



HUMBOLDT-UNIVERSITÄT ZU BERLIN

REAL-ARBEITSBERICHT NR. 16-2

**Modelling language competition for
new minorities**

TORSTEN TEMPLIN

Updated: 2016-11-15

REAL

Research group “Economics and Language”
Forschungsgruppe »Ökonomie und Sprache«

Modelling language competition for new minorities

TORSTEN TEMPLIN

Humboldt-Universität zu Berlin

Modelling Language Competition for New Minorities

-

Working paper

Torsten Templin*¹

¹*Research Group Economics and Language (REAL)
Department of Education Studies
Humboldt-Universität zu Berlin*

Abstract

In this paper we model a language competition situation between a dominant majority language H and a minority language L , where the minority language group is a result of a continuous and ongoing migration process. Migrants enter the society as monolinguals in L or already as bilinguals, form families and produce offspring. Offspring are raised in either one of the two languages or both, are formally educated primarily in the dominant language and will produce a next cohort with its own linguistic repertoire. Furthermore, adult language learning is considered. Since properly designed language competition models allow predictions on the future development of the linguistic composition of a population, we scrutinize the possibility of using language competition models as tools for *ex ante* policy evaluation.

1 Introduction

Language competition models are formal mathematical models that describe the “competition” between two or more language communities. More precisely, it is modeled how the (relative) size of certain language groups or the geographical distribution of certain languages changes over time within a given territory. Therefore, we also use the term ‘language dynamics models’. In the past two decades a growing number of such models were developed and analyzed. They borrow techniques and ideas from biology, chemistry, physics and economics, and use respective metaphors to describe and simulate the (aggregated) linguistic behavior of human beings.¹

Many language competition models are concerned with processes of language decline and language death. This is motivated by the observation that “[a]ll over the world, people are stopping speaking minority languages and shifting to languages of wider communication” (Sallabank, 2012, p. 104). It is predicted that 90 percent of the currently 7000 spoken languages will not survive the end of the century (Krauss, 1992). Based on quantitative (historical) data on specific cases of language competition or, more

*Research supported by the European Community’s Seventh Framework Program under grant agreement No. 613344 (Project MIME).

¹E.g. “[L]anguage competition resembles the competing relation in ecology, where the rise or decline of the population size of a species is influenced by the growth rate of the competing species. This competing relation exists not only between predators and preys, but is common among various species in the biological world” (Zhang & Gong, 2013, p. 9699).

globally, on the changing total number of actively spoken languages, researchers try to develop models that are able to simulate similar data. “The hope is that as more and more quantifiable relations in and among languages are discovered and simulation models are developed which can adequately replicate these distributions, the simulation models will of necessity become more and more adequate as models of actual languages [...]” (Schulze *et al.*, 2008, p. 15).

In contrast to the majority of language competition models that are concerned with the extinction and revitalization of languages historically rooted in specific regions, we here focus on minority languages (from the host country point of view) mostly spoken by “newcomers” or their descendants. According to the United Nations (2013) there were 232 million international migrants worldwide, and in the global north migrants make up 10.8% of the total population. Take the European Union as an example. Within the Union with its currently 24 official languages citizens of EU member states can easily move to and work in other member states. In view of high unemployment rates in some member states and demand for qualified staff in others, (workforce) mobility becomes even more attractive. Moreover, of course, there is migration from outside the EU². Such mobility flows within the EU and from outside the EU have an impact on the linguistic composition of single member states. Therefore, it needs adequate models that can simulate the medium- and long-term linguistic effects of international mobility.

We confine ourselves to states or regions with one official or dominant language facing an ongoing influx of speakers of another language, e.g. the official language of another state or region. Examples could be Spanish in the US, Turkish or Arabic in EU member states or the dominant language of one European state in another European state with a different dominant language. Throughout the paper we call the locally dominant (high status) language H and the (low status) language of the newcomers L . The model developed in this paper describes how the number of speakers of H and L change over time. One mechanism that drives the language dynamics and that is explicitly modeled is external mobility, i.e. the influx of new individuals with certain language repertoires. A second mechanism that affects the evolution of the linguistic composition of the population is intergenerational language transmission. Individuals form couples and families, have children and transmit one or more of the languages they speak to their children. For these processes *emotional* and *instrumental* motives as well as external conditions such as linguistic concentration/segregation play an important role and are included in the model. A third mechanism taken into account is language learning. On the one hand, young newcomers or children of newcomers enjoy formal education in one of the two languages or both – in practice predominantly in H – and might become bilingual or even monolingual in H . On the other hand, adult newcomers might acquire the locally dominant language over the years to improve their social and economic opportunities. The model proposed here comprises these three mechanisms to describe the language dynamics of a population that is constantly changing.

A second aspect in which the approach taken here differs from some of the work in the literature is that we do not aim at explaining the ‘how and why’ of language dynamics, e.g. by identifying abstract principles that might explain long-term developments. Instead, we build on the extensive research of socio-linguists and other social scientists to design a preferably informed and realistic model of processes at hand. To do so, we review literature from different fields on language transmission, language acquisition and language learning and identify important factors and variables that drive the language dynamics and should thus be modeled. The aim of our research is to present a model design that enables predictions of the nearer linguistic future of a population facing mobility. Above that, the model shall be able to simulate how this future can be altered through a change of relevant model parameters. For this to be meaningful, such parameters cannot just be abstract and rather arbitrary numbers, but have to correspond to measurable quantities. The hope of our research is that adequate models can be used to simulate

²“A total of 3.4 million people immigrated to one of the EU-28 Member States during 2013 [...]. Among these 3.4 million immigrants during 2013 there were an estimated 1.4 million citizens of non-member countries [...]” Eurostat (2014).

language planning policies and their effects on language dynamics. If this is possible, then language competition models could function as formal tools to ex ante evaluate language policies.³

The paper is organized as follows. In the remainder of Section 1 we give a short overview of the language competition models available in the literature and discuss how far they can be used to simulate meaningfully the effects that changes in the linguistic environment have on the language dynamics. In the first two parts of Section 2 the ‘the number of speakers’ – as the central object of interest – and the conceptualization of the dynamics are discussed. Thereafter, literature on language transmission and acquisition is reviewed in some detail and the operationalization of the linguistic environment is outlined. At the end of this section, population dynamics and time scales are discussed. In Section 3 the general model formulation is presented and specified in Section 4. Section 5 provides an analysis of the steady states of the model, while in Section 6 interventions are discussed. In Section 7 some numerical examples shall illustrate the language dynamics graphically and show how a change in model parameters affects medium- and long-term outcomes. The conclusions are provided in Section 8.

1.1 Language Competition Models: A Short Overview

One important point of departure was a short paper published in 2003: “It was the population dynamics model of Abrams and Strogatz [...] which started the avalanche of physics papers on language competition” (Schulze *et al.*, 2008, p. 13). In their “minimal model” two languages compete for monolingual speakers. The attractiveness of a language is determined by the number of its speakers as well as its “perceived status”.⁴ Hence, it is assumed that the higher the number of speakers of a language and the higher its status, the more speakers it attracts. These two assumptions are formalized by a differential equation that describes the language dynamics.

On the one hand, Abrams and Strogatz could fit their “minimal model” to aggregated empirical data of endangered languages.⁵ On the other hand, the model is quite simplistic, does not really account for the underlying social and linguistic processes of language acquisition, language transmission and language change, neglects bilinguals and predicts that always one of the two competing languages will become extinct in the long run.

Due to such shortcomings the Abrams-Strogatz (AS) model was revised and extended by many authors, especially from the field of (statistical) physics. Patriarca & Leppänen (2004) and Patriarca & Heinsalu (2009) include spatial components in their adaptations of the AS model. They take geographical inhomogeneities into account and could show that it is possible that both languages survive in two different geographical regions. Mira & Paredes (2005) introduced the idea of similarity between competing languages to the original model and could prove that both languages can survive if they are close to each other. Stauffer *et al.* (2007) propose microscopic or individual based versions of the AS model and apply simulation techniques instead of averaging over the whole population.

Furthermore, Mira & Paredes (2005), Wickström (2005), Minett & Wang (2008), Heinsalu *et al.* (2014) and others extended the A-S model by additionally considering bilinguals. The role of bilinguals in language shift is crucial, since in practice individuals seldomly change from being monolingual in one language to being monolingual in the other language. Instead, it is far more realistic to consider transitions

³“[I]n my view, the chief usefulness of an economic perspective on language probably is not so much that it helps to understand language-related processes *as such*. [...] The chief usefulness of language economics is derived from its capacity to formulate, evaluate and compare policy options regarding language, and hence to assist in decision-making” (Grin, 2003, p. 5, emphasis in original).

⁴“[A] parameter that reflects the social or economic opportunities afforded to its speakers”.

⁵They investigated four cases studies: Scottish Gaelic in Sutherland, Scotland; Quechua in Huanuco, Peru; Welsh in Monmouthshire, Wales; Welsh in all of Wales.

from monolinguals in one language to bilinguals and - in a second step - from bilinguals to monolinguals in the other language. For such a two step transition it normally takes more than one generation. For that reason, some authors explicitly modeled vertical, i.e. intergenerational, language transmission. Minnett & Wang (2008) develop a uni-parental model for vertical language transmission. Children of monolingual parents are raised monolingual, while bilingual parents raise their children either as monolinguals or bilinguals, depending on the distribution of speakers and the statuses of the languages. In contrast, Wickström (2005), Fernando *et al.* (2010), Templin *et al.* (2016) and others use a two-parents model. Here adult individuals (randomly) form families and bring up their children in either one of the parents languages or in both of the parents languages. Based on a very general utility maximization framework for individual families Wickström (2005) proved general properties of intergenerational language transmission.

An other feature of the model developed in Minnett & Wang (2008) is horizontal language transmission. During their lifetime monolingual adults can learn a second language and therefore become bilingual. The incentive to learn a second language increases with the number of speakers of that language as with the status variable.

In contrast to Abrams and Strogatz, Pinasco & Romanelli (2006) proposed a Lotka-Volterra type model⁶ inspired by population dynamics to model language competition and also show the possibility of coexistence. Spatial extensions of this model can be found in Kandler & Steele (2008) and Kandler *et al.* (2010). Zhang & Gong (2013) also use a Lotka-Volterra type model, but avoid using a “abstract [status] parameter [that] lacked explicit sociocultural meanings” (Zhang & Gong, 2013, 9698). Instead, they introduce a language diffusion principle⁷ and two language inheritance principles⁸ to estimate model parameters. To fit their model to historical data, the authors introduce a (linear) time transformation of the original model, but do not really theorize this transformation. In the end, they avoid using the status parameter, but get a new time-transformation parameter, which is again hard to interpret.

A third strand of research takes an economic approach to model language dynamics. One of the first dynamic language economics models was presented in Grin (1992). Grin considers two languages. Bilingual agents can conduct activities in both languages and gain utility from them. Activities require the investment of time and goods, while both are limited. The model can be characterized as economic, since “language use is analyzed as an object of choice by individuals who maximize a utility function” (Grin, 1992, p. 74). Individual utility maximizing choices at time t affect choices at time $t + 1$ due to network effects: the more the minority language is used at time t , the more attractive is the minority language at the next point in time. These network effects drive the dynamics. A second model to be mentioned here was developed in John & Yi (1996). There, a two period model for two monolingual regions with different languages is developed. In every period individuals can either learn the other language or engage in production. Between the periods, they are able to move to the other region and engage in production there. The authors assume network externalities, which connect one individuals decision to all other individuals: “per capita production of an agent is (weakly) increasing in the number

⁶A set of differential equations that are used in evolutionary biology to describe the dynamics of biological systems in which two species compete with each other, normally one as a predator and one as a prey.

⁷“People are language carriers; language diffuses along with the diffusion of the population [...]” (Zhang & Gong, 2013, p. 9699). People diffuse from a center with a high population density to regions with lower population densities. With them the language of the center diffuses. The further away a region is from the center, the lower is the impact of this center-language. The mathematical principle that models this diffusion is inspired by Fourier’s law of heat conduction.

⁸The “inheritance rate of a language reflects the inheritance capacity of this language during learning” (Zhang & Gong, 2013, 9698). The concept of inheritance capacities is borrowed from genetics. The first inheritance principle roughly states that if language learning is random, then the populations speaking different languages remain constant across generations. Therefore, the inheritance rates depend on the sizes of the three language repertoire groups. To apply the model to real-life situations information on the number of speakers is needed. If such data are not available, then the second more abstract inheritance principle can be used. This second principle is inspired by lexical diffusion dynamics in computational linguistics.

who speak the same language” (John & Yi, 1996, p. 4). A dynamic version of the two-period model with successive generations can be found in John & Yi (2001). In a similar fashion, Tamura (2001) considers two regions with monolingual populations. Individuals do not choose whether or not to learn the other language. Instead, parents decide on spending time to raise their children as bilinguals, which might have a positive impact on the child’s income, or to spend their time on consumption. Since time is a scarce resource, parents have to make an optimal decision. Another economic approach is presented in Kennedy & King (2005). They construct a three generation model with overlapping generations, where adults vote for public language policies. Two languages are considered and people are either monolinguals or bilinguals. As before, for production adults profit from other adults speaking their language(s) and are hence interested in educational policies that teach their language to children. At the same time education is costly, such that adults have to weight costs against benefits. The voting decisions then drive the language dynamics. A last set of models to be mentioned here applies a game theoretic approach. Although they are no actual dynamic models (they do not describe how the linguistic composition changes over time), they inspired many of the previously listed economic models. In 1991 Selten and Pool published a model that can account for an arbitrary number of languages. They assume that the distribution of native languages is fixed and that individuals can learn additional languages. Importantly, not all of the languages considered by the model have to have native speakers. Individuals (or players in the game theoretic jargon) gain communicative benefit from knowing additional languages while learning them produces costs. The communicative benefit a player gains from a language depends on the number of speakers of that language weighted by an “importance function”. “The importance function is intended to reflect the fact that economic, geographical, and other attributes of a person affect the value of being able to communicate with this person” (Selten & Pool, 1991, p. 67). The authors show that for this multilingual setting an equilibrium exists. Adoptions, extensions and applications of the Selten-Pool model or comparable game-theoretical approaches can, for example, be found in Church & King (1993), Ginsburgh *et al.* (2007), Gabszewicz *et al.* (2011) and Iriberry & Uriarte (2012).

For more extensive overviews of most of the papers listed above see Patriarca *et al.* (2012), Gong *et al.* (2014) and John (2016).

1.2 Phenomenological Versus Mechanistic Models

We are interested in a model design that enables us to simulate the language dynamics of a population given certain information on the linguistic environment. Since certain aspects of the environment can be seen as a result of language policies, e.g. language(s) of instruction in schools or languages used in public administration, such a model design could help policy makers to compare and evaluate different policy options *ex ante*.

Some of the existing language competition models already deal with language policies, especially with revitalization policies for endangered minority languages. Even Abrams and Strogatz express the hope that their minimal model “may be useful in the design and evaluation of language-preservation programs” (Abrams & Strogatz, 2003, p. 900). Minett & Wang (2008) go further and introduce a simple policy to illustrate that with state intervention the minority language can be maintained. They “assume that a community can bring about a change in the value” of relevant model parameters whenever the number of speakers of the minority language falls below some threshold. Due to such a change of parameters, the numbers increase again and the minority language is saved from extinction.⁹ However, it is not quite clear

⁹A similar notion can, for example, be found in Heinsalu *et al.* (2014): “Importantly, specific linguistic policies aimed at defending and revitalizing minority languages can be described by a change in the values of α and β ” (p. 14). These parameters “may be interpreted as the importance of the bilinguals as the representatives of Y, X to the monolinguals of X, Y respectively” (p. 11), where X and Y denote the two competing languages.

what these parameters are and hence what such an intervention actually means. The parameters indicate the “peak rate” at which individuals with one linguistic repertoire switch to another repertoire, which “can reflect, variously, the rate of contact between pairs of individuals, the propensity for individuals to learn a new language based on their existing linguistic skills, or the provision of language resources to children” (Minett & Wang, 2008, 22).

As already pointed out by (Fernando *et al.*, 2010, p. 50), these parameters can not be measured directly and hence their values can not be justified. Therefore, neither the model of Abrams and Stogatz nor the more sophisticated model by Minett and Wang can be used realistically to simulate language policies. This is the problem with most available language competition models, where some rather artificial parameters are used to describe language dynamics. Even if such parameters are estimated through a fit of the model to empirical data, there is no proper interpretation for their concrete values. This is especially true for a parameter present in almost every model, called the status of a language or prestige of a language. Here, Fernando *et al.* ask “what were the characteristics of a language having a prestige value, say 1.2, and what was the sociocultural condition corresponding to the difference between two languages having prestige values, say 1.2 and 1.3, respectively (Fernando *et al.*, 2010, p. 50)”. On the basis of such vague concepts as the status, models like the one by Abrams and Stogatz are “unable to suggest anything more explicit [to save endangered languages] than the obvious advantage of increasing the status of the low-status language” (Fernando *et al.*, 2010, p. 50). A similar criticism is expressed in Zhang & Gong (2013, p. 9698): “However, this abstract [status] parameter lacked explicit sociocultural meanings”. Fernando *et al.* (2010) call models that are based on parameters whose values are arbitrary or just obtained by model fits, but lack explicit meaning, “phenomenological” models. Instead, they promote a more “mechanistic” approach, which means to develop models “characterized by parameters that are measurable in the field” (p. 52) and therefore have “predictive power” (p. 49). Since we are interested in models that allow more or less realistic evaluations of language policies, we follow Fernando *et al.* in their call for “mechanistic” conceptions.

2 Towards a *Mechanistic* Model

This section is devoted to a number of central aspects of language competition models and how they are dealt with in the literature. The aim of the following deliberations is twofold. On the one hand, we would like to make the modeling approach more accessible for scholars that are not familiar with all the mathematical concepts used in language competition models. In our opinion we rely on the expertise and research of scholars from various fields to design informed and *realistic* models. Hence, it should be made possible at least to assess the underlying theoretical assumptions incorporated in the mathematical models. Enhancing interdisciplinary exchange could furthermore increase the influence of the modeling approach in the social sciences. On the other hand, we want to give a more detailed overview on how central aspects of language dynamics are approached differently in the literature. This examination yields the foundation for the design of our own model.

We start with ‘the number of speakers’ as the central object of interest and how it is operationalized. Thereafter, we discuss how the dynamics are conceptualized. As the main drivers for language dynamics, processes of language transmission and acquisition are discussed in greater detail. Furthermore, we outline how population dynamics and age structure are taken into account. At last we discuss time scales and time frames in language competition models.

2.1 Numbers of Speakers

Language competition models are concerned with the distribution of certain language repertoires throughout a population or a given territory. The vast majority of models available in the literature consider two languages, a high status language H and low status language L . Hence, the first step in the development of a language competition model is to specify the different language repertoires¹⁰ that are taken into account. In the following we will use the term language repertoire group (LRG), which is defined as the collection of individuals having the same language repertoire (LR).¹¹ Every individual within the population has to be assigned to only one language repertoire group. As in most language competition models we consider three language repertoire groups: monolinguals in L , monolinguals in H and bilinguals.

At a first glance it might seem straight forward for non-linguists to assign every individual to one of those three groups. But the “question of what it means to know a second language turns out, like all good questions, to have many possible answers” (Spolsky, 1989, p. 140). For example, what about an individual that grew up in language H and acquired some of language L later on in her life? How well does she have to speak L to count as bilingual? Is a basic knowledge of L sufficient, or does she have to be fluent in L ? These questions indicate that the task of assigning individuals to language repertoire groups it is not an easy one. In principle, one could introduce a whole range of different language repertoire groups, including e.g. individuals speaking only one language, individuals fluent in one language with some basic knowledge of the other language and individuals fluent in both languages.

For simplicity and mathematical tractability, we restrict ourselves here to the three language repertoire groups. How individuals are grouped together then largely depends on the underlying definitions of monolingualism and bilingualism, and there are various ways to define these concepts, see e.g. (Baker, 2011, pp. 2-17). Most of the papers that provide mathematical models for language competition do not explicitly discuss this issue.¹² In contrast, Wickström (2005) and Fernando *et al.* (2010) provide two possible conceptions of bilingualism. To be a bilingual in the sense of Wickström (2005), one has to be raised bilingually: “By monoglot we understand a native speaker of only one language. Such a person might be able to communicate very well in the other language, but not with the ease or comfort of a native speaker. [...] A bilingual person is, hence, someone who is fully socialized into both language groups” (p. 83). According to (Fernando *et al.*, 2010, p. 53), however, bilingualism is defined as “the ability to function confidently in two languages, that is, the ability to have communicative competence in two languages” (p. 53). Using such a definition, Fernando *et al.* allow language acquisition outside the family. Although Minett & Wang (2008) do not provide any explicit definition of bilinguals, it is apparent that they have a similar conception in mind, since it is the case that languages can be learned later in life as well. Since we emphasize the role of language as a tool for communication and as human capital and want to take different modes of language acquisition into account (see below), a functional understanding of proficiency appears to be the most appropriate. We thus follow the definition of bilingualism given by Fernando *et al.* (2010).¹³

¹⁰Here, we use the term ‘linguistic repertoire’ in the very narrow sense as the set of languages a person commands, while only the two languages H and L are taken into account, see below.

¹¹Note that this definition neither implies nor precludes any kind of group identity.

¹²Exceptions are e.g. Grin (1992) or John & Yi (1996).

¹³For the development of language dynamics models and hence for the current paper such a definition is sufficient. Unfortunately, if one would like to use the model to analyze a real-life language competition situation, then linguistic competences would have to be determined throughout a larger population, which requires extensive language testing and is hence expensive and tedious. Therefore, for any real-life application of the model we mostly rely on information compiled through quantitative surveys or studies as ‘censuses’. In such surveys we can find self-reported data on the level of language proficiency. Although self-reported data are by far not ideal to reliably access linguistic competences, see (Baker, 2011, Chapter 2) for example, in most cases these are the only large-scale data available. Hence, for applications we would call an individual a speaker of a language if she reports to speak/know the language very well.

Now we can define the central dependent variables (or state variables) of the language competition model. In the literature we find two different approaches: microscopic models and macroscopic models, cf. Patriarca *et al.* (2012). In a microscopic model a finite number of individuals is considered. These individuals form a network and are connected either to all other individuals within the network or only to some of their neighbors, see e.g. Stauffer *et al.* (2007). In microscopic models the language repertoire of every individual is monitored. In contrast, macroscopic models only consider the sizes of the different LRGs and, in some cases, their geographical distribution. To illustrate the macroscopic approach we introduce some notation that is used throughout the paper. Let X_H denote the fraction of H -monolinguals of the population, X_L denote the fraction of L -monolinguals and X_B denote the fraction of bilinguals. Note that $X_B = 1 - X_H - X_L$, i.e. that all language repertoire groups together make up the whole population. Hence, the vector $X := (X_H, X_L)$, which is the central dependent variable in our model, fully characterizes the linguistic composition of the population. A macroscopic model – as proposed in this paper – is only concerned with the linguistic composition of the population as a whole, not with the language repertoires of single individuals. One reason for choosing a macroscopic over a microscopic approach is that the linguistic composition X can be obtained from quantitative surveys. Although such surveys also provide information on the language repertoires of a number of single individuals, normally these individuals are not directly connected to each other.

2.2 Dynamics

Language competition models are concerned with how the linguistic composition of a population changes over time. Models make (probabilistic) statements about the composition of “tomorrow”, given today’s composition as well as other relevant information. In formal terms this reads as follows. Let t denote time. We write $X_{LR}(t)$ for the fraction of the population having language repertoire LR , $LR = H, L$ or B , at time t . Given $X(t) := (X_H(t), X_L(t))$ and other relevant information, a language dynamics model makes statement about $X(t + 1)$.¹⁴ This subsection aims at explaining the general idea of how the dynamics are modeled.

The evolution of the linguistic composition can be seen as an aggregated result of individual language related behavior or decisions. This includes the communication of parents with their children, decisions on schools and foreign languages in formal education, adult language learning, language use in multilingual contexts and consumption of cultural products, to name some important aspects. Every such individual behavior is framed by a given linguistic environment, which includes the linguistic composition of the population.¹⁵ Take language acquisition in formal education as an example. Which languages are learned depends, among other factors, on the language(s) of instruction, on which foreign languages are offered and on how schools deal with speakers of minority languages. At the same time, individuals actively shape the linguistic environment through their behavior. To put it differently, individual language related behavior is not isolated but affected - at least to some extent - by the behavior of the rest of the population. For example, the decision of a monolingual speaker of a minority language on learning the majority language might be different in a situation where 30% of the population speaks her language

¹⁴Gong *et al.* (2014) classify models either as *rule-based* or as *equation-based*. Providing formulas for $X(t + 1)$ based on $X(t)$ and other relevant information refers to the *equation-based* approach, on which we concentrate in this paper. Furthermore, as outlined in the previous section, some models additionally have a spatial component. Accordingly, such models make statements about the next point in time ($t + 1$) for every point in space.

¹⁵Linguistic environment is a “[t]heoretical construct used for analytical purposes. It subsumes in an extensive (but obviously not exhaustive) fashion all the relevant information about the status, in the broadest sense of the word, of the various languages present in a given polity at a certain time. This includes the number of speakers, individual proficiency levels in the various languages, the domains of use of each language by different types of actors (individuals, corporations, the state, civil society organizations), and their attitudes towards the languages considered” (Grin & Vaillancourt, 1997, p. 49).

compared with a situation where only 1% speaks the minority language. We therefore speak of network effects or network externalities.

In short, we conceptualize language dynamics as follows. At each point in time individuals in a population show some linguistic behavior or, respectively, make some language related decisions. This behavior is affected by the linguistic environment. The linguistic behavior of the population, understood as aggregated individual decisions, may change the linguistic environment at the next point in time. This influence of today's behavior on the behavior of tomorrow reflects the network effects.

To derive an explicit model we start with the individual level. We need to clarify which behavior is considered, which elements of the linguistic environment are taken into account and how individual behavior is affected by the environment. As a result we obtain models for individual behavior. Next, individual behavior is aggregated and properties of the aggregated behavior are derived.¹⁶ Last, a (simple) mathematical model that satisfies these properties is proposed and analyzed. Hence, we use a methodological individualistic approach to model language competition.

2.3 Individual Level

The linguistic composition of a population, e.g. the inhabitants or citizens of an EU member state, evolves due to a number of complex processes or mechanisms. As every model builds on a necessary simplification of the complex reality, we restrict ourselves to four important ones. First, individuals enter or leave the population. On the one hand, children are born and people die. On the other hand, individuals from other states move to the state of consideration, or vice versa. Second, children acquire one or several languages mostly within the family context. Third, pupils are educated in a certain language and can learn additional languages in formal education. Fourth, adult individuals can learn new languages, improve their skills in a language, or can forget a language they do not use regularly.¹⁷ For the moment we neglect external mobility and take a closer look at language acquisition and learning.

In the literature we find two different but related ways of how these three modes of language acquisition are conceptualized. One could call them an 'economic conception' and a 'sociological conception'. We see the former represented in Church & King (1993), Tamura (2001) or Wickström (2005), while the latter can be found Fernando *et al.* (2010).¹⁸ What characterizes the former models as "economic" is that individuals are conceptualized as (rational) actors that evaluate different options according to a cost-benefit or cost-utility analysis. The three models differ in the actors they consider as well as in their operationalization of utility. While Church and King consider individuals, which are already equipped with a certain language repertoire, deciding on whether or not to learn a second language, Tamura and Wickström consider parents that decide on a language repertoire in which to socialize their children in. In Fernando *et al.* (2010) language acquisition within the family, as well as language learning in formal education, are modeled. For language acquisition at home their focus is not on the parents' decision, but

¹⁶The easiest way to illustrate the aggregation is to use a probabilistic framework. Assume that for a given environment the probability that a L -monolingual learns language H is given by p . Thus, the probability that she remains monolingual is $(1 - p)$. Then, on average, a fraction p of all B -monoglots learn H , while a fraction $(1 - p)$ remains monolingual. A different but compatible framework is used in Wickström (2005). The author assumes that actors choose the option, which yields the highest utility for them. Deriving properties of the aggregated level is than 'just' a matter of counting individuals that choose the same option.

¹⁷For research on different channels of acquisition of second and third languages see e.g. Gazzola *et al.* (2016, pp. 6ff), citing results from a quantitative study presented in Grin (1999).

¹⁸"Both sociologists and economists have studied immigrants' language proficiency [...]. Sociologists have mainly focused on the impact of immigrants' exposure to the language and social opportunities for language learning [...]. Economists, on the other hand, have noted the importance of immigrants' self-selection and the difficulty of learning a language" Van Tubergen & Kalmijn (2005a).

on the child that hears conversations in both languages at home as well as outside the home. Roughly, it is the amount of exposure to a language that determines whether or not the child will acquire that language successfully. Since part of the exposure occurs within the family, parents do play a role. But – and this is the fundamental difference from the “economic” approaches – parents behavior is not modeled explicitly as conscious utility-maximizing decision making. In the following paragraphs we review research on the three modes of acquisition to work out the strengths and weaknesses of the two approaches. Thereafter, we propose a combination of both conceptions.

2.3.1 Language Acquisition in the Family

Although in some European member states up to 1 out of 4 children lives in a single parent household, the majority of children is raised in families with two parents, see Ruggeri & Bird (2014). For this reason, we restrict ourselves to language acquisition within families with two parents and explicitly model family formation. As in Wickström (2005) and Fernando *et al.* (2010) we do not differentiate between the language repertoire of the mother and the one of the father and hence get six possible family types F ($F = HH, HB, LL, LB, BB, HL$). Family formation is assumed to be the result of a random search and mating process: adult individuals meet randomly, some form couples and some have children. While Wickström (2005) assumes a low success probability for HL couples, i.e. couples of a monolingual speaker of H and a monolingual speaker of L , Fernando *et al.* (2010) neglect this unlikely family type, i.e. assume success probability 0. This shall reflect the (rather realistic) assumption the existence of couples where both partners are not able to communicate with each other is rather unlikely (if one of the partners acquires the other language, then the couple is not HL anymore). Here, we follow Fernando *et al.* (2010). For the mathematical details on how family formation is modeled see Section 3.

Following family formation we need to model intergenerational language transmission or language acquisition within the family. In Tamura (2001) parents can use their spare time to teach their children a second language or to do something else. Both using spare time for something else and acquiring a better economic perspective for their children (due to bilingualism) yield utility for the parents. Recall, in Wickström (2005) parents also choose the language repertoire LR for their children, which yields the highest individual utility. For Wickström it is not the free time parents are weighting against their children’s economic future. Instead, they want to pass cultural values attached to their heritage language(s) to the next generation. If the language that promises economic opportunities and the heritage language are not identical, then there is a similar trade-off as the one between consumption (of free time) and production (of bilingual children) analyzed in Tamura (2001). In Wickström (2005), the more people who speak the language and the higher the status of that language, the higher is the individual utility one gains from that language. Counting the number of all F -type families choosing language repertoire LR yields $q_{LR}(F)$, the fraction of F -type families bringing up their children in language repertoire LR .¹⁹ In Fernando *et al.* (2010) a similar expression is used, but there it is interpreted as the probability that a child growing up in a F -type family develops language repertoire LR . This probability depends on the exposure to LR ; for $LR = B$ it depends on the exposure to H and L . This exposure has two components: family and community. The exposure within the family context depends on the family type. In Fernando *et al.* (2010) it is assumed that in HH families only H is spoken and in LL families only L . Furthermore, they make the simplifying assumption that in HB families language H is spoken five-sixths of the time and respectively that L is used five-sixths of the time in LB families. In BB families both languages are spoken half of the time. In addition to the family context, children are exposed to conversations heard in public. This public exposure depends on the linguistic composition of the community (i.e. population) as well as the prominence of the language in mass media.

¹⁹In Wickström (2005) this fraction is denoted by $\alpha_{LR}(F)$. Since α_{LR} is a different model parameter in Fernando *et al.* (2010), we use the notation $q_{LR}(F)$.

Since parents language decisions and the child's exposure are connected to one another, both approaches share some common features. If parents decide that their children shall acquire a certain language, they might use that language in communication with the child on a regular basis and hence provide a certain amount of exposure. Later we argue that a child's exposure within the family context is partially determined by parent's decisions. In both approaches $q_{LR}(F)$ depends on the family type F and on the linguistic composition of the population. Additionally, in Wickström (2005) the function q_{LR} , $LR = H, L$, increases with the status of H , while the model in Fernando *et al.* (2010) does not contain an explicit status variable. Instead, a parameter called "susceptibility" is introduced. The susceptibility towards L is defined as the "responsiveness of a child born to HH or HB families to speak L as a function of conversation heard" (p. 56) or as "effectiveness of hearing language [B] in motivating its learning" (60). This parameter is not to be understood as an individual trait of the child. Among other things, it represents "the "status" of [B], where status is used to mean the entire constellation of societal factors that motivate the learning of a given language" (p. 60, emphasis in original). In short, the effect of the amount of conversation heard by a child in a certain language on her language repertoire depends on a status related parameter.²⁰ Hence, the main difference between both approaches is how the dependences of $q_{LR}(F)$ on the family type, the linguistic composition of the population and the status are conceptualized and modeled. We now review some empirical research on language acquisition and language transmission to identify the main factors that affect the probabilities $q_{LR}(F)$. On that basis we argue that Wickström underestimates the impact of the community and media on children's language acquisition. At the same time, this review emphasizes the active role of the parents and how economic and institutional factors affect the families language behavior.

The language(s) used within the family context is one key factor in shaping a child's linguistic repertoire. "Since most children do not begin schooling until ages 5 or 6 [...], the child's near total exposure to only the home language (or languages) takes place during the most important linguistically formative years of one's life" (Caldas, 2012, p. 351). A child's exposure at home comprises conversations between the parents, conversations between parents and siblings as well as communication with the child itself. Which language(s) are used at home first of all depends on the language repertoires of all the family members. If the entire family consists of l -monolinguals, $l = H, L$, then - at least with respect to conversations heard in the family - the child is only exposed to language l . In cases where more than one language is available (HB , LB and BB families) parents have to make certain decisions on which languages to use with their children. In the literature this is referred to as 'private language planning' (Piller, 2001) or 'language management' (Spolsky, 2004). So can language acquisition of children outside of formal education be thought of as the result of a utility maximizing decision of the parents, as suggested in Wickström (2005)? Empirical studies show that the answer is 'yes and no'. In the following we concentrate on research on language transmission in first or second generation migrant families and interlinguistic families.

A child is exposed to conversations between other family members, especially between both parents. The language or mix of languages used by the parent-couple might be established before the child was born. For example, Piller (2002) found that everyday language choices may originate from the couples first interactions. If both partners do not share the same L1 (say his L1 is German and her L1 is English), they might choose the language that better enables communication (they choose English since his English is better than her German). This first language choice then might become a habit. Hence, part of the amount of conversation a child is exposed to can not be seen as the result of private language planning towards the child.

²⁰In contrast to the classical status parameter, the parameter use in Fernando *et al.* (2010) can be obtained from empirical data, at least in theory. To assign values to the parameter one needs data on the probability that growing up in an HH or an HB family becomes a speaker of l in various environments with different l frequencies. Such data are not easy to acquire. In particular, measuring the amount of exposure to a language requires a lot of effort. This is one reason why we later argue for a status parameter that can be obtained from current quantitative surveys.

Various studies on bilingual and migrant families show that many parents in fact make conscious language decisions with respect to their children. For example, confronted with a dominant language that differs from their heritage language, many migrants want to pass their language (and culture) on to the next generation.²¹ Other parents (additionally) support the acquisition of the locally dominant language to guarantee better social and economic opportunities for their children.²² This suggests, that if one or both parents are speakers of a low status minority language, the language choice can be seen as “a trade-off between the efficiency aspect of language as a means of communication, on the one hand, and its importance as a carrier of identity, on the other hand” or as a “trade-off between internal, emotional motives,²³ on the one hand, and external, objective criteria, on the other hand” (Wickström, 2005, pp. 83 and 85). Obviously, for monolingual speakers of the dominant language there is no actual trade-off. Furthermore, some families in which at least one of the parents is bilingual explicitly want to raise their children bilingually as they “see bilingual education as an investment into their children’s future” (Piller, 2001, p. 77). In particular, if the minority language has a relatively high economic status (as British or American English), then parents not only transmit the minority language due to ‘emotional motives’ but also due to ‘objective criteria’. Hence, to model language transmission one has to take the specific social and economic situation as well as identity aspects into account.

From the few examples provided above we can see that many parents make language decisions for their children, which supports the ‘economic’ approach. But, and this is one flaw of this approach (at least in the version proposed by Wickström), the above studies also show that the parents decisions do not always result in actual language transmission, i.e. private language planning is by no means always successful. The (partial) failure of private language planning can have a number of reasons. The transmission of a minority language can only be successful if the initial decision translates into everyday practices over a long period of time. “Though some families may explicitly devise an initial bilingual strategy, family language policies are rarely rigid. Rather, they are likely to evolve with the changing dynamics of family life, are adjusted to account for their perceived effectiveness, or are ‘negotiated’ between family members” (Caldas, 2012, p. 354). It is not sufficient to use the minority language in the first couple of years for the child to acquire the minority language successfully.²⁴ Continuous exposure to the minority language within the family might not even be enough. Investigating families with at least one non-Dutch native De Houwer (2007) found that in around 25% of all families where either one parent only used the minority language while the other one used Dutch (one parent one language) or both parents used the minority language *and* Dutch the children did not speak the minority language. One major threat to successful minority language transmission is an external social environment that is dominated by the majority language. Starting with nursery, play-groups, kindergarten and preschool activities, the child potentially gets more and more in contact with other children and adults speaking the dominant language. “Particularly during adolescence, if a child’s peer group does not speak the minority language being spoken in the home, then it is very likely that the child will not speak the home language either”

²¹“When I asked multilingual parents if they would want to pass down their native language to their children, most said that they would” (Tuominen, 1999, p. 59). Tuominen interviewed 18 multilingual families in the US.

“[P]arents in both countries were concerned about their children’s potential loss of the mastery of the Russian language and, thus, applied different learning strategies to preserve their children’s native language skills”(Elias & Lemish, 2011, p. 11). Elias and Lemish interviewed families who migrated from Russia, 30 in Israel and 30 in Russia.

²²“As evident, it is the educational and material advantages acquired through English that elevates one’s status, escaping caste stigma. Therefore, parents encourage the learning of English among their children” (Canagarajah, 2008, p. 160f). Canagarajah investigated Sri Lankan Tamil migrant families in the US, in Canada and in Great Britain.

“Parents may believe that there are economic, employment or educational advantages of speaking a majority language to their children and not the minority language” (Baker, 2011, p. 52).

²³The wish to transmit cultural values and identity is one possible ‘emotional motive’. Being able to communicate with one’s child in the language one is most comfortable in would be another one, cf. Okita (2002, Chapter 4).

²⁴“Children who migrated with their parents when they were five or six years old told me that they had stopped speaking Tamil within a few years of migrating. They had lost their Tamil proficiency when I interviewed them as teenagers” (Canagarajah, 2008, p. 149f).

(Caldas, 2012, p. 356). By the same token, older siblings and media consumption might negatively affect the child, see e.g. Schwartz (2008). Subsequently, at some age children start making their own choices on which languages they want to use and begin to play an important role in the ‘negotiations between family members’. Thereby, children themselves have a strong impact on the language practices in the family and might counteract their parents efforts.²⁵ As a result, some parents give up while others intensify their efforts. Therefore, successful minority language transmission in an environment dominated by the majority language requires continuous efforts and resources;²⁶ language transmission is much more than just a single decision, it is something parents have to *do*.²⁷ Whether or not a family continuously invests time and other resources in language maintenance (partially) depends on its economic and social standing, as pointed out by Tuominen (1999): “It may be that the costs of language maintenance are too high for most multilingual parents, especially if they are struggling financially or lack education” (p. 63). Those with educational and financial resources have a better access to literature on multilingual child rearing, to reading materials and - if available - multilingual nurseries, kindergartens and schools. As a result, such parents have better opportunities to “choose their children’s external environments carefully”, which makes “successful bilingual family language policy” more likely (Caldas, 2012, p. 356). The cost of language transmission are weighted against the benefits, which are related to cultural identity, communication needs and deliberations on the child’s future, as outlined above. And not only parents face costs and benefits: Studying the minority language means for the children “not only doing what parents asked them to do, but losing time to do other things” (Okita, 2002, p. 126).

Summarizing we can state that successful language transmission depends on family characteristics (family type, socio-economic status and attitudes towards language) and on the linguistic environment of the family. The latter includes the strength of the language community, the presence of the minority language in the public and in the media, (the perception of) socio-economic opportunities related to different languages and institutionalized support for the minority language (availability and accessibility of bilingual nurseries, kindergartens, schools and language courses). Since a child’s peer group affects its attitudes towards and exposure to the minority language, linguistic concentration (often connected to ethnic concentration) in neighborhoods and (pre-)schools also has an impact on the success probabilities.

2.3.2 Language Learning in Formal Education

“One truth is certain: formal schooling introduces perhaps the most important outside influence (for better or worse) on a family’s strategy to rear multi-literate/multilingual children [...]” (Caldas, 2012, p. 357). With respect to formal education parents and their children have to make certain choices. First, parents can send their children to a monolingual school in the majority language, a monolingual school in the minority language or to a bilingual school. Clearly, there is only a choice if all these school types are available, which is often not the case.²⁸ Choosing a school type affects the child’s language repertoire in different ways. Sending children to a school where the minority language is used yields a higher amount of exposure to the minority language compared with a monolingual school in the majority

²⁵“Not surprisingly, many parents told me that rather than their language practices influencing their children, their children’s language choice shaped the everyday language use of everyone at home” (Canagarajah, 2008, p. 164). Similar observations can be found in Tuominen (1999) or Schwartz (2008).

²⁶We assume that children growing up in a society with a dominant majority language acquire this language at some point in their life. “Children growing up with two languages invariably learn to speak the majority language. The minority language is the one that is at risk of not being spoken” (De Houwer, 2007, p. 419).

²⁷That is why the concept of family language policies includes planning, ideology and practice, cf. see King *et al.* (2008).

²⁸Whether or not parents can truly choose a language type largely depends on educational policies. “Sadly, a good bilingual school is only an option for richer families prepared to pay for education. Some countries, such as Switzerland, Belgium, Luxembourg and the Nordic countries of Norway, Sweden and Finland are positive towards bilingualism and make an effort to include such policies within the state system” (Barron-Hauwaert, 2004, p. 48).

language. Additionally, such a school choice has an impact on the child's peer group. In comparison to a monolingual school in the majority language, there it is more likely that the child can use the minority language to communicate with her peers.

If children are enrolled in a mostly *H*-monolingual school, maintaining the minority language becomes a difficult task. Over the years children are learning abstract ideas and concepts and their vocabulary in the majority language grows. At some point, they can express themselves better in the majority language, since the daily conversations at home might not be able to compete with this enlarged vocabulary, see Okita (2002, p. 125). This weakens the minority language. To some extent this deficit can be counterbalanced by parental efforts. Practicing reading and writing at home or sending their children to supplementary language classes improves the child's abilities in the minority language. Again, this requires time and other resources from the parents as well as from the child, and supplementary language classes are not necessarily available. Another problem is that with increasing age the child's autonomy grows and more than a few start rebelling against their parents efforts, see, for example, Tuominen (1999), Okita (2002) or Schwartz (2008).

As before, there might be a trade-off for parents. To transmit the minority language in an environment dominated by the majority language a *L*-monolingual might be the best option. At the same time, the majority language is normally related to better socio-economic opportunities for the children, and it is more likely to be the language of upward mobility.²⁹ Hence, from an instrumental point of view, an *H*-monolingual school might be the best option.³⁰ Bilingual schools have the potential to satisfy both needs, but often come with extra costs. Here again, highly educated and economically well off families have more resources to inform themselves about bilingual education. For them it is easier to find appropriate schools and pay for bilingual education.

Another aspect to be mentioned here is that of (foreign) language classes in school. There might be the possibility that children from different family types take languages courses as a regular part of their education. Through such courses it is even possible that *H*-monolingual children can learn the minority language. This possibility, i.e. direct teaching of the minority language to schoolchildren that are monolingual in the dominant language, is considered in Fernando *et al.* (2010).

2.3.3 Language Learning by Adults

The majority of children growing up in an environment dominated by one (official) language (*H*) do acquire this language. This is different for adults moving to such an environment (temporarily or permanently) at a certain age but with little or no proficiency in *H*. Even if they stay for many years, they might never 'fully' acquire the local language. For language transmission we are interested in what affects the successful transmission of the minority language (*L*). Now, for adult language learning, we want to know what affects the successful acquisition of the dominant majority language. As before, the literature provides sociological and economic approaches: "Both sociologists and economists have studied immigrants' language proficiency [...]. Sociologists have mainly focused on the impact of immigrants' exposure to the language and social opportunities for language learning [...]. Economists, on the other hand, have noted the importance of immigrants' self-selection and the difficulty of learning a language" (Van Tubergen & Kalmijn, 2005b, pp. 1414f). Chiswick & Miller (2002) developed a useful model in which proficiency in *H* is a function of exposure to the language, efficiency in learning the language and economic incentives (three E's). The model is based on a human capital approach. Esser (2006) provides

²⁹"Because of this understandable instrumental motivation, parents wanted their children to learn the languages that were economically more advantageous for them" (Canagarajah, 2008, p. 165).

³⁰"For all practical purposes, though, it seems realistic to expect that most parents will enroll their children in the local monolingual majority language school" (Caldas, 2012, p. 359).

a similar model. Also building on a human capital approach, he identifies four important components for the acquisition of H by L -speaking newcomers: motivation for learning the language (including economic incentives), access to the language (including communication opportunities, which are related to exposure, and language classes), efficiency and costs for language learning. We will shortly outline both models as well as the results of the quantitative analysis presented in both papers.³¹

Language skills can be seen as human capital. They “satisfy the three requirements for human capital in that they are productive, costly to produce, and embodied in the person” (Chiswick & Miller, 2014, p. 5). Additionally, the ‘value’ of this human capital depends on the linguistic environment. With proficiency in H in an H -dominated environment one can find better jobs and be more productive in the job. Furthermore, proficiency enables more efficient consumption and communication with locals and local authorities/administration. This yields certain incentives to learn H . At the same time, language learning is costly, it requires the investment of time and other resources.

Exposure to the local language facilitates its learning in two ways. First, being exposed to and addressed in H on a regular basis yields an additional incentive to learn H , compared with a situation in which a newcomer can use L in her everyday live. Living in an area with many other speakers of L reduces the communication costs for L -monolinguals. If there is a local economy offering jobs in L , the incentives to learn H are even lower. Hence, the linguistic concentration (of L -speakers) is found to have a negative effect on H -proficiency,³² cf. Chiswick & Miller (2002) and Esser (2006). Secondly, the more one is exposed to a language, the more opportunities there are to use that language and hence the more practice one can get. Besides linguistic concentration, time since migration is identified as another important factor for exposure to H . As indicated above, children can have a positive impact on H -proficiency, since they might use H at home. Conversely, children can function as translators and hence reduce incentives for parents to learn H , see (Chiswick & Miller, 2014, pp. 35f). Concerning the motivation (incentives) to learn H it was found that those newcomers who plan to stay for a long time are more likely to acquire H . Efficiency variables are age at time of migration (with increasing age language learning decreases) and education.

2.4 Linguistic Environment and Status Variable(s)

Based on the deliberations on the individual level we now want to move on to the societal level. The aim of this section is to propose a number of relevant variables that characterize the linguistic environment and the language repertoire groups.³³ We specify how these quantifiable variables affect the processes of language transmission and learning at the macro level. This shall yield the foundation for our mathematical language competition model.

Throughout the rest of the paper we use the following notation.

- Language l , $l = H, L$
- Language repertoire LR , $LR = H, L, B$
- Language repertoire group LRG , $LRG = H, L, B$

³¹We can not go into great detail here. For extensive overviews on the connections between language and economics see, for example, Grin (2003) and Zhang & Grenier (2013).

³²Note that there might be reversed causality: migrants with little proficiency in H choose to move to linguistic enclaves because of their low proficiency, see Danzer & Yaman (2010).

³³“An economic model is always made up of two categories of ingredients, namely, variables and relationships between them; and at least some of these relationships should be causal ones” (Grin, 2003, p. 10).

Recall, by $X = (X_H, X_L)$ we denote the linguistic composition of the population. The fraction of bilingual individuals X_B is given by $1 - X_H - X_L$. The linguistic composition X evolves due to family formation, language transmission within the family, language education and adult language learning. For the mathematical model this translates into four quantities determining the dynamics of X :

- ψ_F : fraction of F -type families,
- $q_{LR}(F)$: fraction of F -type families bringing up their children with language repertoire LR ,
- s_{LR_1, LR_2} fraction of children entering school with language repertoire LR_1 and leaving school with language repertoire LR_2 ,
- a_{LR_1, LR_2} rate at which adults with language repertoire LR_1 become adults with repertoire LR_2 .

Note, q_{LR} , s_{LR_1, LR_2} and a_{LR_1, LR_2} can also be interpreted as probabilities: $q_{LR}(F)$ can be seen as the probability that a child growing up in an F -type family develops language repertoire LR . These quantities are not just fixed numbers but depend on the linguistic composition and various aspects of the linguistic environment. Let $param$ be a list of model parameters characterizing these aspects of the environment. Then, $\psi_F = \psi_F(X, param)$, $q_{LR}(F) = q_{LR}(F; X, param)$, $s_{LR_1, LR_2} = s_{LR_1, LR_2}(X, param)$ and $a_{LR_1, LR_2} = a_{LR_1, LR_2}(X, param)$. Before proposing a manageable list of model parameters we want to emphasize again that no mathematical model can account for the whole complexity of the social reality. A “model is nothing but a form of discourse about reality”, and the rationale of this discourse is to “constitute a useful instrument helping us to reflect upon and understand reality” (Grin, 2003, p. 10). If designed properly “(algebraic) models can be uniquely powerful and transparent tools in the study of selected relationships between selected sets of variables” (ebd).

As outlined earlier, family formation is modeled as a random search and mating process. On the one hand, the result of this process clearly depends on the linguistic composition of the population. If there are only a few L -monolingual adults, then the possible number of L -monolingual families (LL) is necessarily limited. On the other hand, who actually meets whom is not just a question of the overall numbers. Spatial concentration or even segregation of the minority language group might favor monolingual couples compared with inter-linguistic ones. Furthermore, linguistic concentration affects the amount of exposure to both languages and hence intergenerational language transmission, see above. Therefore we introduce a variable $C \in [0, 1]$, a measure for the concentration of speakers of language L , i.e. L -monolinguals as well as bilinguals. Treating bilinguals rather as speakers of L than as speakers of H is justified by the assumption that most L -monolinguals respectively their descendants acquire the dominant language H , while only a small number of H -monolinguals learn the minority language L . Thus, the vast majority of bilinguals is assumed to have L antecedents and hence some feeling of belongingness to the L -community. Furthermore, such a conception of bilinguals reflects the frequently encountered connection between language and ethnicity. Note, for minority languages learned by a large number of H -monolinguals, such as English, the model has to be adjusted. For $C = 0$ the language repertoire groups are distributed equally throughout the territory of consideration (no concentration). If $C = 1$, then we are dealing with a spatially segregated population (maximal concentration). As explained in Appendix A.1, a variable or index for linguistic concentration can be obtained in different ways from quantitative data. In the literature there is an ongoing discussion about the (dis)advantages of certain well-known indices, see A.1.

Next, we want to turn to the status related and institutional variable(s). We have seen that in a number of language competition models a single *status* or *prestige* variable is used, while it is not explained what it shall actually mean for a language to have status value say 0.3 or 1.3; the values of these status variables are arbitrary. To get a more meaningful status variable that can be obtained from quantitative data, we first

identify different dimensions of the status of a language. For this purpose, the systematic framework first developed in Giles *et al.* (1977) is very helpful. There, the authors suggest three categories of factors that are involved in language vitality: demographic factors, status factors and institutional support factors. Demographic factors are related to the number of speakers as well as their spatial distribution. The main status factors are economic, social and symbolic status. The economic status displays the economic standing of the language groups, the social status is related to prestige, social standing and (political) power and the symbolic status is related to identity and culture, cf. Baker (2011, pp. 55f). Institutional support comprises government (services) on different levels, mass media, business and education. For the present model, we build on this framework and propose two specific variables that describe part of the linguistic environment and reflect status and institutional support factors.

First, the socio-economic position of a language repertoire group is measured by the average socio-economic status (SES) of all individuals belonging to that group. The SES is a widely used variable in econometric and sociological analyses and can be measured in different ways. The main dimensions of the SES are income, education and occupation, see, for example, Bernard (2012, Chapter 11). Let $S_{SE}(LR) \in [0, 1]$ denote the average normalized SES of language repertoire group LR . As in Templin *et al.* (2016) we are interested in the socio-economic advantage individuals gain from speaking a certain language. Hence, instead of just asking how high the status of a certain language is, we ask how high is the status of one language is compared with the other one. Therefore, we define the relative the socio-economic status \bar{S}_{SE} . Let $LR \in \{H, L, B\}$. The relative socio-economic status of language repertoire group LR is defined by

$$\bar{S}_{SE}(LR) := \frac{S_{SE}(LR)}{S_{SE}(H) + S_{SE}(L) + S_{SE}(B)}.$$

Note, $\bar{S}_{SE}(L)$ reflects the socio-economic status of L -monolinguals, while $1 - \bar{S}_{SE}(L)$ is the socio-economic status of individuals speaking H .

Second, we consider the *institutional* or *official* status of a language. This status variable is determined by *official* domains the language can be used for. The *official* status is oriented towards Fishman's GIDS.³⁴ The *official* status clearly belongs to institutional support factors in the Giles *et al.* framework. Above that, the *official* status might reflect the prestige of a language to some extent and, at the same time, affect the prestige of the language. For example, providing governmental services in a minority language can improve its perceived status. This illustrates, how the different categories of variables are strongly connected to one another. For the design of a single variable that captures various aspects of the interaction between individuals with a certain language repertoire and *official* institutions, we first have to provide a list of d domains. For each domain we make the simplifying assumption that a language can either be used or not used. Counting the number of domains a language l , $l = H, L$, can be used for and dividing this number by d , the total number of domains, we obtain a status variable $S_{OF}(l)$ with possible values in $\{0, 1/d, 2/d, \dots, (d-1)/d, 1\}$. If $S_{of}(l)$ equals 1, then the language can be used in all d domains, which might be the case for the dominant local language. Since education and media are treated separately, we here consider three domains: government at national level, government at provincial level and public administration. This yields an *official* status variable $S_{OF}(l)$, $l = H, L$, with possible values in $\{0, 1/3, 2/3, 1\}$. Since H is the locally dominant and official/national language we assume $S_{OF}(H)$ to be one, i.e. H can be used in all domains. As for the socio-economic status, we introduce the relative official status $\bar{S}_{OF}(l) := S_{OF}(l)/(S_{OF}(H) + S_{OF}(L)) = S_{OF}(l)/(1 + S_{OF}(L))$.

Although the socio-economic status and the official status of the minority language are clearly related to *objective criteria*, cf. Section 2.3.1, both status variables are also connected to the prestige of that language, see above. So are these two status variables sufficient to account fully for the *symbolic* status of the minority language, and the *emotional motives* that play an important role in its intergenerational

³⁴Graded intergenerational disruption scale, see Fishman (1991).

transmission? Probably not. We have seen that the majority of migrant parents want to pass their first language on to their children despite its economic and official status, since this is the language they are most comfortable in and that is associated with their (cultural) heritage and (group) identity. But how to quantify and measure this cultural dimension? The major problem lies in the difficulty, if not even impossibility, of proposing a single number that adequately describes the average attachment to, or *value* of, the heritage language for its speakers. “The components of non-market value [of a language] are very difficult to identify theoretically, and no less difficult to measure empirically [...]” (Grin, 2003, p. 38). One theoretically possible strategy to assess the average – for individuals or families of certain types – non-market value is to ask directly a representative sample of individuals or individual families in surveys.³⁵ Another strategy is to use somewhat easier measurable factors that are linked to the non-market value/symbolic status of the minority language. One such factor could be community support for the minority language. We have seen in Section 2.3 that support from the language community, i.e. through language classes, is one factor for the success of parents’ endeavors to transmit a minority language to the next generation. Furthermore, community support can be thought of as a proxy for the importance of maintaining and transmitting the language within the language community. To include (the strength of) community support as a single variable in a formal modeling framework, a simple but sound operationalization is needed. For any operationalization it has to be taken into consideration that community support might depend on available socio-economic resources within the community or support by external language planning programs. Another possible factor is the presence of the minority language in mass media. “Mass media have considerable influence on language prestige and choice, often bringing the dominant language into the home for the first time” (Sallabank, 2012, pp. 117f). Besides the positive effect on its prestige, the presence of *L* in the (local) mass media increases its usefulness and comes with a certain amount of potentially high status exposure to the language. These are some of the ways in which minority language media can help parents with their language transmission efforts, see, for example, Elias & Lemish (2011), and more generally, Cormack (2007). As for community support, to use media presence as a variable in a mathematical model, it has to be quantified, which is quite difficult again. First of all, different types of media are used: print media, radio, television and – to a growing extent – the Internet. Moreover, most migrants not only consume local media (host-country), but also media from the home-country or international media, while thanks to the Internet access to both is easier and cheaper as ever before. Further, the effect of minority language media on language maintenance and transmission is not at all clear, cf. Cormack (2007).

For the reasons outlined in the last paragraph, we restrict ourselves to the socio-economic status and the official/institutional status for the current model. We leave it to future research to extend the current model with further variables that display more dimensions of the status of the minority language to get a more complex model, which is closer to the complex reality of the processes at hand.

The last feature of the linguistic environment included in the language dynamics model proposed here is language education policy. Restricting ourselves to two languages, the language policy determines to what extent the languages *H* and *L* can be used and learned in the education system.³⁶ One of the most important decisions to be made concerns the language(s) of instruction. Roughly, for the three levels of education (primary, secondary and tertiary) schools could use only *H*, only *L* or both languages as medium of instruction.³⁷ In many cases, the default language of instruction is the (locally) dominant,

³⁵Parents could, for example, be asked how much they are *willing to pay* to successfully transmit their heritage language to the next generation. For the economic *willingness to pay* approach for languages see Grin (2003, pp. 40f). Alternatively, one could try to assess parents language ideologies and beliefs in the different communities, see King *et al.* (2008), to construct a single variable that reflects the importance of language transmission.

³⁶It is out of the scope of this paper to delineate the complex and ongoing discussions on language planning in education, mother tongue education, bilingual education and second/foreign language education. For an overview on various perspectives on and approaches to language policy and education see, for example, Hornberger & May (2010) and Baker (2011).

³⁷For a more detailed typology (10 types) see Baker (2011, Chapter 10).

official or national language H . Additionally, schools can offer L as a second/foreign language for H -monolingual pupils or special classes for L -monolingual pupils to learn H (H as a second language). The type of program a student is enrolled in affects her abilities in both languages, see above. To consider a specific language education policy within the model framework, one has to specify the effects of the policy on the fractions s_{LR_1, LR_2} , $LR_1, LR_2 = H, L, B$. Hence, to model a specific policy reliable estimates on the (linguistic) effects on pupils with different linguistic repertoires are needed. Furthermore, if the education policy comprises different programs (e.g. H -monolingual programs as well as bilingual programs) information on the fraction of students enrolled in these programs is needed. For this paper we do not model specific educational language policies. Instead, we work directly with the fractions s_{LR_1, LR_2} . Hence, we are only able to investigate the long-term effects of changes in these fractions. To analyze the effects of certain policies on the long-term development of the linguistic composition, one has to make assumption on what fractions correspond to the policy.

Let us summarize the main components of our language competition model. In Giles *et al.* (1977) the authors propose three categories of factors that influence language vitality: demographic factors, institutional support and status factors. The linguistic composition of the population X and the linguistic concentration measure C are the *demographic* factors considered in the present model. Furthermore, the linguistic environment is characterized by two normalized status parameters: the socio-economic status \bar{S}_{SE} and the official status \bar{S}_{OF} . The language education is no explicit model parameter but reflected by the fractions s_{LR_1, LR_2} . The status and education parameters reflect some important institutional support and status factors in the Giles *et al.* (1977) framework.

Next, we propose simplifying assumptions on how the linguistic composition as well as the environment parameters affect the four quantities ψ_F , q_{LR} , s_{LR_1, LR_2} and p_{LR_1, LR_2} driving the language dynamics.

Family formation: The expected number of families of a certain type F depends on the linguistic composition of the population X as well as on linguistic concentration C . Hence, $\psi(F, X, param) = \psi(F, X, C)$. The higher the linguistic concentration, the lower the expected number of mixed families. The exact formulas for $\psi(F)$ are provided in Section 3.

Language transmission: Recall, $q_{LR}(F)$ denotes the fraction of families of type F bringing up children with language repertoire LR . Before stating hypotheses on how these quantities depend on our parameters describing the linguistic environment, we make some simplifying assumptions:

A1: All children in one family have same language repertoire

A2: Both parents shall be able to communicate with their children, i.e. $q_L(HH) \equiv q_L(HB) \equiv 0$ and $q_H(LL) \equiv q_H(LB) \equiv 0$

A3: The more present a language l is in a certain family type, the more likely it is that children become monolinguals in l :

$$A3a: q_H(HH, param; X) \geq q_H(HB, param; X) \geq q_H(BB, param; X)$$

$$A3b: q_L(LL, param; X) \geq q_L(LB, param; X) \geq q_L(BB, param; X)$$

A4: The higher the number of l -monoglots, the higher the incentive to transmit l to ones children.³⁸

A4a: If X_H increases, then q_H and q_B do not decrease and q_L does not increase

A4b: If X_L increases, then q_L and q_B do not decrease and q_H does not increase

A5: In an HH family children grow up as monolingual H speakers, i.e. $q_H(HH) \equiv 1$

³⁸For a derivation of the following assumption from a utility maximization approach see Wickström (2005).

Furthermore, we assume that the $q_{LR}(F, param; X)$ are continuous and almost everywhere differentiable functions with respect to X . Most of these assumption can already been found in Wickström (2005) and Templin *et al.* (2016). Let us proceed with the effect linguistic environment parameters have on the q -fractions.

The status parameters $\bar{S}_{SE}(L)$ and $\bar{S}_{OF}(L)$ reflect the relative disadvantage of L -monolinguals. The lower both parameters, the higher the disadvantage in socio-economic and institutional terms, and, in turn, the lower the utility from speaking the minority language. Consequently, the lower both parameters, the lower the incentive – from an instrumental point of view – to transmit L to the next generation. In Section 2.3.1 a second effect by which the socio-economic status affects language transmission was outlined. Successful language transmission requires financial and educational resources. Parents with high socio-economic status that want to transmit the minority language have better chances than parents with low SES, everything else being equal. A third effect of $\bar{S}_{SE}(L)$ and $\bar{S}_{OF}(L)$ on the transmission of L mentioned earlier has to do with the prestige of L . The higher the average socio-economic status and the more official domains the language can be used for, the higher might be it's symbolic status. Note that all three effects push in the same direction. Therefore, we assume that the higher any of the two status variables, the lower probability that HB or BB families only transmit H . This assumption can be expressed by:

$$\frac{\partial q_H(HB)}{\partial \bar{S}_{SE}(L)}, \frac{\partial q_H(BB)}{\partial \bar{S}_{SE}(L)}, \frac{\partial q_H(HB)}{\partial \bar{S}_{OF}(L)}, \frac{\partial q_H(BB)}{\partial \bar{S}_{OF}(L)} \leq 0. \quad (1)$$

Similar arguments can be put forward for LL and LB families. The lower $\bar{S}_{SE}(L)$ and $\bar{S}_{OF}(L)$, the higher the incentive to transmit or teach (e.g. in kindergarten or preschool) the majority language to children. Therefore,

$$\frac{\partial q_L(LL)}{\partial \bar{S}_{SE}(L)}, \frac{\partial q_L(LB)}{\partial \bar{S}_{SE}(L)}, \frac{\partial q_L(LL)}{\partial \bar{S}_{OF}(L)}, \frac{\partial q_L(LB)}{\partial \bar{S}_{OF}(L)} \geq 0. \quad (2)$$

In (1)-(2) one can see that both status parameters corresponding to the minority language, i.e. $\bar{S}_{SE}(L)$, $\bar{S}_{OF}(L)$, affect fractions $q_L(F)$ in the same direction. To simplify the formulas in the upcoming section we therefore introduce a status indicator composed of the two status parameters. Let α_{SE} and α_{OF} be weights, i.e. $\alpha_{SE}, \alpha_{OF} \in [0, 1]$ and $\alpha_{SE} + \alpha_{OF} = 1$. Given such weights we define the composed status indicators

$$S(L) := \alpha_{SE} \cdot \bar{S}_{SE}(L) + \alpha_{OF} \cdot \bar{S}_{OF}(L), \quad (3)$$

$$S(H) := \alpha_{SE} \cdot \bar{S}_{SE}(H) + \alpha_{OF} \cdot \bar{S}_{OF}(H). \quad (4)$$

Since both single status parameters are normalized, the composed status indicators have values between 0 and 1 as well. Furthermore, we have $\bar{S}_{SE}(H) = 1 - \bar{S}_{SE}(L)$, $\bar{S}_{OF}(H) = 1 - \bar{S}_{OF}(L)$. Therefore, we get $S(L) = 1 - S(H)$. The composed indicator $S(L)$ reflects the relative linguistic disadvantage of L -monolinguals in terms of socio-economic standing, public institutions, politics, media and community support. With the composed status indicator and the linguistic concentration measure C , we end up with two main parameters characterizing the linguistic environment, i.e. $param = \{S(L), C\}$.

Note, that due to assumptions A2 and A3a we have $q_B(HB) = 1 - q_H(HB)$, $q_B(LL) = 1 - q_L(LL)$ and $q_B(LB) = 1 - q_L(LB)$. Plugging this into (1) and (2) yields the respective statements for $q_B(HB)$, $q_B(LL)$ and $q_B(LB)$. Furthermore, $q_L(BB) + q_B(BB) = 1 - q_H(BB)$. Hence,

$$\frac{\partial q_B(BB)}{\partial S(L)} + \frac{\partial q_L(BB)}{\partial S(L)} \geq 0. \quad (5)$$

	F	$S(L)$	C
$q_H(F, S(L), C; X)$	HB	-	
	BB	-	-
$q_L(F, S(L), C; X)$	LL	+	+
	LB	+	+
	BB		

Table 1: Relation between language transmission and different environment parameters. “+” indicates a positive effect of the environment parameter, “-” indicates a negative effect and “ ” indicates no or unclear effect. Note, $q_B(F, param; X) = 1 - q_H(F, param; X) - q_L(F, S(L), C; X)$.

Linguistic concentration affects the exposure to both languages. Strong linguistic concentration keeps the exposure to H low for L -speaking children and vice versa. Since bilinguals are considered rather as L 's than as H 's, we assume that given high concentration LB and BB families predominantly live in L dominated neighborhoods. HB families are assumed to be found in L -dominated as well as H -dominated neighborhoods. Therefore,

$$\frac{\partial q_L(LL)}{\partial C}, \frac{\partial q_L(LB)}{\partial C} \geq 0 \quad (6)$$

$$\frac{\partial q_H(BB)}{\partial C} \leq 0. \quad (7)$$

Due to the last statement we also get

$$\frac{\partial q_B(BB)}{\partial C} + \frac{\partial q_L(BB)}{\partial C} \geq 0. \quad (8)$$

The statements (1)-(7) are displayed in a more accessible fashion in Table 1.

Language education: We assume that neither H -monolingual nor bilingual children will unlearn H through formal education, i.e. $s_{H,L} = s_{B,L} = 0$. Hence, $s_{H,H} = 1 - s_{H,B}$ and $s_{B,B} = 1 - s_{B,H}$. Note, if $s_{H,B} > 0$, then some former H -monolingual children learn L at school. This is only possible if L is taught as a second language. If $s_{B,H} > 0$, then some children entering school as bilinguals leave the school system as H -monolinguals, which might be the case if a H -only language policy is applied. In principle we could include the option that L -monolingual children can get educated fully in L , which translates to $s_{L,L} > 0$. In practice, since H is a dominant majority language, all children learn H at some point in their educational career. Therefore, we set $s_{L,L} = 0$. Hence, $s_{L,B} = 1 - s_{L,H}$. Finally, it is reasonable to assume that bilingual children are more likely to leave school as H -monolinguals than children entering school as L -monolinguals, i.e. $s_{B,H} > s_{L,H}$. Note that the s_{LR_1, LR_2} are conceptualized as constant model parameters. They do not depend on the distribution of speakers but reflect language education policies and the quality of language education.

Adult language learning: We have seen above that the acquisition of the local language H by adult newcomers is a question of exposure, efficiency, (economic) incentives and costs. The same approach can be applied to H -monoglots learning the minority language L . The quantities describing adult language learning in the model are $a_{L,B}$ (L -monoglots become bilingual) and $a_{H,B}$ (H -monoglots become bilingual). Since linguistic concentration C has a strong effect on exposure to the other language, we

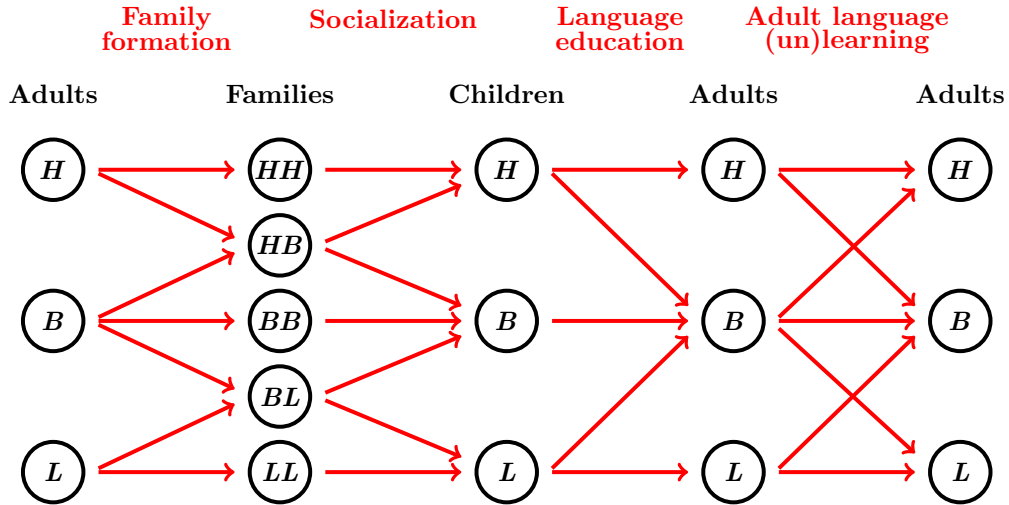


Figure 1: Illustration of general language dynamics.

assume that both $a_{L,B}$ and $a_{H,B}$ decrease if C increases. The role of the status variable \bar{S} is not as easy to pin down. On the one hand, if the economic and official statuses of the minority language group are low, then there are strong (economic) incentive for L -monoglots to learn H , since H promises much better economic opportunities. On the other hand, if the socio-economic status is low, then L -monoglots might have only few financial and educational resources, a potential hindrance for learning a new language (efficiency and costs). Appropriate programs encouraging and supporting the acquisition of the local language (e.g. offering free language classes) might be able to ease the socio-economic disadvantages. Later on, when specific functional expressions are introduced for $a_{L,B}$ and $a_{H,B}$, both effects are taken into account. The support of the local language acquisition is reflected in one parameter of the functional expressions.

2.5 Population Dynamics

In most of the language competition models reviewed so far the population was considered to be constant in the sense that no new/external individuals enter the population (except the ones born within the population) and that individuals do not leave the territory. Thus, migration and emigration movements are neglected.³⁹ Yet, one can observe vast mobility in many parts of this world. According to the United Nations (2013) there were 232 million international migrants worldwide in 2013, and in the global north migrants make up 10.8% of the total population. Looking, for example, at the European Union, single member states face mobility within the Union as well as from and to non-EU states. From the perspective of an individual EU member state, such mobility flows are external. Since this paper is mainly concerned with ‘new minorities’, it will focus on the migration of people with a heritage language that is the dominant or national language of the host country. In models with different growth rates for the different language groups, see below, external mobility can be incorporated into the growth parameters (Malthus rates), cf. Kandler *et al.* (2010).

In some models the population is also constant in a second sense. In Wickström (2005) every family has on average two children,⁴⁰ in Minett & Wang (2008) every adult has one child and in Fernando

³⁹As noted above there are several models concerned with internal mobility, i.e. mobility within the territory.

⁴⁰This assumption (Assumption 4.1) is made later in his analysis. In the more general model introduced before, a family of type F has on average $2\beta(F)$ children, $\beta > 0$.

et al. (2010) the death rate equals the birth rate. As Minett & Wang (2008) state for the original Abrams/Strogatz model: “The population size [is assumed to be] constant” (p. 21). The same holds true for John & Yi (2001) or Mira & Paredes (2005). Hence, in these models population dynamics are neglected. This is different for an other set of models, starting with Baggs & Freedman (1993). There, the authors include different birth and death rates for the three groups (two monolingual groups and a bilingual one). An other way to incorporate population dynamics can be seen in Pinasco & Romanelli (2006). Their model is inspired by the Lotka-Volterra model, which describes the dynamics in a predator-prey system. A maximal carrying capacity is assumed, i.e. at every point in time only a limited amount of resources is available. If the population size is below this capacity, the population increases, and decreases whenever it is above this capacity. Extensions of this model are proposed in Kandler & Steele (2008) and Kandler *et al.* (2010). The authors in Heinsalu *et al.* (2014) argue that given homogeneous growth rates for all groups, the population dynamics terms in their differential equation disappear asymptotically. Hence, assuming homogeneous growth the population dynamics do not affect the steady states. Therefore, they neglect population dynamics.

For various reasons, see Sleebos (2003) for example, several developed countries show fertility rates that are not sufficient to preserve the population in the long-term. For example, in the European Union the average number of children per women fluctuated around 1.5 in the last decades, while approximately 2.1 children are necessary to reproduce the population, see e.g. Eurostat (2014). To investigate different scenarios for the future linguistic development in European states as well as in other states facing a comparable situation, the model here takes external mobility and population dynamics into account. Instead of assuming some maximal carrying capacity, we work with birth and death rates that could be obtained e.g. from Eurostat. Such a model could also be used to explain past developments where mobility and population dynamics played an important role. In contrast to Kandler & Steele (2008) and Kandler *et al.* (2010), we do not develop a spatial model. A spatial extension might be one object of future research.

2.6 Time Scales and Time Frames

In our opinion, time is an underestimated issue of language dynamics models. In Abrams & Strogatz (2003) the relative number of speakers of the minority language x evolves according to a differential equation. This equation describes how x changes over time t . Although it is not made explicit, the plots suggest that time is measured in years. The authors fit their model to data for up to 100 years. For their extension Minett & Wang (2008) use a similar kind of differential equation but introduce a mortality rate μ , the rate “at which adults are replaced by children” (p. 25). Since such rates are often yearly rates, it could be assumed time is also measured in years.⁴¹ But again, it is not specified how long one period of time $((t + 1) - t)$ actually is. In their simulations t is not called years but “iterations” and 2000 such iterations are calculated. It remains unclear, if 2000 iterations means 2000 years or stands for some other but not specified amount of time.

These two examples shall illustrate some ‘problems’ that can be found in most of the models available in the literature. First, the time scale is either not specified or not justified. This might not be problem if models are fitted to data, since in most cases there is some parameter regulating the speed of the dynamics.⁴² In the absence of such a parameter Zhang & Gong (2013) solve the problem of fitting their model to empirical data via optimizing over the time scale, see above. Furthermore, for some authors the time scale seems irrelevant, since they are only interested in long-term stable equilibria. From this point of view it does not matter whether such an equilibrium is reached within 50 or 1000 years. But, and this

⁴¹As an example Minett & Wang (2008) use $\mu = 2\%$, which would support this conjecture.

⁴²In Abrams & Strogatz (2003) Mira & Paredes (2005) and Minett & Wang (2008) this parameter is called B or $c_{L_1L_2}$.

is one key point of our critique, the latter is only true if parameters are constant over time. The status is one such parameter. Why should the status of a dying minority language be constant over a century or even longer? For that reason Templin *et al.* (2016) developed a model where the status of the minority language shrinks over time if no revitalization measures are applied to increase the status. Since it is only asked whether or not the language can be saved and at what costs, the speed of change used in the examples is also not justified.

For a “mechanistic” model, we would argue, one should have a more precise idea of how fast things happen and what the time horizons are in which it is plausible that the parameters remain constant.

3 General Model Formulation

In this section we state the formal language dynamics model. As before, X_H , X_L and X_B denote the fractions of H -monoglots, B -monoglots and bilinguals. Since the population size might not be constant we also work with absolute numbers. Let N_H be the number of H -monoglots, N_L be the number of L -monoglots and N_B be the number of bilinguals. The total population size is given by $N := N_H + N_L + N_B$. Note that $N_H = N \cdot X_H$, $N_L = N \cdot X_L$ and $N_B = N \cdot X_B$. We use both, relative numbers (X_H) and absolute numbers (N_H), since ψ and q only depend on relative numbers, while the overall dynamics are described in absolute terms.

The language dynamics models describes in mathematical terms how N_H , N_L and N_B change over time as a result of family formation, language transmission, language learning in formal education and adult language learning. Let $X := (X_H, X_L)$.

Family formation is the result of a random search and mating process affected by linguistic concentration C . As in Templin *et al.* (2016) we assume that both adults shall be able to communicate properly with one another. Hence, we neglect families of type HL . The random family formation process yields the following distribution of family types ψ :

$$\psi(HH, C; X) = (C + (1 - C)X_H)X_H + (1 - C)X_HX_L \quad (9)$$

$$\psi(HB, C; X) = 2(1 - C)X_HX_B \quad (10)$$

$$\psi(LL, C; X) = \left(1 + C \frac{X_H}{1 - X_H}\right) X_L^2 + (1 - C)X_HX_L \quad (11)$$

$$\psi(LB, C; X) = 2 \left(1 + C \frac{X_H}{1 - X_H}\right) X_LX_B \quad (12)$$

$$\psi(BB, C; X) = \left(1 + C \frac{X_H}{1 - X_H}\right) X_B^2. \quad (13)$$

See the Appendix A.2 for a justification of the family type distribution. Note, for $C = 0$, i.e. no linguistic concentration, the family type distribution is the same as given in Templin *et al.* (2016).

The process of language transmission from one generation to the next is determined by the distribution of family types and the fractions $q_{LR}(F)$. First of all, for all family types F these fractions have to add up to one. Given the distribution of language repertoires (X) in one generation and the relevant environment parameters (the status indicator $S(L)$ and the linguistic concentration C), the fraction of the next generation equipped with the language repertoire LR is given by

$$\sum_F q_{LR}(F, S(L), C; X) \cdot \psi(F, C; X). \quad (14)$$

In the above formula $q_{LR}(F, S(L), C; X) \cdot \psi(F, C; X)$ is the (relative) number of F -type families ($\psi(F)$) times the fraction of F -type families raising their children with repertoire LR ($q_{LR}(F)$). This is the summed over all family types F .

In the following let t be time measured in years. $N(t)$ is the overall population size at time t . $N_H(t)$, $N_L(t)$ and $N_B(t)$ denote the sizes of the language repertoire groups. Furthermore, the vector

$$X(t) = (X_H(t), X_L(t)) := \left(\frac{N_H(t)}{N(t)}, \frac{N_L(t)}{N(t)} \right)$$

describes the linguistic composition of the population at time t (relative instead of absolute numbers). By λ we denote the annual birth rate and by μ the annual death rate. It is assumed that birth and death rates are the same for all language repertoire groups. The model could easily be adjusted to cases of differing death and birth rates for the different language repertoire groups and family types (just substitute μ by μ_{LR} and λ by λ_F in (15)). Neglecting mobility, language education and adult language learning for the moment, the dynamics of the basic model can be described by the following three differential equations:

$$\dot{N}_{LR}(t) = -\mu N_{LR}(t) + \lambda N(t) \sum_F q_{LR}(F, S(L), C; X(t)) \psi(F, C; X(t)), \quad (15)$$

$LR = H, L, B$. The first summand represents the number of people with language repertoire LR dying at time t . The second summand represents all the children raised with language repertoire LR at time t . The overall population size changes according to $\dot{N}(t) = (\lambda - \mu)N(t)$. Note, to fully describe the dynamic system $\dot{N}(t)$, $\dot{N}_H(t)$ and $\dot{N}_L(t)$ are sufficient, since $N_C = N - N_H - N_L$.

We now extend the basic model by stepwise including the other three processes affecting the language dynamics. Before that we introduce the following notation for better readability.

$$f_{LR}(S(L), C; X) := \sum_F q_{LR}(F, S(L), C; X) \psi(F, C; X), \quad LR = H, L, B. \quad (16)$$

The functions f_{LR} are associated with raising children with language repertoire LR .

Formal education: s_{LR_1, LR_2} denotes the fraction of children entering school with language repertoire LR_1 and leaving school with LR_2 . We exclude the possibility that children enter school as monolinguals in one language and leave school as monolinguals in the other language, i.e. $s_{H,L} = s_{L,H} = 0$. Furthermore, we assume that bilingual children do not unlearn the majority language during their education, i.e. $s_{B,L} = 0$. In contrast, we include the option that some bilingual children might switch to H -monolinguals due to an H -dominated education, i.e. $s_{B,H} \geq 0$. Monolingual children of both language might learn the other language, i.e. $s_{H,B}, s_{L,B} \geq 0$. The first extension of the basic model described by (15) is then given by

$$\dot{N}_H(t) = -\mu N_H(t) + \lambda N(t) ((1 - s_{H,B}) f_H(S(L), C; X(t)) + s_{B,H} f_B(S(L), C; X(t))) \quad (17)$$

$$\dot{N}_L(t) = -\mu N_L(t) + \lambda N(t) (1 - s_{L,B}) f_L(S(L), C; X(t)). \quad (18)$$

To simplify notation even further we introduce

$$g_H(S(L), C; X) := (1 - s_{H,B}) f_H(S(L), C; X) + s_{B,H} f_B(S(L), C; X), \quad (19)$$

$$g_L(S(L), C; X) := (1 - s_{L,B}) f_L(S(L), C; X). \quad (20)$$

Adult language learning: We only consider acquisition of an additional language, i.e. the transition from monolingualism to bilingualism ($a_{H,B}$ and $a_{L,B}$). Note, in contrast to the constants $c_{H,B}$ and $c_{L,B}$,

the transition parameters $a_{H,B}$ and $a_{L,B}$ depend on C , $S(L)$ and X . Using the notation introduced in (19)-(20) we add one additional term in (17) and (18). Let $LR = H, L$. The second extension is given by

$$\dot{N}_{LR}(t) = -[\mu + (1 - \mu)a_{LR,B}(S(L), C, X(t))]N_{LR}(t) + \lambda N(t)g_{LR}(S(L), C; X(t)). \quad (21)$$

To shorten notation we write $g_{LR}(t)$ for $g_{LR}(S(L), C; X(t))$ and $a_{LR,B}$ for $a_{LR,B}(S(L), C, X(t))$.

Mobility: The absolute number of people equipped with language repertoire LR migrating to the population is denoted by $M_{LR}(t)$. The total number of migrants is given by $M(t) := M_H(t) + M_L(t) + M_B(t)$. Note, in principle $M_{LR}(t)$ could be negative, which would indicate net emigration of LR 's. Later we investigate two special cases. In the general case we get $\dot{N}(t) = (\lambda - \mu)N(t) + M(t)$ and hence

$$\dot{N}_{LR}(t) = -[\mu + (1 - \mu)a_{LR,B}(t)]N_{LR}(t) + \lambda N(t)g_{LR}(t) + M_{LR}(t), \quad (22)$$

$LR = H, L$.

Additionally, we define $m_{LR}(t) = M_{LR}(t)/M(t)$, the share of LR mobility. In the following paragraphs we consider two special cases for migration. In the first case, migration is assumed to be constant over time, i.e. every year the same absolute number of migrants enter the population (e.g. 100,000 migrants per year). In the second case, migration is constant relative to the population size. If e.g. the government allows for an annual migration of 2% of the total population size, then migration is constant relative to the population size. This implies that the number of migrants increases with an increasing population size.

Special case 1 (Constant absolute migration flow): Let $M_{LR}(t) = M_{LR}$ be constant. Trivially, if $M = 0$ and $\lambda = \mu$ the total population size remains constant. Furthermore, if $M > 0$ and $\lambda < \mu$, then the population size $N(t)$ converges to the steady state $N(\infty) := M/(\mu - \lambda)$. Let $\lambda < \mu$ and $M > 0$. Applying the quotient rule to $X_{LR} = N_{LR}/N$ and setting $N = N(\infty)$ we obtain for the steady state population size

$$\dot{X}_{LR}(t) = -[\mu + (1 - \mu)a_{LR,B}(t)]X_{LR}(t) + \lambda g_{LR}(t) + (\mu - \lambda)m_{LR}. \quad (23)$$

Special case 2 (Constant relative migration flow): Let $M(t)/N(t) =: \nu$ be constant as well as the fractions $M_{LR}(t)/N(t) =: \nu_{LR}$. Then, $\nu = \nu_H + \nu_L + \nu_B$, $\dot{N}(t) = (\lambda - \mu + \nu)N(t)$ and

$$\dot{N}_{LR}(t) = -[\mu + (1 - \mu)a_{LR,B}(t)]N_{LR}(t) + \lambda N(t)g_{LR}(t) + \nu_{LR}N(t). \quad (24)$$

Using the last two equations and applying the quotient rule to $X_{LR} = N_{LR}/N$ we obtain

$$\dot{X}_{LR}(t) = -(1 - \mu)a_{LR,B}(t)X_{LR}(t) + \lambda g_{LR}(t) + \nu_{LR} - (\lambda + \nu)X_{LR}(t). \quad (25)$$

Note that $\nu_{LR} = M_{LR}/N = \nu M_{LR}/M = \nu m_{LR}$. If $\lambda < \mu$ and $(\mu - \lambda) = \nu$, i.e. if the total population size remains constant, then (23) and (25) yield the same formula.

4 Specific Functional Forms

In this section we specify the functions $q_{LR}(F, param; X)$ and $a_{LR,B}(param; X)$. For intergenerational language transmission we build on the functional expressions proposed in Templin *et al.* (2016). Here, these functional expressions are denoted by $\tilde{q}_L(F)$, see Appendix A.3. These functions are constructed

in such a way that if X_H is sufficiently high and $S(L)$ is sufficiently low, then bilingual (BB) and even mixed couples (LB) do not raise their children as L -monoglots. Earlier we made the assumption that $q_H(HH) = 1$. To preserve symmetry, we now also assume that $q_L(LL) = 1$, i.e. $\eta = 0$. Both assumptions imply that parents can only transmit languages they speak properly.

In Templin *et al.* (2016) the authors neglected linguistic concentration, which corresponds to $C = 0$. Therefore, we set $q_{LR}(F, S, C; X) = \tilde{q}_{LR}(F, S; X)$ whenever $C = 0$. For $C = 1$, i.e. maximal linguistic concentration, we assume that LB and BB families predominantly live in L -dominated areas. Therefore, given maximal concentration, children in BB families do not grow up as H -monoglots, and children in LB families become L -monoglots. Furthermore, we assume that half of the HB families live in H -dominated areas, while the other half resides in L -dominated areas. Given maximal concentration, we thus assume that one half of all HB children grow up as H -monoglots, while the other half becomes bilingual. For the mathematical formulae see Appendix A.4.

Next, we specify the functions $a_{l,B}(S(L), C; X)$, $l = H, L$. Let θ and ϕ be a parameters between 0 and 1. Support for the acquisition of H and L for monolingual speakers of the other language is modeled by the parameters u_H and u_L . We set

$$\begin{aligned}\tilde{a}_{L,B} &:= \max \{0, \theta S(H)X_H - \phi(1 - u_H)\}, \\ \tilde{a}_{H,B} &:= \max \{0, \theta S(L)X_L - \phi(1 - u_L)\}.\end{aligned}$$

For maximal support for the acquisition of the local language H for L -monoglots, i.e. $u_H = 1$, the function $\tilde{a}_{L,B}$ is strictly increasing in $S(H)$ and X_H . This reflects that the higher the socio-economic and communicational incentives associated with H as well as the exposure to H , the more L -monolinguals learn H . If no support is made available, then $S(H)$ and X_H have to be high enough such that L -monoglots acquire H . Since linguistic concentration hinders learning of the other language, we set $a_{l,B} = (1 - C) \tilde{a}_{l,B}$, $l = H, L$. Hence, given maximal concentration/segregation, no monolingual individuals learn the other language.

5 Steady States

In this section we want to investigate possible steady states of the systems described in the previous two sections. First, we will look at monolingual steady states, that is situations without any bilingual speakers. Thereafter we are concerned with the more interesting case of steady states containing a bilingual subpopulation. Steady states are denoted by \hat{X}_H , \hat{X}_L and \hat{X}_B . We only consider the special cases with constant absolute and constant relative migration flows and assume that $\lambda \leq \mu$.

5.1 Monolingual steady states

In a monolingual steady state we have $\hat{X}_B = 0$ and hence $\hat{X}_H + \hat{X}_L = 1$. Clearly, this can only yield a steady state if no bilingual speakers enter the society any more, i.e. $M_B \equiv 0$. Moreover, neither children nor adults of either language group can learn the other language, otherwise some would become bilingual. Since HL couples have probability zero, all families are of types HH and LL . This implies that a monolingual steady state is only possible if either there is only one language group left ($\hat{X}_H = 1$ or $\hat{X}_L = 1$,) or the population is entirely segregated, which would include L -only education for L -pupils and H -only education for H -pupils. For $\hat{X}_H = 1$ to be stable there can not be any L -migration and for $\hat{X}_L = 1$ to be stable (economic) incentives to learn H have to be sufficiently low. We are predominantly interested in cases with a positive inflow of L -speaking newcomers and where the status of the locally

dominant language H is higher than the one of L , i.e. $S(L) < 1/2$. Therefore, we neglect the cases for which monolingual equilibria exist.

5.2 Bilingual steady states

Steady states without any monolingual speakers of the minority language, i.e. $\hat{X}_L = 0$, are only possible, if all migrants already enter the society as H -monoglots or bilinguals ($M_L \equiv 0$). Given a highly restrictive migration regime, this could be an imaginable scenario. In reality, though, most states face an influx of at least some people with no or only little knowledge of the locally dominant language. For the numerical examples in the subsequent section we therefore assume $M_L > 0$.

In principal, steady states without a monolingual H -subpopulation ($\hat{X}_H = 0$) are possible as well. Vast migration of L -monoglots combined with a low fertility rate (small λ) and strong linguistic concentration can lead to a situation in which low rates of H -acquisition by L 's can not compensate for the shrinking monolingual H population. Whether or not then part of the population remains bilingual, i.e. whether this yields a steady state with $\hat{X}_B > 0$, depends on the status of H . If the status is relatively low, it is theoretically possible that H is not learned or transmitted anymore, such that $\hat{X}_L = 1$, but this is a very unlikely scenario.

In a third possible type of steady states there are monolingual speakers of both languages as well as bilinguals, i.e. $\hat{X}_H, \hat{X}_L, \hat{X}_B > 0$. Given a steady influx of L -monolinguals combined with low linguistic concentration and low status of L , a sufficient number of HB and BB couples raise their children only in language H , such that a certain H -monolingual subpopulation remains stable. In the following section we present numerical examples with and without such an H -monolingual subpopulation.

6 Intervention

Some aspects of the linguistic environment such as the official status of H and L or the language(s) used in education are already a result of language planning policies. Governmental interventions can thus be thought of as alterations in the linguistic environment. In terms of the mathematical model, governmental interventions can be conceptualized as changes in the parameters characterizing the environment. The most relevant parameters in our model are probably the official status of L , the educational parameters s_{LR_1, LR_2} and the parameter u_H reflecting (institutional) support for the acquisition of H for L -monolingual newcomers.

In principle, at all levels of the model intervention is conceivable. The state could promote bilingual child raising and provide information such that the rates $q_B(LR)$ increase. Since decision makers can not or should not force individual families to raise children with a certain language repertoire and since socio-economic as well as identity related considerations play a major role here, the effect of such a promotion might be marginal. Much more relevant is education, since it “tends to be the single most important channel of government intervention in the sphere of language” and in “education in general remains a major area of competence of states, typically representing between 10% and 20% of aggregate government expenditure in developed countries” (Grin, 2003, p. 17). Special classes for L -monolingual pupils could increase the fraction of children successfully acquiring the locally dominant language ($s_{L,B}$) and providing bilingual education or at least L -language classes might have the effect that some H -monolingual pupils learn L ($s_{H,B} > 0$). Providing support for the acquisition of H for L -monolingual adults ($u_H > 0$), e.g. in the form of cheap or free language classes, could increase the rate at which adult L -monoglots become bilingual. Granting some official status to the minority language, e.g. providing

administrative information and forms in L , can increase the status of language L . This effects decisions made by individuals and families at all levels of the model. Note, an increase of $S(L)$ can lead to a decrease in the number of L 's learning H , since they can then perform more activities in L . Last, governmental intervention could also influence the linguistic composition of the people entering the population (m_H, m_L, m_B). For example, mandatory language test for newcomers are one instrument to reduce the fraction of L -monolingual migrants.

7 Numerical Examples

The aim of this section is to provide a few numerical examples for the model described above and to illustrate the language competition graphically. We restrict ourselves to the special case of constant absolute migration flow and assume that all newcomers enter the population as L -monoglots ($m_L = 1$).

For intergenerational language transmission, i.e. for the functions $q_{LR}(F)$, we use the same parameter constellation as in the numerical examples of Templin *et al.* (2016): $\zeta = 0.8, \varepsilon = 0.4, \beta = 0.4, \delta = 0.7, \gamma = 0.1$. The yearly birth rate is assumed to be at 1.5% ($\lambda = 0.015$), while the yearly death rate is at 2% ($\mu = 0.02$). For adult language learning we set $\theta = 0.3$ and $\phi = 0.05$. Furthermore, we assume that there is no or only very little support for the acquisition of the newcomer language L by monolingual speakers of the dominant language, i.e. $u_L = 0$. Similarly, for the few numerical examples investigated below it is assumed that H -monolingual pupils do not learn L at school, i.e. $s_{H,B} = 0$. Setting $u_L = 0$ or $s_{H,B} = 0$ would have a negative effect on the relative number of H -monoglots.

Hence, for the numerical examples only four parameters are varying: the status variable $S(L)$, the linguistic concentration measure C , the fraction of pupils entering the school as L -monoglots and additionally learn H during formal education and the support for the acquisition of the locally dominant language H by newcomers u_H .

In Table 2 parameter values for six numerical examples are specified. Additionally, the corresponding steady states are provided. Table 3 displays the distribution of speakers after 50 and 100 years, starting at $X_H(0) = 0.85$ and $X_L(0) = 0.15$ and at the steady state population size. We choose the status of the minority language $S(L)$ in the first four examples to be 0.2, while for the last two examples the status is 0.4 and hence considerably higher. In most examples the linguistic concentration measure C is at $1/2$, 80% of all L -monolingual children learn the dominant language H at school and maximal support for the acquisition of the dominant language by adults is provided ($u_H = 1$). For Example 1, which shall function as a reference case, the steady states consists of 25% H -monoglots, 12% L -Monoglots and 63% bilinguals. Comparing the steady states with the first line of Table 3 it becomes apparent that even after 100 years the linguistic composition is not close to its steady state yet. In Example 2 no support is provided for the acquisition of H by L -monolingual adults. As expected this increases the number of L -monoglots in the steady state by 20%, while after 100 years we only observe an increase of 2.5%. Example 3 shows results for zero linguistic concentration. Due to high rates of exposure to H and extensive contact with H -monoglots, sufficiently many HB families emerge and a relatively large number of children in those and BB families are raised as H -monoglots. This yields a steady state in which almost half of the population is monolingual in H . In contrast, Example 4 illustrates the case of full linguistic concentration. For $C = 1$ no HB couples form and none of the BB families raise their children only in language H . As a result, since $\lambda < \mu$, the H -monolingual population dies out in the long run. The higher status of language L in Example 5 yields, compared with Example 1, a decrease in the number of H -monoglots and an increase in the number of L -monoglots. This can be explained by the higher attractiveness of language L and, respectively, by the lower attractiveness of H . Fewer L -monoglots become bilingual and, at the same time, fewer bilingual parents only transmit H to their

	$S(L)$	C	$s_{L,B}$	u_H	\hat{X}_H	\hat{X}_L	\hat{X}_B
Example 1	0.2	0.5	0.8	1	0.252	0.122	0.626
Example 2	0.2	0.5	0.8	0	0.126	0.327	0.547
Example 3	0.2	0	0.8	1	0.456	0.040	0.504
Example 4	0.2	1	0.8	1	0	0.4	0.6
Example 5	0.4	0.5	0.8	1	0.128	0.210	0.662
Example 6	0.4	0.5	0.5	1	0.060	0.385	0.554

Table 2: Parameter values for six numerical examples and corresponding steady states.

	$X_H(50)$	$X_L(50)$	$X_B(50)$	$X_H(100)$	$X_L(100)$	$X_B(100)$
Example 1	0.681	0.052	0.267	0.562	0.062	0.376
Example 2	0.679	0.069	0.252	0.557	0.085	0.357
Example 3	0.702	0.027	0.271	0.610	0.030	0.360
Example 4	0.664	0.237	0.099	0.517	0.286	0.197
Example 5	0.670	0.066	0.264	0.536	0.079	0.385
Example 6	0.670	0.074	0.256	0.534	0.092	0.374

Table 3: Distribution of speakers after 50 and 100 years for the numerical examples defined in Table 2. The initial linguistic composition is $X_H(0) = 0.85$, $X_L(0) = 0.15$ and $X_B(0) = 0$ and the population size is assumed to be in the steady state.

children. In the last example only half of all L -monolingual pupils learn H at school. This results in a even lower number of H -monolinguals and an increased number of L -monolingual adults.

Looking at the numbers in Table 3 and comparing them with the steady states in Table 2 two important properties of the model can be observed. First, one can see that it takes many years until the linguistic composition reaches its steady state. Hence, for the analysis of short- and medium-term developments – for which the parameters can be assumed to be constant – it is not sufficient just to consider steady states. Second, one can see that in the short- and medium-term a change in one ore more parameters mainly affects the fraction of L -monolingual and bilingual speakers. This can be explained by the fact that few H -monolinguals learn L despite of its status. Furthermore, $s_{L,B}$ and u_H only affect L 's. On top of that it takes some generations for intergeneration language transmission and for low fertility to change the linguistic composition noticeably.

Last, we provide a few trajectories that shall graphically illustrate the results depicted in Tables 2 and 3. In Figure 2 the Examples 1 and 4 are displayed for initial composition $X_H(0) = 0.85$ and $X_L(0) = 0.15$ and for steady state population size. Figure 3 shows trajectories for the same examples with the same initial composition but with changing population size. Here, an initial population of 100 million people and a constant yearly migration inflow of 200,000 L -monolinguals is considered. Compared with the situation in which the total population size is at its steady state (40 million), here the fractions of L -monolgot and bilinguals are increasing much slower. To get a better understanding of how the different language groups evolve in the medium-term, Figure 4 shows the absolute number of speakers for Examples 1 and 4. It can be seen that although the fraction of H -monoglots decreases only slightly (see Figure 3), the absolute number of H -monoglots and with them the overall population size shrinks dramatically over the first 100 years due to the low fertility rate.

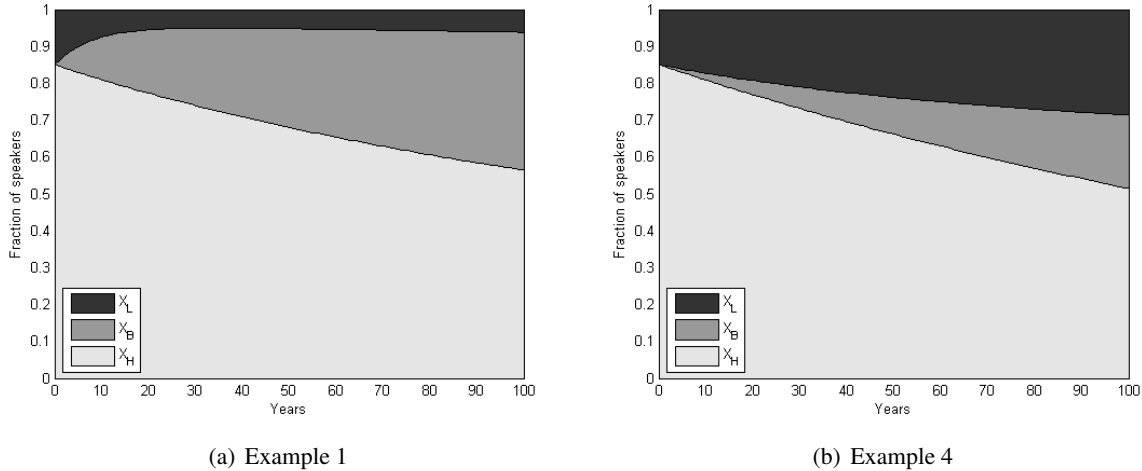
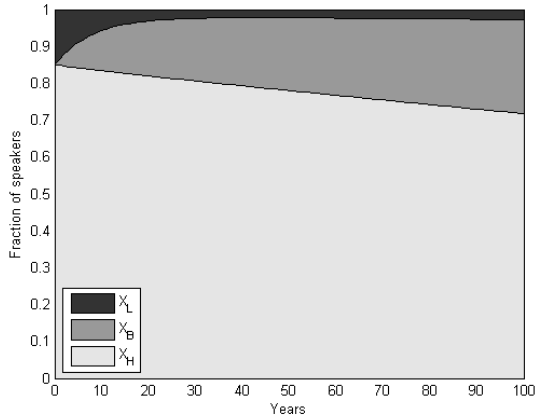


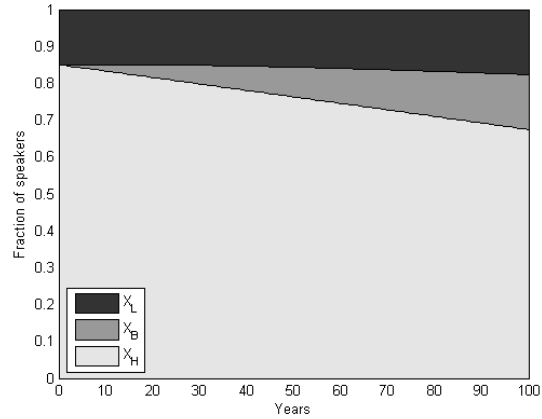
Figure 2: Trajectories for fractions of speakers for Examples 1 and 4 for initial composition $X_H(0) = 0.85$ and $X_L(0) = 0.15$ and for steady state population size.

8 Conclusion and Outlook

In this paper we develop and analyze a language dynamics model for “new minorities” that includes intergenerational language transmission, formal language education and adult language learning. Speakers of the newcomer language L enter a society with one (main) official language H , which is spoken by most society members. To improve their socio-economic position within the host country, some of the newcomers learn the locally dominant language H . The more they get in contact with speakers of H , the higher their exposure to H and consequently the more likely they are to acquire this language. The model includes a parameter measuring linguistic concentration. The more segregated both language groups are, the lower is the exposure to the other language and thus fewer people become bilingual. More important for longer-term language dynamics than adult language learning is intergenerational language transmission and language education for the descendants of newcomers. Individuals form families and transmit one or all of their languages to their children. The decisions on which languages to transmit are driven by instrumental and emotional motives, as well as costs. While the locally dominant language usually promises better socio-economic opportunities for their children, many newcomers want to transmit their heritage language to the next generation. Especially if children live in a social environment dominated by H and if they are only educated in H , then transmitting the minority language might not be successful and after two or three generations the language is lost. Therefore, linguistic concentration supports the transmission of L , but might hinder a proper acquisition of H . We argue that to design a mathematical model, which could be used to compare medium- and long-term outcomes of certain policy options, its variables and parameters have to be measurable in the field. This takes particular regard of the status parameter. On the one hand we follow Fernando *et al.* (2010) in their criticism of the abstract status parameter used in a number of language competition models. But in contrast to them we emphasize the role of factors related to the socio-economic and the institutional status issues for individual language related decisions. Instead of omitting the status from the model we therefore suggest a measurable status parameter composed of the socio-economic status of the language repertoire groups and the institutional status, which is operationalized as the number of official domains the languages can be used for. This yields an indicator of (1) the socio-economic opportunities related to a language, (2) the usefulness of a language in communication with state authorities and (3) the socio-economic resources available to



(a) Example 1

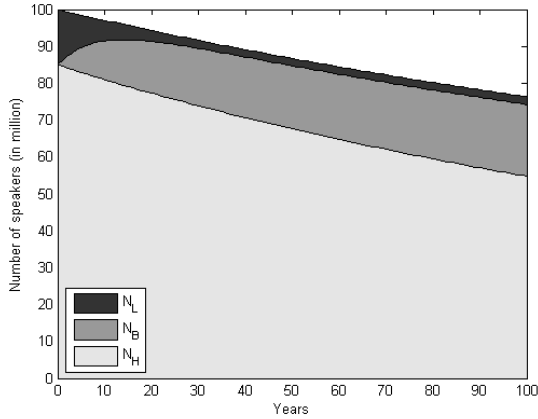


(b) Example 4

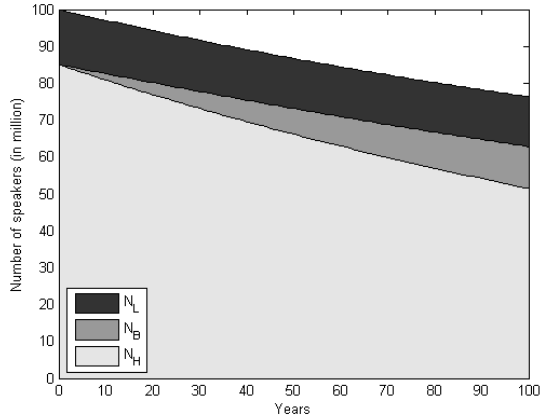
Figure 3: Trajectories for fractions of speakers for Examples 1 and 4. The initial population consists of 100 million people (85 million are H -monoglots and 15 million are L -monoglots) and every year 200,000 new L -monolinguals enter the population.

members of a language repertoire group. Two other important aspects that have to be taken into account are population dynamics and time scales. Using birth, death and migration rates obtained from empirical surveys allows for the model to be closer to real developments. Moreover, for the model to be realistic one has to know how fast things are happening. Furthermore, short- and medium-term analyses are probably more interesting for policy makers than developments over hundreds of years, for which it is, on top of that, hard to argue for parameters to be constant.

The analysis of long-term dynamics shows that monolingual steady states are only possible for fully segregated populations. Due to the continuous influx of L -speakers on the one hand and the relatively high status of H on the other hand, steady states with only one language group are highly unlikely. Much more likely are, however, steady states with monolinguals of both languages as well as bilinguals. This can be seen in Table 2, where steady states are calculated for different parameter constellations characterizing the linguistic environment. In all cases, except for Example 4, for which the H -monolingual population dies out in the long-run as a result of linguistic concentration and low fertility rates, the steady state consists of monolinguals of both languages as well as bilinguals. Table 3, in which the linguistic composition after 50 and 100 years is depicted for all the numerical examples, illustrates the insufficiency of an analysis concentrating only on steady states. Although steady states differ notably for the six examples, the difference after 50 and even 100 years is not as dramatic. It can be seen that during the first 50 years the H -monolingual population is barely affected by language policy measures targeting speakers of L . Since the “original” H -monolingual population dies out (low fertility) and “new” H -speakers have L -descendants, effects of such policies on X_H can only be observed in the long-run. For the fractions of L -monolinguals and bilinguals the short- and medium-term effects of changes in relevant model parameters are more significant. Here, in particular, the role of linguistic concentration must be emphasized. For maximal linguistic concentration ($C = 1$) after 50 years the fraction of L -monoglots is almost five times higher compared with the case of moderate linguistic concentration ($C = 0.5$). The effects of changes in model parameters related to language education and support for adult language learning are not as strong but can be observed as well. The same holds true for changes in the status parameter. In Section 7 we furthermore illustrate that neglecting population dynamics and assuming a steady state population size might have a strong impact on the analysis of short- and medium-term developments. If the overall



(a) Example 1



(b) Example 4

Figure 4: Trajectories for absolute number of speakers for Examples 1 and 4. The initial population consists of 100 million people (85 million are H -monoglots and 15 million are L -monoglots) and every year 200,000 new L -monolinguals enter the population.

population size is shrinking towards its steady state, then the influx of L 's has a weaker effect on the linguistic composition of the population compared to the steady state case. At the same time, the absolute number of H -monoglots shrinks rapidly, while their relative number shrinks much slower. This shows that relative as well as absolute numbers should be considered in order to describe fully the population's linguistic development.

We see this paper as a point of departure for future research on language dynamics in societies with new language minorities. To get closer to the complexity of the processes at hand, the model can be extended and refined in several ways. First, in the current version of the model we only consider one language minority, while normally one can observe a variety of minority language groups. In the case of multiple linguistic minorities the locally dominant language H is not only a mean of upward social mobility, but can also function as a vehicular language used in communication between speakers of different minority languages. Second, different levels of linguistic competences or actual language use (instead of proficiency) could be modeled. Third, the current model concentrates on permanent migration and mostly neglects mobile individuals which stay only temporarily and leave the country after a couple of years. Particularly in view of highly skilled workers or international students, who frequently belong to this category, a model taking into account international vehicular languages that are (partially) spoken by locals as well as newcomers – as in the case of English – would be of interest.

So far the model allows us to simulate the future effects of changes in the linguistic environment, which might be the result of language planning policies. For the model to be used in policy analysis, the costs of language policies have to be included. If estimates on (per-capita) costs are available, then the costs and outcomes of certain language policies can be compared, for example in the form of cost-effectiveness analyses. Furthermore, if benefits are defined and operationalized, for example as the propensity to pay for being able to communicate with state authorities in ones heritage language, then the model could also be used for cost-benefit analyses.

References

- Abrams, Daniel M., & Strogatz, Steven H. 2003. Linguistics: Modelling the dynamics of language death. *Nature*, **424**(6951), 900.
- Baggs, I., & Freedman, H. I. 1993. Can the speakers of a dominated language survive as unilinguals?: a mathematical model of bilingualism. *Mathematical and computer modelling*, **18**(6), 9–18.
- Baker, Colin. 2011. *Foundations of bilingual education and bilingualism*. Bilingual education and bilingualism, vol. 54. Bristol, Buffalo, Toronto, Sydney: Multilingual matters.
- Barron-Hauwaert, Suzanne. 2004. *Language strategies for bilingual families: The one-parent-one-language approach*. Clevedon, Buffalo, Toronto: Multilingual Matters.
- Bernard, H. Russell. 2012. *Social Research Methods: Qualitative and Quantitative Approaches*. 2. edn. SAGE Publications.
- Caldas, Stephen J. 2012. Language policy in the family. *Pages 351–373 of: Spolsky, Bernard (ed), The Cambridge Handbook of Language Policy*. Cambridge: Cambridge University Press.
- Canagarajah, A. Suresh. 2008. Language shift and the family: Questions from the Sri Lankan Tamil diaspora. *Journal of Sociolinguistics*, **12**(2), 143–176.
- Chiswick, Barry R., & Miller, Paul W. 2002. *Do enclaves matter in immigrant adjustment?* Tech. rept. IZA-DP-449. Institute for the Study of Labor.
- Chiswick, Barry R., & Miller, Paul W. 2014. International migration and the economics of language. *Handbook of the Economics of International Migration*. Amsterdam: Elsevier, 211–269.
- Church, Jeffrey, & King, Ian. 1993. Bilingualism and network externalities. *Canadian Journal of Economics*, 337–345.
- Cormack, Michael J. 2007. The Media and Language Maintenance. *Pages 52–68 of: Cormack, Michael J., & Hourigan, Niamh (eds), Minority Language Media: Concepts, Critiques and Case Studies*. Clevedon, Buffalo, Toronto: Multilingual Matters.
- Danzer, Alexander M., & Yaman, Firat. 2010. Immigranten in Deutschland: Ethnische Enklaven schwächen die Sprachkompetenz, mehr Bildung stärkt sie. *IAB-Kurzbericht*, **17**, 2010.
- De Houwer, Annick. 2007. Parental language input patterns and children's bilingual use. *Applied psycholinguistics*, **28**(03), 411–424.
- Elias, Nelly, & Lemish, Dafna. 2011. Between three worlds: Host, homeland, and global media in the lives of Russian immigrant families in Israel and Germany. *Journal of Family Issues*, **32**(9), 1245–1274.
- Esser, Hartmut. 2006. *Migration, Sprache und Integration*. Berlin: WZB Berlin.
- Eurostat. 2014. *Fertility statistics*.
http://ec.europa.eu/eurostat/statistics-explained/index.php/Fertility_statistics#Main_statistical_findings.
Accessed: 2015-08-27.

- Eurostat. 2014. *Migration and migrant population statistics*.
http://ec.europa.eu/eurostat/statistics-explained/index.php?title=Migration_and_migrant_population_statistics&oldid=216019.
 Accessed: 2015-08-27.
- Fernando, Chrisantha, Valijärvi, Riitta-Liisa, & Goldstein, Richard A. 2010. A model of the mechanisms of language extinction and revitalization strategies to save endangered languages. *Human biology*, **82**(1), 47–75.
- Fishman, Joshua A. 1991. *Reversing language shift: Theoretical and empirical foundations of assistance to threatened languages*. Vol. 76. Clevedon: Multilingual matters.
- Gabszewicz, Jean, Ginsburgh, Victor, & Weber, Shlomo. 2011. Bilingualism and Communicative Benefits. *Annals of Economics and Statistics*, 271–286.
- Gazzola, Michele, Grin, François, & Vaillancourt, François. 2016. Evaluating Language Policy and Planning: An Introduction to the Economic Approach. *Working paper*.
- Giles, Howard, Bourhis, Richard Y., & Taylor, Donald M. 1977. Towards a theory of language in ethnic group relations. *Pages 307–384 of: Giles, Howard (ed), Language, ethnicity and intergroup relations*. London: Academic Press.
- Ginsburgh, Victor, Ortuño-Ortín, Ignacio, & Weber, Shlomo. 2007. Learning foreign languages: Theoretical and empirical implications of the Selten and Pool model. *Journal of Economic Behavior & Organization*, **64**(3), 337–347.
- Gong, Tao, Shuai, Lan, & Zhang, Menghan. 2014. Modelling language evolution: Examples and predictions. *Physics of Life Reviews*, **11**(2), 280–302.
- Grin, François. 1992. Towards a threshold theory of minority language survival. *Kyklos*, **45**(1), 69–97.
- Grin, François. 1999. *Compétences et récompenses: la valeur des langues en Suisse*. Editions universitaires.
- Grin, François. 2003. Language planning and economics. *Current issues in language planning*, **4**(1), 1–66.
- Grin, François, & Vaillancourt, François. 1997. The economics of multilingualism: Overview and analytical framework. *Annual review of applied linguistics*, **17**, 43–65.
- Heinsalu, E., Patriarca, M., & Léonard, J. L. 2014. The role of bilinguals in language competition. *Advances in Complex Systems*, **17**(01), 1450003.
- Hornberger, Nancy H., & May, Stephen (eds). 2010. *Language Policy and Political Issues in Education*. Encyclopedia of Language and Education 2nd Edition, vol. 1. New York: Springer.
- Iriberry, Nagore, & Uriarte, José-Ramón. 2012. Minority language and the stability of bilingual equilibria. *Rationality and Society*, **24**(4), 442–462.
- John, Andrew. 2016. Dynamic Models of Language Evolution: The Economic Perspective. *Pages 101–120 of: Ginsburgh, Victor, & Weber, Shlomo (eds), The Palgrave Handbook Economics and Language*. New York: Palgrave Macmillan.
- John, Andrew, & Yi, Kei-Mu. 1996. Language, learning, and location. *Thomas Jefferson Center for Political Economy Discussion Paper*, **264**. Reprinted in: Federal Reserve Bank of New York Staff Report No. 26 (1997).

- John, Andrew, & Yi, Kei-Mu. 2001. Language and Location. *Unpublished Manuscript*.
- Kandler, Anne, & Steele, James. 2008. Ecological models of language competition. *Biological Theory*, **3**(2), 164–173.
- Kandler, Anne, Unger, Roman, & Steele, James. 2010. Language shift, bilingualism and the future of Britain's Celtic languages. *Philosophical Transactions of the Royal Society B: Biological Sciences*, **365**(1559), 3855–3864.
- Kennedy, Peter, & King, Ian. 2005. *Dynamic Language Policy*. Unpublished Manuscript.
- King, Kendall A., Fogle, Lyn, & Logan-Terry, Aubrey. 2008. Family Language Policy. *Language and Linguistics Compass*, **2**(5), 907–922.
- Krauss, Michael. 1992. The world's languages in crisis. *Language*, **68**(1), 4–10.
- Massey, Douglas S., & Denton, Nancy A. 1988. The dimensions of residential segregation. *Social forces*, **67**(2), 281–315.
- Minett, James W., & Wang, William SY. 2008. Modelling endangered languages: The effects of bilingualism and social structure. *Lingua*, **118**(1), 19–45.
- Mira, Jorge, & Paredes, Ángel. 2005. Interlinguistic similarity and language death dynamics. *EPL (Europhysics Letters)*, **69**(6), 1031.
- Morrill, Richard L. 2016. On the measure of geographic segregation. *Pages 25–36 of: Geography research forum*, vol. 11.
- Okita, Toshie. 2002. *Invisible work: Bilingualism, language choice and childrearing in intermarried families*. Vol. 12. Amsterdam, Philadelphia: John Benjamins Publishing.
- Patriarca, Marco, & Heinsalu, Els. 2009. Influence of geography on language competition. *Physica A: Statistical Mechanics and its Applications*, **388**(2), 174–186.
- Patriarca, Marco, & Leppänen, Teemu. 2004. Modeling language competition. *Physica A: Statistical Mechanics and its Applications*, **338**(1), 296–299.
- Patriarca, Marco, Castelló, Xavier, Uriarte, J. R., Eguíluz, Víctor M., & San Miguel, Maxi. 2012. Modeling two-language competition dynamics. *Advances in Complex Systems*, **15**(03n04).
- Piller, Ingrid. 2001. Private language planning: The best of both worlds. *Estudios de Sociolingüística*, **2**(1), 61–80.
- Piller, Ingrid. 2002. *Bilingual couples talk: The discursive construction of hybridity*. Vol. 25. Amsterdam, Philadelphia: John Benjamins Publishing.
- Pinasco, Juan Pablo, & Romanelli, Liliana. 2006. Coexistence of languages is possible. *Physica A: Statistical Mechanics and its Applications*, **361**(1), 355–360.
- Ruggeri, Kai, & Bird, Chloe E. 2014. *Single parents and employment in Europe*. Short statistical report 13.
- Sallabank, Julia. 2012. Diversity and language policy for endangered languages. In: Spolsky, Bernard (ed), *The Cambridge handbook of language policy*. Cambridge: Cambridge University Press.

- Schulze, Christian, Stauffer, Dietrich, & Wichmann, Søren. 2008. Birth, survival and death of languages by Monte Carlo simulation. *Communications in Computational Physics*, **3**(2), 271–294.
- Schwartz, Mila. 2008. Exploring the relationship between family language policy and heritage language knowledge among second generation Russian–Jewish immigrants in Israel. *Journal of Multilingual and Multicultural Development*, **29**(5), 400–418.
- Selten, Reinhard, & Pool, Jonathan. 1991. The distribution of foreign language skills as a game equilibrium. *Pages 64–87 of: Game equilibrium models IV*. Springer.
- Sleebos, Joëlle. 2003. *Low Fertility Rates in OECD Countries*. OECD Labour Market and Social Policy Occasional Papers. Organisation for Economic Co-operation and Development, Paris.
- Spolsky, Bernard. 1989. Communicative competence, language proficiency, and beyond. *Applied Linguistics*, **10**(2), 138–156.
- Spolsky, Bernard. 2004. *Language policy*. Cambridge: Cambridge University Press.
- Stauffer, Dietrich, Castelló, Xavier, Eguiluz, Victor M., & San Miguel, Maxi. 2007. Microscopic Abrams–Strogatz model of language competition. *Physica A: Statistical Mechanics and its Applications*, **374**(2), 835–842.
- Tamura, Robert. 2001. Translators: Market makers in merging markets. *Journal of Economic Dynamics and Control*, **25**(11), 1775–1800.
- Templin, Torsten, Seidel, Andrea, Wickström, Bengt-Arne, & Feichtinger, Gustav. 2016. Optimal language policy for the preservation of a minority language. *Mathematical Social Sciences*, **81**, 8–21.
- Tuominen, Anne. 1999. Who decides the home language? A look at multilingual families. *International journal of the sociology of language*, **140**(1), 59–76.
- United Nations. 2013. *International Migration Report 2013*. Tech. rept. United Nations, Department of Economic and Social Affairs, Population Division, New York.
- Van Tubergen, Frank, & Kalmijn, Matthijs. 2005a. Destination-Language Proficiency in Cross-National Perspective: A Study of Immigrant Groups in Nine Western Countries. *American Journal of Sociology*, **110**(5), 1412–1457.
- Van Tubergen, Frank, & Kalmijn, Matthijs. 2005b. Destination-Language Proficiency in Cross-National Perspective: A Study of Immigrant Groups in Nine Western Countries. *American Journal of Sociology*, **110**(5), 1412–1457.
- Wickström, Bengt-Arne. 2005. Can Bilingualism be Dynamically Stable?: A Simple Model of Language Choice. *Rationality and Society*, **17**, 81–115.
- Zhang, Menghan, & Gong, Tao. 2013. Principles of parametric estimation in modeling language competition. *Proceedings of the National Academy of Sciences*, **110**(24), 9698–9703.
- Zhang, Weiguo, & Grenier, Gilles. 2013. How can language be linked to economics?: A survey of two strands of research. *Language Problems & Language Planning*, **37**(3), 203–226.

A Appendix

A.1 Measuring segregation

Segregation is a “multidimensional phenomenon varying along five distinct axes of measurement: evenness, exposure, concentration, centralization, and clustering” (Massey & Denton, 1988, p. 281). This is one reason why in the literature a number of different indices for spatial segregation can be found. It is out of the scope of this paper to give a detailed overview on the discussion of how to measure segregation. Instead, we refer the interested reader to the literature, see e.g. Massey & Denton (1988) or Morrill (2016), and just provide three well-known examples of segregation indices.

Within a given territory we consider two distinct groups Y and Z . Let N_Y be the total number of Y 's and N_Z be the total number of Z 's. Let $N := N_Y + N_Z$. Furthermore, the territory is divided into K areas (sub-territories). For $k = 1, \dots, K$ let $N_{Y,k}$ (resp. $N_{Z,k}$) denote the number of Y 's (resp. Z 's) in area k . Finally, set $N_k := N_{Y,k} + N_{Z,k}$.

First, we consider the *classical* index of dissimilarity ID :

$$ID = \frac{1}{2} \sum_{k=1}^K \left| \frac{N_{Y,k}}{N_Y} - \frac{N_{Z,k}}{N_Z} \right|.$$

The index of dissimilarity can be interpreted as the proportion of the minority population Z that has to move such that all K areas have the same proportion of Z 's. One problem with the ID is that it is a structural measure of evenness across the K areas, “not of segregation in a geographic sense” (Morrill, 2016, p. 27). For example, the measure is unable to differentiate between cases where areas with a high Z concentration are close to each other or far apart. This is well illustrated in Morrill (2016).

Second, we consider an index of isolation:

$${}_Z II_Z = \sum_{k=1}^K \frac{N_{Z,k}}{N_Z} \frac{N_{Z,k}}{N_k}.$$

This index measures the percentage of Z 's in the own geographical area for the average Z -speaker. While the ID is symmetrical, in general ${}_Z II_Z$ and ${}_Y II_Y$ do not yield the same results. This way it can e.g. be displayed that as the majority Y 's are more segregated from Z 's than the other way around.

Third, we consider an index of exposure:

$${}_Z IE_Y = \sum_{k=1}^K \frac{N_{Z,k}}{N_Z} \frac{N_{Y,k}}{N_k} = 1 - {}_Z II_Z.$$

This (second) index of exposure (or interaction index) measures the extend to which average Z 's are exposed Y 's. As the index of dissimilarity, the measures ${}_Z II_Z$ and ${}_Z IE_Y$ do neither account for the geographical sizes of the areas nor their distance to one another. Nonetheless, they are able segregation and expose to some extend and are relatively easy to obtain from census-like data. In the literature spatial and other extensions of these basic indices are provided, see e.g. Morrill (2016).

A.2 Family formation

Consider a population of size $2N$, where N is large. Suppose that the population consists of N female and N male individuals. Let N_H denote the number of female H -monolinguals, N_L the number of female L -monolinguals and N_B the number of female bilinguals. We assume that these numbers are the same for the male population. The distribution of speakers is given by $X_H = N_H/N$, $X_L = N_L/N$ and $X_B = N_B/N$.

Family formation can be conceptualized as a repeated random procedure of choosing pairs. Let us start with one randomly chosen pair consisting of a female Y and a male Z . The probabilities that a certain pair is chosen depends on the numbers of speakers as well as on the linguistic concentration. For maximal concentration, i.e. $C = 1$, H monolinguals will only meet other H monolinguals. Let us start with a pair consisting of two H monolingual adults. The probability for such a pair is given by

$$\begin{aligned}\mathbb{P}[Y = H, Z = H] &= \mathbb{P}[Y = H]\mathbb{P}[Z = H|Y = H] \\ &= \mathbb{P}[Y = H](C + (1 - C)\mathbb{P}[Z = H]) \\ &= X_H(C + (1 - C)X_H)\end{aligned}\tag{26}$$

Next, we consider pairs with one H monoglot and one adult who (also) speaks L . Let $LR \in \{L, B\}$. Then

$$\begin{aligned}\mathbb{P}[Y = H, Z = LR] &= \mathbb{P}[Y = H]\mathbb{P}[Z = LR|Y = H] \\ &= \mathbb{P}[Y = H]((1 - C)\mathbb{P}[Z = LR]) \\ &= (1 - C)X_HX_{LR}\end{aligned}\tag{27}$$

Finally, we consider pairs where both adults speak L . Let $LR_1, LR_2 \in \{L, B\}$. Then

$$\begin{aligned}\mathbb{P}[Y = LR_1, Z = LR_2] &= \mathbb{P}[Y = LR_1]\mathbb{P}[Z = LR_2|Y = LR_1] \\ &= \mathbb{P}[Y = LR_1]\mathbb{P}[Z = LR_2] \left(1 + C \left(\frac{1}{\mathbb{P}[Z = L] + \mathbb{P}[Z = B]} - 1\right)\right) \\ &= X_{LR_1}X_{LR_2} \left(1 + C \frac{X_H}{1 - X_H}\right)\end{aligned}\tag{28}$$

Note that a family type LR_1LR_2 , $LR_1 \neq LR_2$, is obtained either by $Y = LR_1$; $Z = LR_2$ or by $Y = LR_2$; $Z = LR_1$ (for the family type we do not take the gender of the parents into account). Hence, the probability of obtaining a pair of type LR_1LR_2 , $LR_1, LR_2 \in \{H, L, B\}$ is given by $\mathbb{P}[Y = LR_1, Z = LR_2]$, if $L_1 = L_2$, and by $2\mathbb{P}[Y = LR_1, Z = LR_2]$, if $L_1 \neq L_2$.

All N pairs are chosen randomly after one another. The total expected number of LR_1LR_2 -type pairs equals $N \cdot \mathbb{P}[LR_1LR_2]$ and hence the expected fraction of LR_1LR_2 -type pairs is $\mathbb{P}[LR_1LR_2]$. We assume that all couples of types other than HL form families. Recall, we assume that parents shall be able to properly communicate with each other, and therefore we exclude HL families. Hence, after this first step we have N pairs with $2N(1 - C)X_HX_L$ of them being of type HL . Splitting these HL pairs again and repeating the random selection we obtain new pairs of types HH , LL and HL . This procedure is repeated until only HH and LL pairs remain. This way, half of the $2N(1 - C)X_HX_L$ pairs of type HL will be transformed into HH pairs, while the other half will form LL pairs. As a result, we obtain the numbers presented in equations (9)-(13). Note, due to the law of large numbers (N is assumed to be large), the realized number of LR_1LR_2 -type of pairs can be approximated by the expected number.

A.3 Functional expressions for $\tilde{q}_L(F)$

For non-negative parameters $0 \leq \eta < \beta < \delta$ and $\varepsilon + \gamma < \zeta < 1$ consider the functions

$$\begin{aligned}\tilde{q}_H(HH, S; X) &:= 1 - \eta S(L)X_L \\ \tilde{q}_H(HB, S; X) &:= \max\{0, \zeta S(H) - \beta S(L)X_L\} \\ \tilde{q}_H(BB, S; X) &:= \max\{0, \varepsilon S(H) + \gamma S(H)X_H - \delta S(L)X_L\} \\ \tilde{q}_L(LL, S; X) &:= 1 - \eta S(H)X_H \\ \tilde{q}_L(LB, S; X) &:= \max\{0, \zeta S(L) - \beta S(H)X_H\} \\ \tilde{q}_L(BB, S; X) &:= \max\{0, \varepsilon S(L) + \gamma S(L)X_L - \delta S(H)X_H\}.\end{aligned}$$

Note, $S(H) = 1 - S(L)$.

A.4 The functions $q_L(F)$

For $C \in [0, 1]$ the functions $q_L(F, S, C; X)$ are given by

$$\begin{aligned}q_H(HH, S, C; X) &= 1 \\ q_H(HB, S, C; X) &= (1 - C) \cdot \tilde{q}_H(HB, S; X) + C/2 \\ q_H(BB, S, C; X) &= (1 - C) \cdot \tilde{q}_H(BB, S; X) \\ q_L(LL, S, C; X) &= 1 \\ q_L(LB, S, C; X) &= (1 - C) \cdot \tilde{q}_L(LB, S; X) + C \\ q_L(BB, S, C; X) &= (1 - C) \cdot \tilde{q}_L(BB, S; X) + C.\end{aligned}$$