karjus and Ehala, 2018). These models, furthermore enhanced by the use of data for calibration, allow to unravel the multiple pathways through which language shift can occur, and what can be done about it.

Second, the importance of bilingualism was recognised relatively early by linguists interested in questions relating to language competition (Minett and Wang, 2008). This recognition comes from observations made on the field stating that speakers rarely shift directly from one language to another, but instead go through a phase of bilingualism (Appel and Muysken, 2005). Increasingly, however, it seems that formal and natural scientists are recognising this fact as more of their models published recently are considering bilingualism (Heinsalu et al., 2014; Seoane and Mira, 2017).

Third, it speaks for itself that linguists and social scientists rely more on sociolinguistic parameters than formal and natural scientists do. Two parameters appear much more often in publications in journals covering humanities and social sciences: homophily and competence. The former refers to the greater likelihood that unions are formed among people who speak a similar language than among people who speak different languages. This is important since bilingual unions often lead to bilingual children and as noted above, bilingualism is an important vector of language shift (Appel and Muysken, 2005). The other parameter, competence, is probably equally important since the degree of fluency of speakers in one language will have a direct impact on the frequency of its usage, and the less a language is used, the more likely it is to become extinct. Other sociolinguistic parameters often encountered in models developed by linguists and social scientists include network typologies, which refers to the fact that not all speakers are equally likely to talk to each other due to, for example, shared interests and acquaintances, and loyalty, which we discuss below.

Fourth, one crucial aspect of the efforts made towards preserving the world's linguistic diversity consists in making inventories of extant languages in terms of their number of

**Table 3 Languages covered in studies based on empirical data, with corresponding countries (and regions between parentheses, if applicable) and references.**

**Panel A: Competition between two languages**

| Dominant language | Threatened language | Country (region) | References |
|---|---|---|---|
| English | French | Canada | Wyburn and Hayward (2009) |
| | French | Canada (Quebec) | Wyburn and Hayward (2009;2010) |
| | Irish | Ireland | Wyburn and Hayward (2008;2019) |
| | Mandarin | Singapore | Zhang and Gong (2013) |
| | Maori | New Zealand | Wyburn (2019; Barrett-Walker et al. (2020) |
| | Scottish Gaelic | Scotland | Wyburn and Hayward (2008); Zhang and Gong (2013) |
| | | Scotland (Sutherland) | Abrams and Strogatz (2003); Kandler et al. (2010); Isern and Fort (2014) |
| | | Scotland (Argyll) | Kandler et al. (2010) |
| | | Scotland (Inverness) | |
| | | Scotland (Ross and Cromarty) | |
| | Spanish | United States | Templin (2019) |
| | Welsh | Wales | Abrams and Strogatz (2003); Wyburn and Hayward (2008, 2009, 2010, 2018, 2019); Fort and Perez-Losada (2013; Zhang and Gong (2013); Isern and Fort (2014); Barrett-Walker et al. (2020) |
| | | Wales (Monmouthshire) | Abrams and Strogatz (2003); Isern and Fort (2014) |
| French | Breton | France (Brittany) | Wyburn and Hayward (2008) |
| German | Romansch | Switzerland (Canton of Grisons) | Civico (2019) |
| | Slovenian | Austria (Carinthia) | Prochazka and Vogl (2017) |
| Greek | Aromanian | Greece (North-Western Greece) | Bakalis and Galani (2012) |
| Mandarin | Malay | Singapore | Zhang and Gong (2013) |
| Spanish | Catalan | Spain (Valencia) | Beltran et al. (2009) |
| | | Spain (Catalonia) | Seoane et al. (2019) |
| | Galician | Spain (Galicia) | Mira and Paredes (2005); Mussa Juane et al. (2019) |
| | Quechua | Peru (Huanuco) | Abrams and Strogatz (2003); Isern and Fort (2014) |

**Panel B: Competition between more than two languages**

| Language A | Language B | Language C | Language D | Country (region or community) | Reference |
|---|---|---|---|---|---|
| English | Hinglish | Hindi | | India (Hindi belt) | Parshad et al. (2016) |
| | Dialect | Mandarin | | Singapore | Zhou et al. (2020) |
| | | | | Singapore (Chinese community) | |
| | Tamil | Malay | | Singapore (Indian community) | |
| | Hakka | Hoklo | Sze Yap | Hong Kong | |
| | Estonian | Russian | | Estonia | Karjus and Ehala (2018) |

**Table 4 Overview of models characteristics with respect to the discipline covered by the journal of publication.**

| | Formal and natural sciences | | Humanities and social sciences | |
|---|---|---|---|---|
| | Count | Percentage[a] | Count | Percentage[a] |
| Macroscopic approach | 28 | 53.8 | 5 | 41.7 |
| Microscopic approach | 11 | 21.2 | 7 | 58.3 |
| Bilingualism as separate state[b] | 14 | 26.9 | 7 | 58.3 |
| No. of geodemographic parameters | 135 | 2.6[c] | 20 | 1.7[c] |
| No. of sociolinguistic parameters | 74 | 1.4[c] | 32 | 2.7[c] |
| Based on data | 17 | 32.7 | 5 | 41.7 |
| Total number of models[d] | 52 | 100.0 | 12 | 100.0 |

[a]Percentages relate to the total number of models in each group.
[b]Bilingualism is considered as a separate state when it forms an intermediate state in the shift from A to B. Thus: A - > AB - > B.
[c]Figures correspond to the mean rather than the percentage.
[d]Totals sum up to 64 since they concern the number of models rather than studies (see Table 2).

speakers (Ethnologue, 2020). Language preservation is, in this sense, a highly applied and quantitative field. Though validation against empirical data is not a prerequisite for mathematical models of language competition to be insightful, it remains an important step that has often not been made, especially in studies published in journals covering the fields of physics and mathematics.

Is the discrepancy between sciences in the techniques used for modelling of language competition due to a lack of engagement among formal and natural scientists with the intellectual framework used by linguists? Part of the explanation for this discrepancy, which was previously raised by Fernando et al. (2010), could lie in the origins of mathematical models of language competition. Many of these models were first developed to answer questions relating to natural phenomena, for example about the transmission of infectious diseases among living species, the spread of particles in space, or predator-prey dynamics (Kandler and Unger, 2018; Prochazka and Vogl, 2018). It will not come as a surprise, therefore, that these models focus in the first place on population dynamics and interactions between individuals and their environment—thus on geodemographic parameters—rather than on sociolinguistics phenomena. Obviously, when applying their models to questions relating to language competition, formal and natural scientists included parameters that reflect sociolinguistic realities. For example, early models included, next to population size, the notion of prestige as an important factor influencing shift from one language to another (e.g., Abrams and Strogatz, 2003). However, these models often failed to also consider the fact that speakers sometimes resist to shifting to a more prestigious language due to loyalty towards their heritage language, an important factor in efforts made towards language preservation (Thomason, 2015). Other early models considered language similarity (e.g., Mira and Paredes, 2005) or language change (e.g., Stauffer and Schulze, 2005). However, these concepts were often implemented based more on mathematical convenience than on phonological or syntactical properties of the languages involved, making them highly vulnerable to criticism from the linguistic perspective.

**Comparison of models over time**. The first mathematical model of language competition was developed thirty years ago, but until ten years ago, almost all models were developed by formal and natural scientists. The last ten years have seen more models developed by linguists and social scientists, but these continue to form only a minority of all models. Taken as a whole, have models evolved to resemble more to those recently published by linguists and social scientists?

Results presented above suggested change over time in a few characteristics inherent to models of language competition.

Table 5 quantifies this change concerning the year 2010, allowing to divide models into two roughly equal numbers. The characteristics considered are the same as in Table 3, which was presented in the previous subsection. Change between the two periods is rather minor concerning the approach taken and the consideration of bilingualism as a separate state. However, there are clear increases in the use of sociolinguistic parameters and on the reliance on data. Part of this increase could be explained by the fact that after 2010, linguists and social scientists started to more regularly develop their own models of language competition. As we saw above, models developed by linguists and social scientists tend to differ considerably in the parameters they use and their use of data. And indeed, as shown in Table 4, the number of publications in journals covering the fields of humanities and social sciences doubled between the period pre-2010 and the period post-2010. Linguists have been increasingly relying on methods developed in the formal and natural sciences over the last years (Bromham, 2017) and this trend seems to apply to the field of language competition as well. By doing so, linguists and social scientists may have contributed to changing practices in the field of mathematical modelling of language competition. However, analyses shown in Table S4 of the Supplementary materials suggest that practices have also started to change among formal and natural sciences, as we notice there an increase in the mean number of sociolinguistic parameters over time. We could thus be witnessing the emergence of a common language between fields concerning mathematical models of language competition.

### Conclusion: a new wave of research on mathematical models of language competition?

Analysing studies published in the last thirty years on models of language competition, two facets can be distinguished. First, mathematical models of language competition, which were initially developed in the formal and natural sciences, have established themselves as a powerful tool for understanding language competition and language death, but have been slow in reaching the linguistic community. This can be seen in the fact that most models are published in journals covering the fields of physics and mathematics, but also in the fact that those models often lack a complete or accurate representation of the key processes affecting language competition, and in the fact that they are rarely validated against data. These assertions, however, seem to represent the reality of models of language competition less and less well. Since the last ten years, the way language competition is modelled has changed considerably. Models published recently typically consider a greater number of sociolinguistic parameters and are more often validated against data than before. This change is in part because more linguists and social scientists have started to show

**Table 5 Comparison of models characteristics during the periods pre-2010 and post-2010.**

| | Pre-2010 | | Post-2010 | |
|---|---|---|---|---|
| | Count | Percentage[a] | Count | Percentage[a] |
| Macroscopic approach | 15 | 44.1 | 18 | 60.0 |
| Microscopic approach | 8 | 23.5 | 10 | 33.3 |
| Bilingualism as separate state[b] | 10 | 29.4 | 11 | 36.7 |
| No. of geodemographic parameters | 90 | 2.6[c] | 65 | 2.2[c] |
| No. of sociolinguistic parameters | 36 | 1.1[c] | 70 | 2.3[c] |
| Based on data | 5 | 14.7 | 17 | 56.7 |
| Published in journals covering formal and natural sciences | 30 | 88.2 | 22 | 73.3 |
| Published in journals covering humanities and social sciences | 4 | 11.8 | 8 | 26.7 |
| Total number of models[d] | 34 | 100.0 | 30 | 100.0 |

[a]Percentages relate to the total number of studies in each group.
[b]Bilingualism is considered as a separate state when it forms an intermediate state in the shift from A to B. Thus: A -> AB -> B.
[c]Figures correspond to the mean rather than the percentage.
[d]Totals sum up to 64 since they concern the number of models rather than studies (see Table 2).

interest for models of language competition, harnessing their unique set of expertise, but also in part to the fact that formal and natural scientists themselves have started to adopt the concepts and terminology used by linguists. In conclusion, though a gap can still be noticed between disciplines concerning mathematical models of language competition, important steps have been made in recent years towards reducing it. A new wave of research on mathematical models of language competition is forming and is set to prove instrumental in informing efforts made towards preserving the world's linguistic diversity.

## Data availability
The list of all selected articles is provided in Tables S2 and S3 of the supplementary materials.

## References
Abrams DM, Strogatz SH(2003) Modelling the dynamics of language death. Nature 424(6951):900

Appel R, Muysken P (2005) Language contact and bilingualism. Amsterdam University Press

An Z, Pan Q, Yu G, Wang Z (2012) The spatial distribution of clusters and the formation of mixed languages in bilingual competition. Phys A 391 (20):4943–4952. https://doi.org/10.1016/j.physa.2012.05.023. Scopus

Baggs I, Freedman HI (1990) A mathematical model for the dynamics of interactions between a unilingual and a bilingual population: Persistence versus extinction. J Mathematical Soc 16(1):51–75. https://doi.org/10.1080/0022250X.1990.9990078

Baggs I, Freedman HI (1993) Can the speakers of a dominated language survive as unilinguals?: a mathematical model of bilingualism. Mathematical Comput Model 18(6):9–18

Bakalis E, Galani A (2012) Modeling language evolution: aromanian, an endangered language in Greece. Phys A 391(20):4963–4969. https://doi.org/10.1016/j.physa.2012.05.033. Scopus

Barreira da Silva Rocha A (2018) Social outcomes due to the interplay between language competition and ideology struggle. Phys A 492:1340–1351. https://doi.org/10.1016/j.physa.2017.11.061. Scopus

Barrett-Walker T, Plank MJ, Ka'ai-Mahuta R, Hikuroa D, James A (2020) Kia kaua te reo e rite ki te moa, ka ngaro: Do not let the language suffer the same fate as the Moa. J R Soc Interface 17(162). https://doi.org/10.1098/rsif.2019.0526

Beltran FS, Herrando S, Ferreres D, Adell M-A, Estreder V, Ruiz-Soler M (2009) Forecasting a language shift based on cellular automata. J Artificial Soc Soc Simulat 12(3):5

Bracken LJ, Oughton EA (2006) 'What do you mean?' The importance of language in developing interdisciplinary research. Trans Institute Br Geogr 31 (3):371–382

Bromham L (2017) Curiously the same: swapping tools between linguistics and evolutionary biology. Biol. Philos. 32:855–886

Caridi I, Nemiña F, Pinasco JP, Schiaffino P (2013) Schelling-voter model: an application to language competition. Chaos, Solitons and Fractals 56:216–221. https://doi.org/10.1016/j.chaos.2013.08.013

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

Castellano C, Fortunato S, Loreto V (2009) Statistical physics of social dynamics. Rev Modern Phys 81(2):591–646. https://doi.org/10.1103/RevModPhys.81.591

Castelló X, Loureiro-Porto L, Miguel MS (2013) Agent-based models of language competition. Int J Sociol Lang 221:21–51. https://doi.org/10.1515/ijsl-2013-0022. Scopus

Castelló X, Eguíluz VM, Miguel MS (2006) Ordering dynamics with two non-excluding options: bilingualism in language competition. New J Phys 8 (12):308–308. https://doi.org/10.1088/1367-2630/8/12/308

Civico M (2019) The dynamics of language minorities: evidence from an agent-based model of language contact. JASSS 22(4). https://doi.org/10.18564/jasss.4097

Clingingsmith D (2017) Are the world's languages consolidating? The dynamics and distribution of language populations. Economic J 127(599):143–176

Crystal D (2000) Language death. Ernst Klett Sprachen

De Oliveira VM, Campos PR, Gomes MA, Tsang IR (2006a) Bounded fitness landscapes and the evolution of the linguistic diversity. Phys A 368 (1):257–261

De Oliveira VM, Gomes MA, Tsang IR (2006b) Theoretical model for the evolution of the linguistic diversity. Phys A 361(1):361–370

De Oliveira PMC, Stauffer D, Lima FWS, Sousa AO, Schulze C, De Oliveira SM (2007) Bit-strings and other modifications of Viviane model for language competition Phys A 376:609–616

Dorian NC (1981) Language death: the life cycle of a Scottish Gaelic dialect. University of Pennsylvania Press

El-Owaidy HM, Ismail M (2002) A mathematical model of bilingualism. Applied Mathematics and Computation 131(2–3):415–432

Ethnologue (2020) Ethnologue: Languages of the World [WWW Document]. Ethnologue. https://www.ethnologue.com/ (Accessed 6.23.20)

Fernando C, Valijärvi R-L, Goldstein RA (2010) A model of the mechanisms of language extinction and revitalization strategies to save endangered languages. Human Biol 82(1):47–75

Fort J, Pérez-Losada J (2013) Front speed of language replacement. Human Biol 84 (6):755–772. https://doi.org/10.3378/027.084.0601. Scopus

Gong T, Shuai L, Zhang M (2014) Modelling language evolution: examples and predictions. Phys Life Rev 11(2):280–302

Grenoble LA, Whaley LJ (2006) Saving languages: an introduction to language revitalization. Cambridge University Press.

Grin F (1992) Towards a threshold theory of minority language survival. Kyklos 45 (1):69–97

Hadzibeganovic T, Stauffer D, Schulze C (2008) Boundary effects in a three-state modified voter model for languages. Phys A 387(13):3242–3252

Hale K, Krauss M, Watahomigie LJ, Yamamoto AY, Craig C, Jeanne LM, England NC (1992) Endangered languages. Language 68(1):1–42

Heinsalu E, Patriarca M, Léonard JL (2014) The role of bilinguals in language competition. Adv Complex Syst 17. https://doi.org/10.1142/S0219525914500039

Iriberri N, Uriarte J-R (2012) Minority language and the stability of bilingual equilibria. Rational Soc 24(4):442–462. https://doi.org/10.1177/1043463112453556

Isern N, Fort J (2014) Language extinction and linguistic fronts. J R Soc Interface 11(94). https://doi.org/10.1098/rsif.2014.0028

Kandler A, Unger R, Steele J (2010) Language shift, bilingualism and the future of Britain's Celtic languages. Philos Trans R Soc B 365(1559):3855–3864. https://doi.org/10.1098/rstb.2010.0051. Scopus

Kandler A (2009) Demography and language competition. Human Biol 81(3):181–210

Kandler A, Steele J (2008) Ecological models of language competition. Biol Theory 3(2):164–173

Kandler A, Unger R (2018) Modeling language shift. In: Diffusive spreading in nature, technology and society. Springer, pp. 351–373

Karjus A, Ehala M (2018) Testing an agent-based model of language choice on sociolinguistic survey data. Language Dynamics and Change 8(2):219–252. https://doi.org/10.1163/22105832-00802004

Kosmidis H, Argyrakis (2005) Language evolution and population dynamics in a system of two interacting species. Phys A 2005(353):595–612

Krauss M (1992) The world's languages in crisis. Language 68(1):4–10

Laruelle A, Navarro N, Escobedo R (2019) Proficiency, attitude, and conventions in minority languages. Sociol Methods Res 48(2):350–368. https://doi.org/10.1177/0049124116672679

Lipowska D, Lipowski A (2017) Language competition in a population of migrating agents. Phys Rev E 95(5):052308

Minett JW, Wang WS (2008) Modelling endangered languages: the effects of bilingualism and social structure. Lingua 118(1):19–45

Mira J, Paredes Á (2005) Interlinguistic similarity and language death dynamics. EPL 69(6):1031

Moher D, Liberati A, Tetzlaff J, Altman DG, Group P (2009) Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. PLoS Med 6(7):e1000097

Moseley C (2010) Atlas of the World's Languages in Danger. UNESCO

Mussa Juane M, Seoane LF, Muñuzuri AP, Mira J (2019) Urbanity and the dynamics of language shift in Galicia. Nat Commun 10(1). https://doi.org/10.1038/s41467-019-09688-8

Parshad RD, Bhowmick S, Chand V, Kumari N, Sinha N (2016) What is India speaking? Exploring the "Hinglish" invasion. Phys A 449:375–389. https://doi.org/10.1016/j.physa.2016.01.015. Scopus

Patriarca M, Heinsalu E (2009) Influence of geography on language competition. Phys A 388(2–3):174–186

Patriarca M, Leppänen T (2004) Modeling language competition. Phys A 338(1–2):296–299

Pinasco JP, Romanelli L (2006) Coexistence of languages is possible. Phys A 361(1):355–360

Prochazka, K., & Vogl, G. (2017). Quantifying the driving factors for language shift in a bilingual region. Proc Natl Acad Sci USA 114(17), 4365–4369. https://doi.org/10.1073/pnas.1617252114

Prochazka K, Vogl G (2018) Are languages like atoms? On modelling language spread as a physicist. Glottotheory 9:77–88. https://doi.org/10.1515/glot-2018-0005

Schulze C, Stauffer D (2005) Monte carlo simulation of the rise and the fall of languages. Int J Modern Phys C 16(05):781–787

Schulze C, Stauffer D (2006a) Computer simulation of language competition by physicists. In Econophysics and Sociophysics: Trends and Perspectives. Wiley, pp. 307–332

Schulze C, Stauffer D (2006b) Monte Carlo simulation of survival for minority languages. Adv Complex Syst 9(03):183–191

Schulze C, Stauffer D (2007) Competition of languages in the presence of a barrier. Phys A 379(2):661–664

Schwämmle, Queirós (2009) Competition and fragmentation: a simple model generating lognormal-like distributions. New J Phys 11(9):093006

Schwämmle V (2006) Phase transition in a sexual age-structured model of learning foreign languages. Int J Modern Phys C 17(01):103–111

Schwämmle Veit (2005) Simulation for competition of languages with an aging sexual population. Int J Modern Phys C 16(10):1519–1526

Scopus (2020) What are the most frequent Subject Area categories and classifications used in Scopus?: Access and use Support Center. https://service.elsevier.com/app/answers/detail/a_id/14882/supporthub/scopus/~/what-are-the-most-frequent-subject-area-categories-and-classifications-used-in/

Seifart F, Evans N, Hammarström H, Levinson SC (2018) Language documentation twenty-five years on. Language 94:e324–e345

Seoane LF, Loredo X, Monteagudo H, Mira J (2019) Is the coexistence of Catalan and Spanish possible in Catalonia? Pal Commun 5(1):1–9

Seoane LF, Mira J (2017) Modeling the life and death of competing languages from a physical and mathematical perspective. Preprint at http://arxiv.org/abs/1703.10706

Solé RV, Corominas-Murtra B, Fortuny J (2010) Diversity, competition, extinction: the ecophysics of language change. J R Soc Interface 7(53):1647–1664

Stauffer D, Castelló X, Eguiluz VM, San Miguel M (2007) Microscopic Abrams–Strogatz model of language competition. Phys A 374(2):835–842

Stauffer D, Schulze C (2005) Microscopic and macroscopic simulation of competition between languages. Phys Life Rev 2(2):89–116

Stauffer D, Schulze C, Lima FWS, Wichmann S, Solomon S (2006) Non-equilibrium and irreversible simulation of competition among languages. Phys A 371(2):719–724

Tanaka CM, Lee J-H, Iwasa Y (2018) The persistence of a local dialect when a national standard language is present: an evolutionary dynamics model of cultural diversity. Bullet Mathematical Biol 80(10):2761–2786. https://doi.org/10.1007/s11538-018-0487-2. Scopus

Templin T (2019) A language competition model for new minorities. Rational Soc 31(1):40–69. https://doi.org/10.1177/1043463118787487. Scopus

Templin T, Seidl A, Wickström B-A, Feichtinger G (2016) Optimal language policy for the preservation of a minority language. Mathematical Soc Sci 81:8–21. https://doi.org/10.1016/j.mathsocsci.2016.03.006

Teşileanu T, Meyer-Ortmanns H (2006) Competition of languages and their hamming distance. Int J Modern Phys C 17(02):259–278

Thomason SG (2015) Endangered languages. Cambridge University Press

Tuncay Ç (2007) A new model for competition between many languages. Int J Modern Phys C 18(07):1203–1208. https://doi.org/10.1142/S0129183107011236

Vogt P (2009) Modeling interactions between language evolution and demography. Human Biol 81(3):237–258

Wang WS, Minett JW (2005) The invasion of language: emergence, change and death. Trends Ecol Evol 20(5):263–269

Wickstrom B-A (2005) Can bilingualism be dynamically stable? A simple model of language choice. Rational Soc 17(1):81–115

Wyburn J (2018) Media pressures on welsh language preservation. J Mathematical Sociol 42(1):37–46. https://doi.org/10.1080/0022250X.2017.1396984

Wyburn J (2019) The need for language planning to address English-language media pressures on minority language survival in bilingual populations. J Operational Res Soc https://doi.org/10.1080/01605682.2019.1609880

Wyburn J, Hayward J (2010) A model of language-group interaction and evolution including language acquisition planning. J Mathematical Soc 34(3):167–200. https://doi.org/10.1080/0022250X.2010.485706

Wyburn J, Hayward J (2008) The future of bilingualism: an application of the Baggs and Freedman model. J Mathematical Soc 32(4):267–284

Wyburn J, Hayward J (2009) OR and language planning: modelling the interaction between unilingual and bilingual populations. J Operational Res Soc 60(5):626–636

Wyburn J, Hayward J (2019) An application of an analogue of the partition function to the evolution of diglossia. Phys A 516:447–463. https://doi.org/10.1016/j.physa.2018.10.047

Zhang M, Gong T (2013) Principles of parametric estimation in modeling language competition. Proc Natl Acad Sci USA 110(24), 9698–9703. https://doi.org/10.1073/pnas.1303108110

Zhou Z, Szymanski BK, Gao J (2020) Modeling competitive evolution of multiple languages. PLoS ONE 15(5). https://doi.org/10.1371/journal.pone.0232888

## Competing interests

The authors declare no competing interests.

## Additional information

**Correspondence** and requests for materials should be addressed to M.B.

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.