universität wien

DISSERTATION / DOCTORAL THESIS

Titel der Dissertation / Title of the Doctoral Thesis

Diffusion modeling of language shift in Austria(-Hungary)

verfasst von / submitted by

Katharina Prochazka, BA BSc MA MSc

angestrebter akademischer Grad / in partial fulfilment of the requirements for the degree of

Doktorin der Naturwissenschaften (Dr. rer. nat.)

Wien, 2019 / Vienna 2019

Studienkennzahl lt. Studienblatt / degree programme code as it appears on the student record sheet:	A 796 605 411
Dissertationsgebiet lt. Studienblatt / field of study as it appears on the student record sheet:	Physik
Betreut von / Supervisor:	emer. o. Univ.-Prof. Dr. Gero Vogl Mag. Dr. Manfred Michael Glauninger, Privatdoz. Univ.-Prof. Mag. Dr. Christoph Dellago

List of publications 7

A note on names 9

1 Introduction: Now you use a language, now you don't 11
1.1 Research directions/paradigms/traditions 13
1.2 Aims of this thesis 15
1.3 Structure of this thesis / Road map to the contents 16

2 Motivation and background 18
2.1 Linguistics: Why do people shift their language? 18
2.2 Physics: Can we use physics in linguistic research? 23
2.3 The analogy between diffusion of matter and of languages: Are languages like atoms? 25

3 Methods: Modeling language shift / How do we use physics in (this) linguistic research? 29
3.1 Mathematical models as a tool 29
3.2 Modeling language shift as a physicist 30
3.2.1 Models of physical diffusion 31
3.2.2 Models of language spread/shift 32
Reaction-diffusion equations 32
Agent-based models and cellular automata 34
3.3 The atoms of language 35
3.4 Incorporating social and psychological influence factors 36

4 Data material 39
4.1 Methods of collecting data on language use 40
4.1.1 Censuses 41
4.1.2 Non-governmental surveys 41
4.1.3 Other statistics on language use 42
4.2 Limits on the informative value of language surveys 42
4.2.1 What is being counted? 42
4.2.2 Who asks? 43
4.2.3 What do the results tell us? 44
4.3 Legal provisions in connection with language surveys 44
4.4 Mixing language and ethnicity in surveys 46
4.5 Criticism of language surveys by minorities 48
4.6 Data on language use in this thesis 49

5 **The lab: Language contact in Austria(-Hungary)** **53**

6 **Comparative data evaluation: What consequences do local conditions have for modeling?** **56**

6.1 Historical background: multilingual Austria-Hungary / Setting the data in context 56

6.2 Influence of language policy 61

6.3 Influence of settlement and population structure 64

7 **Situation I: Slovenian–German in Carinthia** **69**

7.1 Historical background 69

7.2 Preparing the data 71

7.3 Macroscopic view: The language border 73

7.4 Microscopic view: Grid cells and cellular automata 79

7.4.1 Judging the model: Simulation and evaluation procedure 83

Reliability of the model per category 85

Total number of Slovenian speakers 88

Deviation of simulated data from census data over space . 89

7.4.2 Results 89

Language shift in the period 1880–1910 89

Extension of the model through habitat parameters 92

Urban centers 92

Bilingual schools 93

Parishes 93

Language shift in the period 1971–2001 93

Local differences in the language diffusivities 95

7.4.3 Discussion 95

7.5 Preliminary conclusion 97

8 **Situation II: German–Hungarian in southern Hungary** **100**

8.1 Historical background 100

8.2 Preparing the data 103

8.3 Applying the microscopic language shift model 105

9 **Discussion: On different language shift situations and models thereof** **111**

10 **At the end: Summary, lessons learned and outlook** **115**

Interactive maps **122**

Sources and references	**123**
List of figures	**140**
List of tables	**143**
Acknowledgements	**144**
English abstract	**147**
Deutsche Zusammenfassung	**148**

List of publications

This thesis is based on the following papers and manuscripts:

[1] Prochazka, Katharina/Vogl, Gero (2017): Quantifying the driving factors for language shift in a bilingual region. *Proceedings of the National Academy of Sciences* 114 (17): 4365–4369.

[2] Prochazka, Katharina/Vogl, Gero (2018): Are languages like atoms? On modelling language spread as a physicist. *Glottotheory* 9 (1): 77–88.

[3] Prochazka, Katharina (2018): Minderheitensprachen zählen! Über Sprachzählungen und Minderheiten(-sprachen). *Wiener Linguistische Gazette* 83: 1–26.

[4] Prochazka, Katharina (in print): Sprachwechsel in Südkärnten: Quantitative Beschreibung und Modellierung als Diffusionsprozess. In: Bülow, Lars/Fischer, Ann-Kathrin/Herbert, Kristina (eds.), *Dimensionen des sprachlichen Raums: Variation – Mehrsprachigkeit – Konzeptualisierung*. Wien: Peter Lang.

As the first or sole author of these publications, I was responsible for designing research, collecting and analyzing data, performing computer simulations and writing manuscripts. Publications [1] and [2] were co-written with Gero Vogl who advised the project throughout and provided input at all stages.

Verbatim quotation of passages is indicated at the beginning of the respective section. Numbering of headings, figures, equations and tables has been adapted in quoted passages.

A note on names

As we are dealing with language contact, i. e. multiple language on the same territory, most of the places mentioned in this thesis have distinct names in multiple languages. The German Klagenfurt is Celovec in Slovenian, the Hungarian Pécs is Fünfkirchen in German and so on. Throughout this thesis, I use the English versions of place names (e. g. "Carinthia" instead of "Kärnten" or "Koroška")—and in many cases, especially for smaller villages, these correspond to the equivalent in the majority language (German or Hungarian), which is unfortunate considering the topic at hand. I have therefore generally included the name in both languages wherever this is the case.

*

There is much debate on what the "correct" term in English is: "Slovene" or "Slovenian". The convention I am using is this:

- I use "Slovenian" when referring to the language (*Slovenščina*) and its dialects, regardless of whether the language/dialect is spoken on the territory of today's Republic of Slovenia (Republika Slovenija) or not. I see no benefit in making a further distinction for this research project which is concerned with the Slovenian language in Carinthia, outside of Slovenia (and the data used does not differentiate between its dialects anyhow).
- When referring to people, I take "Slovenian" to be anyone living in or having the nationality of the Republic of Slovenia, regardless of their ethnic background or self-identification. "Slovene" refers to the ethnic group of Slavic origin whose members do not necessarily live on the territory of Slovenia—and do not necessarily use the Slovenian language for all of their communication. (Obviously, any classification of people according to categories which have a subjective component such as ethnicity is fraught with difficulties, so this is more a loose working definition than an exact anthropological one.) The ethnic group of people in Carinthia is consequently designated "Carinthian Slovenes" or "Slovene minority".

*

The ethnic group in Hungary primarily descended from German-speaking settlers (see section 8.1) is alternatively referred to as "Hungarian Germans", "German Hungarians", or "Danube Swabians" in English texts (with slight nuances in interpretation). In accordance with the self-designation in German, *Ungarndeutsche*, I use the term "Hungarian Germans", which is a direct translation. Again, Hungarian Germans can have varying levels of actual German language use.

1 Introduction: Now you use a language, now you don't

Imagine you are a physicist working in materials science, studying the way material things move. Now imagine asking yourself if immaterial things move the same way, if language behaves similar to atoms and particles.

*

Imagine you use a language in your everyday life (which you probably do already). Now imagine you gradually stop using this language, and start using another language.

Of course, it might not only be you who stops using this language. On a large scale, the result of such a language shift might look like figure 1, where we see the percentage of people using Slovenian in everyday life according to the Austrian census, in 1880 and 2001. The lighter the color, the fewer people use Slovenian. Comparing both figures, the difference is clearly visible; it seems that people have shifted from one language to another on a level far beyond one individual person.

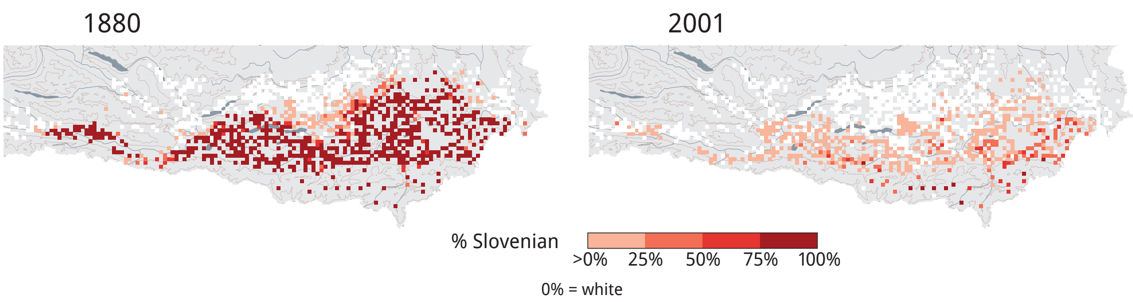


Figure 1: Percentage of "Slovenian" in census results in southern Carinthia, Austria

This, then, is what this thesis is about: Why do these changes in language use happen? Why do people stop using a language? In other words, how do we get from the left to the right picture in figure 1?

Languages are seen by many people as a central part of culture and important for their identity (even if the connection of identity with language is highly individual and takes many forms). Changes in language use can therefore have far-reaching consequences, so there is an interest to understand the process of people changing their language. There are many ways to answer the underlying question (why are some languages being used less and less?) and a number of previous (mostly linguistic) studies have looked into why people change their language. However, few large-scale studies following changes in language use over many years and over a large space exist. I have therefore chosen an alternate path: using physics to study language shift on a large scale over time and space. Why physics? Physics and physical (mathematical) models provide a way to handle a large amount of data and uncover general trends by looking at the bigger picture. Using physics to explore a linguistic research question opens up new ways of looking at existing data which ultimately may help us understand language shift better—and why people stop using a language. It can also give us a better idea on how methods from physics can be applied in other disciplines. The problem is therefore not only a linguistics problem, but just as much a physics problem.

Within this thesis, I develop a mathematical model for language shift to identify the factors driving changes in language use. The model is tested by taking a closer look at two cases of language shift in two specific language contact situations. Language contact situations are situations somewhere in the world—in this case, in Austria and Hungary—where multiple languages and language varieties are being used at the same time in the same place. The different languages and their speakers interact and influence each other over time. This influence can mean that a language's system changes (e. g. new linguistic forms emerge through borrowing, words are transferred from one language to the other and so on), that new languages are created or that people ultimately stop using one language, as in language shift.[1] It is language shift which will be the focus of the research presented in the following.

Up to this point I have written about "language" and "languages" as if both were clearly defined objects by which we all understand the same thing. It will not surprise that this is not the case; that "language" means different things to different people. In research based on empirical data such as this study, the concept of language is construed by the data used: whatever is recorded as "a language" in the data set is a language for the purpose of modeling. In the case of language shift research, it is implied that we work with different languages (rather than different varieties of one language). If I talk about "language use", then this is restricted to "use of which language", i. e. the use of *one* language, not *how* one language and its inner parts or structures are used. Similarly, "people changing their language (use)" refers to people starting use of a different language (shifting from one language to another) rather than people changing a language from within (or changing features such as pronunciation or syntax of one language).

---

1 As noted in section 2.1, the process of "stopping use of one language in favor of another" is more complex than presented in this brief introduction.

1.1 Research directions/paradigms/traditions

Using physics to study language shift classifies this project as an interdisciplinary one merging linguistics and physics, or more broadly humanities/social sciences[2] and natural sciences.

Interdisciplinary (sometimes also called transdisciplinary, crossdisciplinary, multidisciplinary, ...) research can mean different things: Either simply using research methods from one field to explore questions of another field, or combining techniques from two or more fields to answer a question which is beyond the scope of one field. Interdisciplinary research approaches uncover new possibilities, provide new tools and create new insights. And of course, it can create new problems, both in the sense of "new questions to answer" as well as in "obstacles and complications".

Interdisciplinary research does not only mean negotiation of differing stylistic conventions and separate "research cultures" , it can also mean different research paradigms.[3] Physics as a discipline is based in the quantitative research tradition, where things are measured and labeled and put into numbers. More specifically, it employs mathematical tools to evaluate data collected from empirical observations and tries to generalize the results into laws which are expressed as equations. Physics is therefore often seen as an exact science, notwithstanding the fact that measurements can only be as exact as the measurement uncertainty and that physical laws are only true until they are falsified.

Linguistics draws from both quantitative and qualitative research methods [6; 7], depending on the sub-field and the research question–nowadays, both approaches are usually combined to cross-check results (triangulation). Qualitative research includes not only data collection or empirical observations, but also the personal interpretation and discussion of these data sets by the researcher in non-statistical ways—which means the sample size is typically smaller than in quantitative research.[4] In contrast to qualitative methods, the data need not be collected in a

2 There is some debate as to whether linguistics as a discipline is part of the humanities or rather a social science. Depending on the sub-field and the concept of "language" used, the classification is nearer to one or the other. In the case of language shift and sociolinguistics (language in society), the discipline might be seen as more of a social science. Some sub-fields of linguistics e. g. neurolinguistics (the study of language mechanisms in the brain) have close ties to the natural sciences as well.

3 There has long been debate on how science is performed across different fields and especially on the divide between the humanities and the natural sciences, see e. g. [5].

4 Strictly speaking, "quantitative" may be taken to mean: anything which can be measured and put into numbers i. e. which does not rely only on introspection inaccessible to the researcher. In this sense, a survey of ten people about their mother tongue also counts as quantitative research (though a physicist might beg to differ). However, a sample this small cannot in good conscience be used to draw statistical conclusions. "Quantitative" therefore usually implies a larger amount of data involved than in qualitative research. This is also the way I will be using the term "quantitative": involving a large amount of data. For practical reasons, the sample size even in quantitative linguistic studies is still much smaller than the number of atoms in e. g. 1 mole of matter (approximately $6{\cdot}10^{23}$ atoms).

systematic manner (e. g. by conducting an experiment). Data collection can also be guided by the research subject e. g. in unstructured interviews. This is important because studying certain aspects of language is tricky: Many concepts cannot be readily accessed or measured and resist quantification. We cannot climb into people's heads and extract the one feature of language we are interested in, especially when it is an attitude to a particular way of speaking or a subconscious rule to produce acceptable linguistic output.[5] Qualitative approaches to the study of language therefore play an important role in linguistic research.

On the other hand, many branches of linguistics look at language from a quantitative perspective. Efforts to quantify linguistic concepts by introducing a rigorous mathematical framework have been made in different sub-disciplines, most notably in semantics and generative syntax. Quantitative linguistics tries to find the underlying rules for language use which can be formulated mathematically (e. g. the empirical Zipf's law describing the frequency of words). Dialectometry measures the linguistic distance between different language varieties and correlates it with other factors such as geographical distance. Quantitative studies of language variation were pioneered in the 1960s by William Labov who is generally regarded as the founder of sociolinguistics. Labov was one of the first to collect linguistic data on a large scale by making rapid and anonymous observations.[6] In doing so, he was able to correlate phonological variables (pronunciation) with social status [9]. The emergence of computers and their processing power has spawned new sub-disciplines like computational linguistics which model e. g. language learning. Even so, it is difficult to say whether it is possible to establish linguistic laws which have the same properties as physical laws (being stable, universal, expressible in an equation).

Unsurprisingly, both qualitative and quantitative research methods have advantages and disadvantages. Coming back to the topic of this thesis: A qualitative approach is able to characterize language shift in a small region precisely and take into account the inner subjective perspective of the speakers. At the same time, it means a sizable amount of fieldwork and is thus limited in time and space. A quantitative approach is able to recognize trends in the bigger picture and uncover more general rules underlying language shift—but it may mean that the development in some settlements is described less accurately.

In other words: Quantitative research gives us a close-up but narrow view, quantitative research gives us a wide but distant view. Both are necessary for a complete picture. Using physics to explore a linguistic research question is therefore intended not to replace existing linguistic methods, but rather to complement them.

---

5 Take the concept of "word": Each of us *knows* what a word is, yet there is no agreed-upon scientific all-encompassing definition [8].

6 Labov recognized that people adapt their language based on the social situation, making the prevalent tape-recorder interview a poor choice to gain a complete understanding of a person's speech.

1.2 Aims of this thesis

The goal of this thesis is to model linguistic processes, namely language shift in Austria(-Hungary) over time and space. The development of such a model based on methods from physics as well as new possibilities and pitfalls of this approach will be presented within this thesis. In applying physical models to linguistics, the project brings together two different perspectives and thus opens up new opportunities in physical research on spread phenomena and quantitative linguistic research.

It should be pointed out that the aim is not to study language evolution (as in: evolution and change of one specific language system e. g. German) or to contribute to dialectological research (because the thesis does not focus on one language and its dialects but on several different languages and their geographic distribution). It is, however, a contribution to the sociolinguistic (historical) study of languages in Austria and Hungary.

In contrast to existing sociolinguistic studies of language contact situations in Austria and Hungary, the thesis also does not directly investigate the social impact of language shift and the way social parameters influence variation of one language (intralanguage variation). Rather, it studies language shift as a large-scale phenomenon using methods from the natural sciences to get a complete overview of past (and possibly future) spatio-temporal development.

The primary scientific aim of the thesis is this:

- development of a microscopic[7] mathematical model for language use and its change over time and space as language shift
- in different language contact situations in Austria(-Hungary)
- based on methods from physics

The aim described above leads to two central questions:

- **How can language shift be described by mathematical-physical models?**
  By trying to answer this question, we test the applicability of mathematical approaches to linguistic research and the limits of quantifying social processes.
- **Why do people change their language, what are the factors influencing language shift?**
  Answers to this question not only help us understand the process of language shift better, but may also impact language documentation and preservation efforts (and more generally the promotion of multilingualism).

At some point, it is desirable for a model to be tested on data—or even to simply create hypotheses based on data for implementation in the model. In this case, this meant working with large-scale data on language use which was digitized and evaluated in detail for two language contact

7 "Microscopic" here means on the basis of the smallest population units: Towns, villages and hamlets are the starting point for modeling.

situations before being used for model testing. Model testing with empirical data is crucial, as it ensures that both the applicability and validity of the new model can be examined. However, empirical data on language as used in this thesis are quite different from data obtained from physics experiments, and historical data even more so:

– When it comes to asking a person what language they use, we are mostly reliant on their answer, not on a piece of equipment which gives an "objective" reading (within the measurement uncertainty).[8]
– We cannot go back in time to create more data, or different data as in a physics experiment which might be repeated under different conditions.

From this, it follows that linguistic data cannot and should not be treated the same as data on physical processes by simply applying the same types of model. The development of a mathematical model for language shift will therefore be explored on two layers:

- the physical layer, where a large amount of empirical data on language use is treated as "objective" and processed: How do we best model the data? How can we define linguistic concepts formally using mathematics? Is it possible to make predictions about language shift in the future, just like we can calculate probabilities for the movement of an atom? Which factors influence language shift and how much?
- the linguistic layer, where the context of the data is taken into account: What does it mean when someone states that they use one language or another? How do these statements change and what factors influence them? In other words, what can a mathematical model actually tell us about language use and language shift?

1.3 Structure of this thesis / Road map to the contents

This thesis is structured as follows:

The idea is to go from general aspects to specific examples, in that I will first give a background and motivation for the problem posed in this thesis from two sides, the linguistics and the physics side, in section 2. This section gives an overview over the linguistic background and terminology of language shift before going on to discuss analogies between physics and linguistics, between languages and atoms.

In section 3, the methods of this thesis are established: the concept of mathematical modeling as a tool is presented, and classes of mathematical/physical models applicable to language shift are examined. As we are primarily interested in building a model based on empirical data, the next section, section 4, considers available data sources and the manifold ways of asking people

8 Certain features of language can be "objectified" more easily than others. For example, differences in pronunciation (phonetic variation) can be recorded and the sound waves can be evaluated by their physical characteristics.

about their language as well as the consequences involved. This section also introduces the data used for modeling, the Austro-Hungarian (and Austrian and Hungarian) census.

Section 5 sets the scene and historical background for the "laboratory" used in this thesis, two language contact situations in Austria-Hungary and today's Austria and Hungary. We perform a comparative data evaluation in section 6 to explore the consequences of local circumstances for mathematical models of language shift. In this section, we compare the data with regard to two factors: (1) language policy & (2) settlement and population structure. We then go on to the actual modeling, first for Slovenian–German in Carinthia in section 7. After a brief historical background, we first consider a macroscopic (large-scale) approach to language shift—which is found to be inadequate—before we try a microscopic (small-scale) modeling approach which looks at individual cities. Data preparation is described in this section as well as the actual step-by-step model creation. Results of the modeling procedure are presented and discussed.

In the next part, section 8, the model developed for Carinthia is applied to another language shift situation, German–Hungarian in southern Hungary. Again a brief historical background is given, then the data preparation and modeling procedure are briefly reviewed before presenting results. Section 9 discusses the results obtained from modeling both situations, different language shift situations and what these differences mean for modeling. As a conclusion of this work, section 10 gives a complete summary of everything covered and connects the results back to the research questions posed.

2 Motivation and background

People are constantly changing their language: Depending on who we talk to when and about what topic, we use different languages, different varieties, different registers[9]. However, this language use itself can change. In extreme cases, we might even completely stop using a particular language—we shift our language. How does this happen? Why do people shift their language?

This central question can be examined from various points of view. From the linguistic point of view, language shift may be considered in connection with influencing factors such as speaker numbers, or as language shift in connection with the social networks of speakers. On the other hand, we might look at this problem as an opportunity for developing a new research method. We can try to see if we can find an answer by working with approaches which are successfully employed in other disciplines (in this case, physics). In this context, we are using a situation of language shift as our laboratory to test the new method.

Consequently, there are two motivations in studying changes in language use the way it is presented in this thesis: trying to answer the question "why do people shift their language?" itself (or rather: "how do statements about language use change over time and space?"), and seeing if methods from physics can be used in linguistic research and how.

2.1 Linguistics: Why do people shift their language?

Language is an important part of human communication and plays a central role in shaping human cognition. As a system, languages in the world are both heterogeneous and dynamic: They co-exist in a number of shapes and forms, and all of them are continually under transformation. An interesting question in such a situation of abundant variation is now not only why and when which language is used, but also the opposite: why a certain language is no longer used.

The study of changes in language use such as in language shift deals with two questions that are closely related:

(1) Description of the process: When and where does the language use change?

(2) Analysis: What led to this process?

---

9 "Variety" is a catch-all term for any language variation, e. g. geographic variation or variation by sex/gender. "Register" is a term for a functional variety of language, used in a certain context at a certain time. In other words, register is determined by the situational context (what the person is doing), not by a person's background (who the person is) [10]. Examples of different registers are legal English ("legalese"), or the language used in religious ceremonies, or any kind of discipline-specific technical language. Registers typically vary depending on the formality of the situation.

A wide array of terminology is available for the description of changing language use. The terms differ in which aspect of language as a dynamic system they emphasize and how "language" is conceptualized (see also section 4).

Language change Language change is the general term for change within a language, that is, changes in a language's vocabulary, syntax (grammar), morphology, etc.

Language shift Language shift on the other hand is defined as the (mostly permanent) change from one (single) language to another by a speaker [11]: "the change from the habitual use of one language to that of another" [12, p. 68]. This process happens progressively and individually, for one person, but over time also collectively, for a society which gradually changes its language use. Studies of language shift typically examine the shift from one language to another rather than from one variety of a language to another variety of the same language.[10] Language shift usually happens in a language contact situation, a situation where two or more languages (or varieties of a language) are used at the same time and place (*in contact* with each other) and thus interact with each other.[11]

In some cases, language shift can occur when a new language is introduced forcefully to an area where it had not been used before, i. e. language shift as a consequence of language contact when there previously was none, as in colonialism. More generally, language shift is not something that necessarily "just happens", it can also be accelerated when one language (as well as the people using that language) is systematically disadvantaged or not granted legal rights to using that language. Speakers then do not merely change their language use, they actively (have to) give up their previous language. Language shift can therefore have wide-reaching consequences. For many people, language is seen as an important part of their culture and identity (although the interplay between language and identity is complex, see [15]). Changes in language use might therefore affect their own identity as well as their linguistic abilities and knowledge.

Both language change and language shift look at changes mostly from a speaker's perspective (the speaker's use of a language is changed or the speaker actively changes their use). It is also

10 The distinction between "language" and "variety of a single language" is notoriously difficult [13], since there is ultimately no clear line to draw. However, in some cases the distinction is clear, especially when it comes to languages from different language families that are not mutually intelligible. This situation applies to the cases studied in this thesis: Slovenian and German in southern Carinthia coexist in the same geographical area and are influenced by language contact [14], but are distinguishable as two different languages and not varieties of a single language (although they are of course varieties of two different languages). The same holds true for the second case study of German and Hungarian in Hungary, which are again languages from different language families and cannot be mutually understood.

11 Language shift is not the only possible outcome in a language contact situation. Language contact can similarly lead to language change in various ways. See [**muysken**] for an overview of the different possible scenarios.

possible to look at changing language use from a perspective which does not focus on the speakers but rather the language(s) itself, and the sociological processes influencing language use [16].

Language dynamics Language dynamics is a general term for the way different languages or speakers of the same language interact and by which factors this interaction is driven.
Language spread Language spread looks at how the space where a language is used changes, i. e. at the interplay of language and space in the widest sense.[12] If a language spreads, then the area where it is used becomes larger. "Area" and "space" here do not necessarily refer to geographical space, they can also mean domains of language use such as home, work, etc. [19] or social space [20]. Spread of a language across space(s) in turn means that a larger and larger part of a speaker community starts to adopt a certain language.

No matter which terminology is used, the process of changing language use is not an absolute, abrupt change. It is a gradual change which might start with changing the language in one area of use, then in another—or only in one area of use.[13] This means that after a language shift, the original language might still be used for one communicative function e. g. for communication within the family. Even when a language A is completely replaced by a language B, speakers tend to go through intermediate stages (following the speaker classification by Haugen [23, p. 370]):

$$A > A/b > A/B > a/B > B \tag{1}$$

where A, B denotes the primary use and a,b the secondary use of that language.

As a final consequence, language shift/spread can lead to a language no longer being actively used at all, by no-one. Languages affected by this are often metaphorically called "endangered languages" which should be saved before they succumb to language death (without going into detail here on when a language can be considered extinct/dead, see [24]). Likewise, the "vitality" of languages can be quantified and classified on scales (e. g. [25; 26]) and language maintenance efforts can be made to prevent language loss and even revive languages [11]. There can also be language competition and much more (see [27, p. 9f] for an overview). In these metaphors, language is presented as a living organism [28; 29]—but it should not be forgotten that language is only possible through speakers and it is these speakers who do something, not the language "by itself".

12 See also [17; 18] for a more general overview of language and space.

13 This follows a view of multilingualism and language learning where speakers do not alternate between two separate, "complete", enclosed languages, but rather mix and match parts from both languages for different functions. The different languages of a multilingual person then do not form separate systems which are combined without influencing each other, but instead they form one integrated system (individual multilingualism as "multi-competence", see [21; 22]).

Many—though not all—endangered languages are languages which have a small number of speakers or can be disadvantaged in the places where they are used. They are then regarded as "minority languages". Language shift is therefore associated with the discourse on minorities in general. "Minority" as a definition has many aspects (see e. g. [30, p. 32–33] and [31]) and is characterized by one group being unequal to another group. The minority group differs from the majority group in one or more features, e. g. in the number of speakers or the rights granted to the speakers. The latter also indicates inequality as a characteristic of the minority: minority languages are frequently lower in the societal hierarchy than the majority language and have less "power", less prestige and/or fewer speakers.

Minority languages are often associated with a minority group—these are mostly cultural or ethnic minority groups ("Volksgruppen" in German, see [30, p. 26–29]). Such minority groups can be found in many countries and are often protected by law. Legal protection is sometimes dependent on numerical thresholds (a certain number of minority group members must exist for rights to be granted) which are drawn from the results of language surveys (sections 4.1 and 4.3).

So, what drives language shift? Multiple possible explanations have been identified in the literature e. g. [32, p. 60–62]; [25–27; 33; 34]: differences in languages' prestige, lack of interaction with other speakers of the same language, economic reasons, language policy, demographic factors (village size, rural depopulation, ...) as well as the ecological context (ecolinguistics)[14] and many more. A more detailed overview of factors influencing language shift is given in table 1.

Additionally, different mechanisms which influence language shift or more generally language change[15] have been proposed [40–44]: leveling, wave theory and the gravity model. Leveling refers to "the reduction or attrition of *marked* variants" [40, p. 98] (emphasis in original), i. e. the reduction of unusual minority forms. Wave theory assumes that the spread of linguistic features happens in waves from one culturally dominant center of innovation (first demonstrated by Schmidt in 1872 [45]). According to Bloomfield [46], this spread primarily happens where the communication density is strongest because it happens through contact between individuals. The gravity model (also termed cascade [47] or hierarchical model) is an extension of Bloomfield's observation. In the simplest form, it proposes that the interaction M_{ij} between two settlements i,j depends on their distance d_{ij} and population sizes P_i, P_j [34]:

$$M_{ij} = \frac{P_i \cdot P_j}{(d_{ij})^2} \tag{2}$$

14 Ecolinguistics/ecological linguistics studies how language use is shaped not only by social context (speaker community, speaker numbers, etc.) but also the ecological circumstances such as the physical environment [35]. See e. g. [36–39].

15 We can view language shift as a case of language change, where it is a "comprehensive" behavior (use of a *language* that was not used previously) that spreads rather than a change in an existing system (use of a *linguistic feature within one language* that was not used previously).

geographical area in which a language is used
isolation of the community
number/percentage of speakers in the population
marriage patterns
settlement structure: rural or urban
community structure (living closely together or dispersed)
possibility of employment in the place of residence
migration
correlation of social status with language use
group identity and its ties to that language
availability of educational, cultural and religious opportunities in a language
availability of material for language teaching and literature
quality of documentation of a language
level of standardization of a language
writing system, compatibility with new technology e. g. the internet
institutional support
official status, legal policies
prestige of a language (within its own community and other communities)
interaction with speakers of the same language
interaction with speakers of other languages, language contact
mass media

Table 1: A non-exhaustive overview of factors influencing language use and possibly leading to language shift (based on [32, p. 60–62]; [25–27; 33; 34])

The equation for interaction is inspired by Newton's theory of gravitation, where the force of attraction F between two masses m_i, m_j depends on their product and the squared distance $(d_{ij})^2$ between them, multiplied by the gravitational constant G:

$$F = G \cdot \frac{m_i \cdot m_j}{(d_{ij})^2} \tag{3}$$

The interaction M_{ij} resulting from equation (2) is an index score which can be used to compare between different settlements of different sizes rather than an absolute judgment. Communication decreases with higher (physical) distance and increases with higher population size.[16] Increased contact between individuals leads to language change, so language change should start in bigger cities (i. e. from several centers rather than only one) and gradually spread to smaller towns. Accordingly, each town has a sphere of influence for propagating language change. The model is then extended [34] to include linguistic factors such as linguistic similarity and to measure the influence of one center on another (rather than the interaction between them).

Starting from scratch and creating hypotheses on what influences language use and language shift is therefore not the hard part in building a mathematical model for language shift, because this knowledge is well established from previous studies. The challenge is rather to identify which factors play a role in a given language contact situation, and to incorporate these into the model by quantifying their influence.

2.2 Physics: Can we use physics in linguistic research?

Physics at its core is the study of matter in all its forms. To do so, there is a wealth of approaches, theoretical and experimental (and nowadays even computational). Many of these approaches have been adopted by other disciplines, leading to the emergence of interdisciplinary research fields: biophysics, chemical physics, geophysics, astrophysics and so on. Most of these fields still study matter in some way, but the use of physical methods is not restricted to the study of material things. Apart from the connection of physics with philosophy in some fields (quantum physics in particular), there is the branch of social physics (sociophysics) which uses tools inspired by physics to study human behavior e. g. voting in elections or marital interaction (divorce and marriage repair) [48]. Social physics in turn has close ties to computer science and information theory. Social physics can be seen as part of a larger research effort concerned with understanding all kinds of social phenomena, one of which is language.

16 While this was certainly true when the gravity model was introduced by Trudgill in 1974, the rise of the internet, video chat, text messaging, etc. allows speakers to easily keep in contact even if they are geographically separated. "Distance" thus cannot be taken to mean only physical distance anymore.

Physicists looking at language [49] is not a new phenomenon, and neither is the other way around: linguists looking at physics. Many linguistic theories draw motivation from physical (or chemical, depending on how we look at it) concepts, some of which are:

- Coherence and cohesion in text linguistics. Linguistic coherence refers to what makes a text meaningful e. g. logical/thematic structure. For a text to be coherent, the sentences must have logical organization and semantic continuity. Similarly, physical coherence refers to wave sources which have the same frequency/wavelength and a constant phase difference, making it possible for them to interfere.

  Linguistic cohesion on the other hand refers to what holds a text together on its surface, i. e. linking words between sentences whereas cohesion in chemistry and physics refers to the attractive forces between particles of a substance which makes them stick together.
- Verb valence. Valence is the number of arguments needed for a verb to form a grammatical sentence, i. e. the number of arguments which can be "bound" to a verb. For example, "to love" needs two arguments (*I love words*) and therefore has valence two. This concept is inspired by valence in physics/chemistry, i. e. the number of hydrogen atoms which can be bound to an atom.
- The theory of language dynamics, which is explicitly motivated by classical mechanics and dynamics in physics. Just like dynamics is the science of forces and influences responsible for motion of physical objects, language dynamics is intended to be the science of influences on ever-changing language which is constantly "in motion" [50, p. 20].
- Glottochronology which seeks to uncover genetic relations between languages by assuming that the base vocabulary in a language "decays" (is replaced by other words) over time, as do atoms in radioactive decay.

Another analogy that comes to mind concerning language shift is one central to this thesis: that of diffusion. Diffusion in physics means the movement of particles as well as the transport of matter through this movement. This movement is studied over time and space. It can be measured experimentally and simulated using mathematical models implemented on the computer (see e. g. [51] for an overview and section 2.3 below where modeling of physical diffusion is discussed more in detail and in relation to linguistics).

"Diffusion" in the broader sense as "the spread of something" can be found in many fields, sometimes under the term "interdisciplinary diffusion". Models of physical diffusion have been successfully applied in biology and medicine, for example to evaluate the spread of diseases, genetic markers or invasive species (and assess the effectiveness of management strategies to control them), or to follow the path of animals searching for food [48; 52–58]. Not all diffusion theories concern material things: Models on the diffusion of innovations aim to explain the spread of new ideas or technological innovations [59]. Similarly, in conjunction with archeology and history, the spread of culture can also be investigated as diffusion [60; 61]; this can include the spread of beliefs like religion or of skills like farming. Models of cultural diffusion[17] can also be used to study where a certain part of culture originated e. g. by modeling genetic relationships between languages (language trees) to locate the Indo-European "homeland" where the Indo-European language family presumably originated [63; 64].

Likewise, language shift can be seen as a spreading process—a new behavior (use of a certain language) spreads as "language diffusion". This linguistic diffusion can thus be modeled similarly to physical diffusion.[18]

2.3 The analogy between diffusion of matter and of languages: Are languages like atoms?

[*This section is based on [2].*]

Now that we have found an analogy between physical diffusion and "language diffusion" which might be suitable to use for the mathematical modeling of language shift, let us examine this analogy in more detail with regard to two aspects: the characteristics of the elements involved, and their dynamics (movement).

(1) Any physical system has a state it is currently in, depending on the arrangement of its particles. Likewise, a particle itself has a state which determines its characteristics. We can posit the same with regard to people and languages: a person is in the state of using one, two or multiple languages (for all or only specific communicative purposes), and any combination of languages is possible. In both physics and linguistics, these states can change through external or internal forces. Speakers of one language can adopt another. Let us take a look at what happens in a situation where two languages are unequal, i. e. we have a majority and a minority language. Speakers of a majority language usually learn the minority language in addition and they become bilingual (for some communicative functions and possibly for a limited time). On the other hand, speakers of the minority language are most likely already bilingual because they function in an environment dominated by the majority language. This means that the majority language can only spread to areas where it is not already used: to communicative domains still reserved for the minority language. If this has already taken place ("complete bilingualism", i. e. in all domains and areas), the majority language cannot

17 Here, I mean only mathematical models for cultural diffusion inspired by physics. In cultural anthropology, there are different models for cultural diffusion as well as a typology [62, p. 27ff], but these do not have a strict mathematical formulation behind them.

18 Linguistic diffusion usually refers to the change and spread of individual features in a language ("innovations" like the new pronunciation of a word), through contact between varieties of a language or different languages, i. e. it refers to language change rather than language shift. Therefore, there are also models directly inspired by physical diffusion which look at the spread of dialects e. g. [65; 66]. Linguistic diffusion can also look at the diffusion (spread) of individual words e. g. [67].

spread—but the minority language can decrease in use (assuming that the majority language by definition has more power). In analogy with physical diffusion: If the majority language completely replaces the minority language, particles vanish from the system. This means that the number of speakers of a language is not a conserved quantity. However, this applies only to the languages and their manifestation as "speaker number equivalents", but not the speakers themselves: No one disappears if they stop using a language, although their linguistic behavior does.

What about the *number* of states a system (the particles or the speakers) can be in? We have already noted that speakers can change their state by learning additional languages, or dropping ones they no longer use. In case of language shift, they can be e. g. Slovenian speakers and then become German speakers.[19] Similarly, atoms can change their "identity" (chemical element) through radioactive decay. For atoms this is a one-way process whereas speakers are able to change back and forth between languages.[20] However, in contrast to speakers atoms can only be one thing, one element at a time.[21] A lead atom is only a lead atom, but never a lead atom and a gold atom at the same time. People, on the hand, usually have no problem being bilingual or multilingual.[22]

(2) In physical diffusion, it is the particles which are in motion. In language diffusion, a behavior (use of a language) spreads among a community, but it is not the languages themselves which are moving. Even though languages may sometimes be described metaphorically as living organisms which can become "endangered" or "dying" [28; 29] (especially in discussion of minority languages), they are not inherently and independently alive and thus cannot move. This means we cannot directly equate particles with speakers or languages, but should rather

---

19 We will not go into detail here about what it means to be a speaker of a language. At this point, self-identification of someone as a speaker of language X suffices.

20 Another restriction placed on atoms is that not all chemical elements can be transformed into each other: some elements cannot be turned into a certain other element no matter how much energy is put into the system. There is no such general restriction on speakers: they are theoretically able to acquire any new language (even though it might be more difficult if one starts later in life or if there are other factors such as being hard of hearing), although one might not achieve the same proficiency level as someone who learned the language from birth or had no such factors involved.

21 For every chemical element, there can be multiple atomic representations. These are called isotopes and have different inner structures (numbers of neutrons), but are grouped together as the same element. Applied to linguistics, isotopes correspond to speakers with different proficiency levels of the same language. Isotopes can develop into different elements through radioactive decay. This is not the case with language: If someone forgets their knowledge of English, they do not turn into 日本語話者 (a Japanese speaker).

22 This tends to complicate mathematical descriptions, which is why many mathematical models consider only "binary speakers", i. e. those who use language A or B and directly go from one to the other without the intermediate stages in equation (1). More complex models include an intermediate bilingual stages.

focus on similarities in the spreading process. In other words, we do not have an object-based analogy but rather a process-based analogy [68].

The movement of particles in physics is determined by the laws of thermodynamics. Following these laws, a system of particles aims to arrange itself in such a way that its free energy is minimized. Free energy determines the amount of work available in the system. Its value for various states of the system is described by the so-called free energy function. The minimum free energy can be accomplished in several ways, depending on the components (chemical composition) and the form of the free energy function of the system. An example: If all components have "full miscibility", i. e. they can be mixed without restriction to form a homogeneous solution (as is the case e. g. with water and ethanol), the free energy is minimized when an equal concentration of particles is achieved everywhere. Accordingly, movement (diffusion) goes from areas with many particles (high concentration) to areas with few particles (low concentration). However, generally a minimum of free energy coincides with a phase separation (particles separated by their type on different sides) rather than an equal concentration of particles, so diffusion runs into the other direction than in our example. In order to change the direction of diffusion determined by thermodynamics, additional energy must be added to the system.

(3) Diffusion in physics is the transport of matter through the movement of atoms, which, as the subject area "physics" implies, is physical. In linguistics, the transport of behaviors or features is not physical but virtual (even if these behaviors of features may have physical manifestations such as sound waves, or if they are transported physically in the "vessel" of speakers who move). We can measure and make visible the distribution of atoms in matter and the movement of atoms with various experimental methods [51; 69]. We can do the same with linguistic behavior, but by their nature virtual elements are harder to pin down. What we see as a physical manifestation changes depending on how we look at it and which medium is used for expression (speech, writing or signs in the case of sign languages), so we obtain a slice of reality—but there are many realities.[23] This is a key point in collecting data on language use and linguistic behaviors for the study of language spread: Language is first and foremost located in people's heads, so we rely on the same people telling us what they do and how they behave. We also rely on how they interpret our questions and what a person accepts as "a language" (a subject of debate even among linguists).

*

23 The same predicament arises in physics: the famous example of wave-particle duality in quantum mechanics, whereby energy and most physical entities such as atoms have characteristics of both waves and particles. Depending on the interaction with the detector in an experimental set-up, we observe either the wave or particle character.

Evidently, there are many ways to use physics in linguistic research. However, as Altmann and Meyer [49] caution, the abundance of possible analogies between physics and linguistics do not mean that physical argumentation can (and should) be used to explain linguistic phenomena and draw conclusions within the linguistic system. In their words [49, p. 49]: "(i) Use mathematical methods current in physics but do not use physical argumentation. (ii) Search for analogy as much as possible, and if found, do not reduce one discipline to another but try to find a common mathematical structure and a reason for this structural commonality!" Ultimately, the question thus becomes not "*Can* we use physics in linguistic research?", but "*How* do we use physics in linguistic research?" The next section will attempt to answer this question, at least for the question posed at the beginning of this thesis: Why do people shift their language?

3 Methods: Modeling language shift / How do we use physics in (this) linguistic research?

When physics is used in other disciplines, this can mean that the same experimental techniques are used, that theoretical concepts are transferred as analogies or that mathematical models based on these concepts are adapted for application in that discipline. In the case of linguistics (as well as all the fields cited in the preceding section which study interdisciplinary diffusion), the latter two are usually the case. They are also the tools employed in this thesis: applying the concept of diffusion from physics to linguistics to the study of language shift (or the spread of languages), and using mathematical models to describe the process on a large scale and gain insight into the reasons for language shift.

3.1 Mathematical models as a tool

Mathematical models and computer simulation are tools to identify the underlying rules for a process. They can also be used as substitutes for an experiment, when experiments are not possible (see also [70, p. 10]). This is important for research on changing language use (historical sociolinguistics) insofar as that language use does not change overnight. Investigating language shift as a process necessitates several "measurements" over time. However, this means we also cannot go back in time to take more or different measurements, i. e. we usually cannot repeat the study under different conditions as we could with a physics experiment in a lab.[24] In this case, mathematical models are helpful.

Modeling is the abstraction from the "real world", the creation of abstract rules to describe processes, usually expressed by mathematical formulas. Simulation is the application of these abstract rules to empirical data e. g. by implementing the model as a computer program and running it. Put another way, modeling is the imitation of one process through another [70], on an abstract "meta-layer".

Figure 2 shows the general principle of modeling and simulation employed in this thesis. Let us assume we want to study the process of language shift, i. e. how language use changes over time (and space). We expect that people do not change their language randomly, but that there are factors and rules governing the process. We then create hypotheses on what these rules could look like based on existing observations and studies. For example, we might speculate that the number of people who will use a language in the next year depends on the number of speakers

24 Not all physics experiments can be repeated.

we currently have plus interaction with other speakers around them. This rule is then applied to each data point which gives us some result. We can compare this result to empirical data and change the rules based on how well the result from the simulation fits the data. This process of improving the model can be repeated until a satisfactory goodness of fit between result and data is achieved.

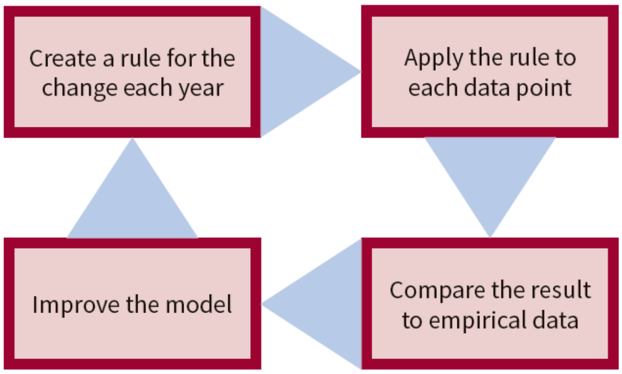


Figure 2: General principle of modeling and simulation employed in this thesis

If the model results and the empirical data fit together, then the model is a possible explanation for the underlying process. However, it is important to keep in mind that even with the best mathematical model, we cannot "prove" anything. We can recreate the course of history the way we assume it happened, and hope that the outcome is the same as in reality. This means that some degree of background knowledge is needed to create useful models. It also means we need to be clear on what part of reality the model imitates to interpret the results, particularly when concepts are transferred from one discipline to another in interdisciplinary research. In the following section, we will consider how different models of physical diffusion can be applied to linguistic research to model different aspects of language shift (language spread and speaker behavior—immaterial and material aspects of change, as it were).

3.2 Modeling language shift as a physicist

[*This section is based on [2].*]

We have seen that the analogy between diffusion in physics (the spread of particles) and diffusion in linguistics (the spread of languages) is not perfect. This means that we need to be especially clear on what is being modeled when we try to apply methods and models from physics to linguistic data.

3.2.1 Models of physical diffusion

Physicists generally study diffusion in two ways: from a macroscopic point of view (looking at the concentration of something, i. e. the amount of substance per unit area or unit volume, and its change), and from a microscopic point of view (looking at how individual particles behave and move).

(1) Macroscopically, the change in the concentration c is described by the so-called diffusion equation (Fick's second law). The concentration of a substance can vary over space and time, i. e. in one dimension it is a function $c = c(x,t)$; in two or three dimensions it is a function $c = c(\mathbf{r},t)$ where $\mathbf{r}$ is a spatial vector. For the simplest case (one dimension, homogeneous medium, no external influences, etc.) we obtain the following diffusion equation[25]:

$$\frac{\partial c}{\partial t} = D \cdot \frac{\partial^2 c}{\partial x^2} \tag{4}$$

In this equation, $\partial c/\partial t$ gives the temporal change in the concentration c of a substance, i. e. the development of the concentration over time. From the other side of the equation, we see that the development over time is proportional to the expression $\partial^2 c/\partial x^2$, the second derivative or curvature of c with respect to x. The curvature represents the change of the so-called concentration gradient over space, i. e. the change of local differences in concentration. The diffusion coefficient or diffusivity D is a measure for both the proportionality (as a multiplicative constant) and the speed of diffusion. Its unit is area covered per time, in SI units m^2/s.

(2) Microscopically, we observe and model the behavior of individual particles on an atomic level. As diffusion is the transport of matter through the movement of atoms, this movement is governed by the boundary conditions imposed by the state of matter. In gases or liquids, particles can move relatively freely and fly or shuffle around. In solids, atoms or molecules jump from one place to another. These places are called sites and are defined by the structure of the solid, e. g. positions on a lattice for solids with crystal structure. Atomic movement usually is random and chaotic. It is generally described by the mean square displacement, which measures how far a particle has moved after a certain time on average. Because atomic diffusion is a stochastic process, i. e. there is no predetermined outcome, only an average travel or jump distance can be given, hence the name "mean square displacement". The mean square displacement $\langle r^2 \rangle$ can be related to the macroscopic diffusion coefficient D by the

25 In two or three dimensions with the Cartesian coordinates (x,y) or (x,y,z) the diffusion equation takes the form

$$\frac{\partial c}{\partial t} = D \cdot \Delta c \tag{*}$$

where Δ is the Laplace operator $\partial^2/\partial x^2 + \partial^2/\partial y^2$ (two dimensions) or $\partial^2/\partial x^2 + \partial^2/\partial y^2 + \partial^2/\partial z^2$ (three dimensions).

Einstein-Smoluchowski relation:

$$\langle r^2 \rangle = 2n \cdot D \cdot t \tag{5}$$

where n is the number of dimensions of the system and t is the elapsed time.

3.2.2 Models of language spread/shift

We have seen that there are two main ways of examining physical diffusion, and we find them both represented in mathematical models of language spread. Such models can be broadly divided into two classes:[26] (1) macroscopic reaction-diffusion equations which follow the fraction of speakers in the population, (2) microscopic agent-based models which follow the behavior of individual speakers.

Reaction-diffusion equations

Examples: [76–80][27]

Reaction-diffusion models expand equation (4) by adding a term to account for external influence, the so-called "reaction term". The result is a differential equation or a system of equations of the following type:

$$\frac{\partial c}{\partial t} = D \cdot \frac{\partial^2 c}{\partial x^2} + f(c) \tag{6}$$

where c is the concentration of speakers, i. e. the fraction in the population, and $f(c)$ is any function of c, the reaction term. Similar to reaction-diffusion equations in chemistry and physics, this equation describes the change $\partial c/\partial t$ in the concentration (here the fraction) of speakers over time. The evolution of the fraction depends on $\partial^2 c/\partial x^2$, the change in the concentration gradient over space, and the diffusion coefficient D. The reaction term $f(c)$ generally describes the transformation of the substances involved in the reaction into one another, i. e. as the concentration of one substance grows, the concentration of another substance becomes smaller. Depending on its sign, it can lead to an increase (positive $f(c)$) or a decrease (negative $f(c)$)

---

26 There are of course multiple other mathematical models of language spread which do not neatly fit into the two categories presented here. Models based on statistical mechanics, network theory, complex systems theory and/or game theory are rather popular for the general study of social dynamics and thus also for language use, see e. g. [71] for a review, or [72–74] for examples of such models for language shift. See also [75] for a general review of mathematical models for language shift. Here, we will restrict the overview to models explicitly inspired by (physical) diffusion models.

27 A similar class of models (often based on the so-called Lotka–Volterra equations) looks at the change in the concentration of speakers without a spatial (space-dependent) component $\partial^2 c/\partial x^2$, e. g. [81–85]. These are not mentioned in detail here as they do not have an exact analogy in physical diffusion.

of the concentration/fraction. For linguistic applications, it describes the interplay between speakers of different languages (for a set of equations), or the transformation of speakers of one language into speakers of another language so that the fraction—and number—of speakers of the second language increases (which leads to a decrease of the speakers of the first language by the same amount, neglecting changes in the population). Reaction-diffusion equations can also be extended further by including birth and death rates.

Let us look at an actual example of a reaction-diffusion model for language spread. In the simplest form (see [79; 86]), we have two monolingual groups c_1, c_2 who use different languages and interact with each other. The evolution of the two groups over time is given by the following equation system:

$$\frac{\partial c_1}{\partial t} = D_1 \cdot \frac{\partial^2 c_1}{\partial x^2} + c_1[\alpha_1 - \beta_1 c_1] + \gamma c_1 c_2 \tag{7a}$$

$$\frac{\partial c_2}{\partial t} = D_2 \cdot \frac{\partial^2 c_2}{\partial x^2} + c_2[\alpha_2 - \beta_2 c_2] - \gamma c_1 c_2 \tag{7b}$$

where $c_1[\alpha_1 - \beta_1 c_1] + \gamma c_1 c_2$ and $c_2[\alpha_2 - \beta_2 c_2] - \gamma c_1 c_2$ represent the growth of the two language (speaker) groups. This growth term consists of two parts: $c_i[\alpha_i - \beta_i c_i]$ is the inherent population growth in a speaker group c_i and $\gamma c_1 c_2$ is the interaction between the two groups. The population growth term follows the logistic growth model originally proposed by François Verhulst [87, p. 35–39]; [48, p. 3–6]. The first part with coefficient α_i represents the general growth rate (birth rate). The second part with coefficient β_i represents the self-limitation of growth (as resources—in this case the number of people who can speak a language—are finite).

In the interaction term, the coefficient γ modifies the strength of the growth happening by exhange between groups (rather than by general population growth). It describes the rate of shift from one language to the other—if a language is "attractive" to speakers (and hence γ is positive), then people from group c_2 will shift their language towards the other language c_1. The term is included in both equations with differing signs (plus or minus) because it is assumed that people who stop using one language will start using the other—no-one is "lost" to a language-less state or a language other than the two being modeled.

Reaction-diffusion equations and diffusion equations look at the evolution of the concentration of something. Applied to language spread, this corresponds to the fraction of speakers of a language in the populations (or even the absolute number of speakers), even if it is the languages we actually want to model. Languages cannot exist without speakers, so whenever we model changes in the concentration (fraction) of a language, we always model the fraction of speakers by proxy.[28] This applies likewise to absolute numbers: We cannot have a result of "537

28 Bilingual speakers, if there are any, are then a separate language group and fraction, i. e. the population splits threewise into two monolingual groups and one bilingual groups.

Hungarian", although a result of 537 Hungarian *speakers* is perfectly acceptable. From this, we see that languages and speakers obviously cannot be directly equated, but they are interwoven and cannot be regarded as completely distinct entities.

Agent-based models and cellular automata

Examples: [88–93]

Agent-based models are the equivalent of a microscopic view on diffusion, in that they model the behavior of individual speakers, the so-called agents. In these models, the individual interactions between agents (speakers of a language) are simulated and at each interaction, the agents change their language with a probability p. This is similar to the jump rate of atoms, who jump only with a certain probability from one place to the next in a solid. Also similar to atoms in a crystal, speakers in agent-based models can be placed on a lattice in fixed places. Changes in the use of a language can then be realized in two ways. Either speakers are immobile and shift only their language, then the result is a model of actual "language movement", i. e. changes in linguistic behavior. Or the speakers do move, then we again have no clear separation of speaker movement and language movement. In any case, at some point any model of "language" needs to incorporate linguistic behavior into a physical person, the speaker. Such a transformation can be done in a number of ways, the one employed in this thesis (see section 7.4) is to operate with the probability of using a language to explicitly model linguistic behavior rather than speaker numbers outright. The probability is then multiplied by the population size and we obtain language as we encounter it in the real world and our data: the number of speakers of a language.

Cellular automata are related to agent-based models in that they provide a way to model the agents' behavior discretely in space and/or time [94; 95], e. g. as atoms placed on a lattice which jump from one lattice site to the next. Applied to language spread, this means our speakers are arranged on a lattice (grid) divided into cells of equal sizes. Each of these cells is updated per time-step according to some mathematical rule which describes the possible change in linguistic behavior. Cells can be filled either with speakers (one speaker per cell) or with larger units which can be inspired by empirical data, e. g. villages with multiple speakers. In the latter case, cells stand for geographic positions and create an abstract representation of real space (unlike in agent-based models, where this is not necessarily so). As the grid cells are flexible in what they represent (individual speakers or something else), cellular automata can be used to model both speakers and languages, and changes in their distribution.

In mathematical terms, cellular automata are generally deterministic (meaning the outcome is pre-determined by a rule and not influenced by probability). They can only have a finite number of states (usually binary, 0 or 1). Both properties hold true for the model based on cellular automata which is introduced in a later section (see section 7.4): The state, i. e. the number of speakers of a language represented by that cell, is given by a mathematical rule which does not have a stochastic element, so it can be calculated straightforwardly if the initial conditions are known. The result for a cell only changes when the initial conditions change, but not due to random effects (meaning there is only one possible outcome per initial condition). The number of states is not binary but finite because it has an upper limit given by a cell's total population—it is impossible to have more speakers of a language than there are people in the population per cell, and it is impossible to have non-integer results for speakers.

*

While reaction–diffusion equations and cellular automata have been presented in separate sections here, they can be seen as two ways of approaching the same problem which converge—meaning one can be transformed into the other (ref. [96] gives examples of transformations in both directions). Differential equations such as reaction-diffusion equations use derivatives of functions. Derivatives are the limits of so-called difference quotients: The difference quotient is defined as the difference between a function value $f(x + \Delta)$ and $f(x)$, divided by Δ. When Δ becomes infinitesimal (it goes to zero: $\Delta \to 0$), the value of the difference quotient is called the derivative of a function. In other words, the derivative gives the rate of change or the slope of a function for infinitely small steps, i. e. the rate of change in one point. This means the discrete rate of change (given by a rule for one timestep) in the cellular automaton is transitioned into a continuous equivalent when written as a differential equation. (Conversely, integrals—the "reverse" of derivatives—are the continuous versions of discrete sums.)

Although reaction–diffusion equations and cellular automata can both (re-)produce complex behavior and represent two sides of the same coin, turning the coin, i. e. the actual transformation, is sometimes not straightforward [96]. It is easier to go from continuum dynamics (reaction–diffusion equations) to discrete dynamics (cellular automata) than the other way around, as no general method is available for obtaining a differential equation from cellular automaton rules. For this reason, a thorough mathematical formulation as a differential equation of the cellular automaton presented in section 7.4 will not be attempted in this thesis.

3.3 The atoms of language

[*This section is based on [2].*]

Up to now, I have not specifically stated what is considered the atoms on the linguistic side of our analogy between the movement/spread of languages and the movement/spread of atoms.[29] I have generally equated "speakers" with atoms, but also "languages" as the bigger entities which

29 I have also not been very specific about the nature of atoms vs. particles in diffusion (although efforts have been made to use the word "atoms" where statements apply only to atoms), simply because it is not a key point for this thesis. Both atoms and particles can move, both transport matter which is the definition of diffusion, and both change the concentration of a substance when they move.

then settle down in the "vessel" of speakers. Concerning language, we can take another point of view: just like matter is made of particles, language is made of smaller elements. Then, words or linguistic features of any kind can be seen as the atoms which spread. This spread can be treated by the procedures outlined above in the same way. Just as we can model the proportion in the population using a certain language with reaction-diffusion equations macroscopically, we can model the proportion in the population using a certain linguistic feature, and the proportion which is not using the same feature (yet, but might adopt it as language change). For a more detailed microscopic view, we can also model the behavior of individual speakers who adopt a linguistic feature with a certain probability (or not), just as they might adopt another language as language shift occurs.

3.4 Incorporating social and psychological influence factors

[*This section is based on [2].*]

Diffusion models of any kind use the tool of mathematics to describe the development of something over time and space, be it the concentration of a substance or atomic movement. It follows that every mathematical variable therefore should correspond to a piece of the thing or process being modeled. Physical models are rather transparent in this regard. They operate with e. g. particle concentration and particle displacement, both of which are accessible by experimental techniques or can be derived from experimental results. So even if a measurement is expensive and labor-intensive, there generally is a physical equivalent to our mathematical term. Linguistics is not as kind and straightforward. Many factors influencing language spread/shift cannot be put directly as variables into our models as they do not have a readily available equivalent in terms of quantity. A prominent example of this are social and psychological factors, e. g. the prestige/status of a language (level of respect speakers have towards a language), potential economic advantages or upward social mobility tied to a certain language.

All of these attitudes towards and feelings about language influence the attractiveness of a language, i. e. its currency for attracting new speakers, and thus also language spread, but they cannot be easily measured or rated on a numeric scale. Attractiveness is one reason majority languages have an advantage over minority languages. Majority languages tend to have a higher attractiveness as speakers expect social or economic benefits in adopting the majority language. If learning the majority language increases the possibility of finding a job locally, then this incentive attracts new speakers and the language spreads. Going into the other direction (majority language speakers shifting to the minority language) requires more effort in the form of energy and/or money to increase the attractiveness of the minority language. As in physics, the barrier resulting from the languages' different attractiveness levels can only be overcome by putting additional energy into the system. Without this additional energy, the process goes into one direction only, as in thermodynamics.

Even though social and psychological factors are difficult to objectively measure and put into numbers (“feelings about language are inaccessible” [9, p. 9], as sociolinguist William Labov put it), various efforts have been made to incorporate them in models of language spread to represent the process more realistically. For example, prestige of a language has been included as a parameter e. g. in [81] with a value of this parameter s from 0 to 1, where 1 is the highest prestige possible (without specifying what this means in terms of linguistic or economic “currency”). In a model with two languages only, it is assumed that the second language then has the prestige $1 - s$. A numerical value for prestige is obtained by fitting the model to empirical data. While the result does not correspond to anything concrete in the “real”, physical world, it can be used to compare different language shift scenarios with a spectrum of prestige differences.

Social and psychological factors influencing language shift can also be subsumed under one parameter. For example, in [77] and equation (7) prestige is assumed to be included in the shift coefficient γ which describes the shift from one language to another, for whatever reason. Again, the shift coefficient does not have an obvious equivalent in the real word, but it can be used to compare different scenarios or find “stable solutions”, i. e. coexistence between languages without further shift. It can also indicate the general attractiveness of a language.

As will be seen later, the model developed in this thesis (see [1] and section 7.4) tries to bypass the issue of introducing an abstract parameter to account for social and psychological factors. It does so by using data on things which correspond to a language’s attractiveness and are present in measurable amounts in the “real” world. Our model includes data on bilingual schools where the minority language is also taught, and on parishes, where the minority language is used during religious ceremonies (e. g. mass). We also try to operationalize the interaction between speakers of the same language, as this interaction is important for a language’s attractiveness. If no one is around to communicate with in the same language, it becomes less attractive to adopt or use that language.

Since no person is an island, we take into account not only the interaction with people in the same village, but also that with people in surrounding villages. Each surrounding village contributes interaction proportional to its number of speakers of a language and indirectly proportional with its distance[30]—the more people in a neighboring village, the higher the interaction and the further away the neighboring village, the smaller the interaction. Such a formulation explicitly incorporates an immaterial social factor in the model, in such a way that its quantity follows from its physical expression. In other words, its numerical value can be calculated from data we already have or can easily obtain.

30 For the mathematical details, see section 7.4. Distance in this simplification only measures the direct path between two points without taking into account the road-rail network, any mountains in-between, etc. or the possibility of communication through electronic media.

Obviously, not all things which influence the general attractiveness of a language can be quantified in the same way as we did with interaction. Building a model is therefore always a trade-off between absolute correspondence to the real world and an improved fit to the data. Abstract parameters e. g. for prestige or attractiveness may lead to a better description of the data, but too many abstract parameters result in an intransparent model which becomes hard to falsify.

4 Data material

Mathematical models are usually at some point tested on empirical data, to see if the model is able to describe reality and to falsify implemented hypotheses. To test a model of language use (and the resulting language shift), we therefore need data on language use. How do we obtain data on what language people use? The easiest way is to simply ask them—flippant as it sounds, this is what language surveys (and researchers) have done and still do. "Language survey" here is a general catch-all term for any questionnaire asking a language-related question and registering the answer. There is not one definite end-all, be-all language survey because there are many ways to ask about language as outlined in section 4.1. Answers to language-related questions strongly depend on how the question is posed, what answers are possible and who is asking.

Language surveys offer information about languages by counting them in one way or another. They can give a numerical result on how many people use the language (the number of speakers)[31] and such results serve as the basis for modeling language shift in this thesis. More specifically, for a model of language shift (the change from the use of one language to another) we need data on

- which language is used
- when (at which time point),
- where (at which point in space) and
- by how many people (the number of speakers).

We also need *enough* data for modeling, i. e. quantitative data with a high temporal and spatial resolution.

Studying language shift frequently means historical research because many processes of language shift have started (or happened) long in the past. This means that in addition to the general limitations on language surveys, we are confronted with the so-called "bad data problem" [98, p. 10f]:

> [W]e do not have access to the same amount and quality of data as we would have if we would do a case study today, for the simple reason that the people whose linguistic and societal behavior we are studying are dead. Getting linguistic data is not merely

---

31 Many languages are characterized by their number of speakers, minority languages even more so: There are only x speakers left, the number of speakers is declining, the language is disappearing. Questions about using a language also open up another discussion: When is one a "real", legitimate speaker of a language? How much competence is needed, and competence in which variety of the language? See [97, p. 11–14].

> a matter of getting a tape recorder and recording how people speak; we will have to make do with what is left of the written record, with all the implications this has for the representativeness of the data. [99, p. 3]

Essentially, we cannot generate any new data and need to make the most out of data which still exists. This limits our approach to mathematical modeling because it means the model should not be (much) more complex than the data: If the data do not offer information on bilingualism and we aim to test all aspects of the model with empirical data, then it becomes more difficult to integrate bilingualism in the model.

In a broader sense, working with only the data on language use already available means working with a fixed concept of "language" defined by the data. In this thesis "language" and "language use" is therefore to be understood as given by the available data. The data set used in the following is the result of a survey on language use (section 4.6). We can only model this data set in the sense of modeling the statements of people who have answered the language survey. Thus it is the "commitment" to a language as evidenced in the statements which changes over time and space. These statements likely overlap with actual language use to some extent, but we cannot be sure how much because as the next section points out, language surveys do not necessarily give an "objective measurement". They can also present a simplified reality if the answer categories are inadequate and e. g. do not make it possible to indicate bilingualism. It is therefore worth looking not only into the methods of collecting data on language use, but also into their drawbacks before diving into the actual modeling.

4.1 Methods of collecting data on language use

[*This section is a translation of section 3 in [3].*]

The following section gives an overview of ways to collect data on language use as well as their limitations, especially with regard to the Austrian/Austro-Hungarian census which is used for modeling later. As this thesis focuses on mathematical modeling of language shift, this overview will be restricted to sources offering quantitative information which can be used to test such a model, and especially language surveys where people have been actively asked about their language use. Data on language use can also be gained in a number of other ways: qualitative interviews, linguistic landscaping (looking at the visibility of different languages in public space [100] e. g. on signs or graffiti or even gravestones[32]) and many more.

32 For example, one might investigate if the spelling of names has changed over time to reflect different languages and ethnic identities tied to language e. g. Germanization of Slovenian names in cemeteries in Carinthia to remove markers identifying people as members of the Slovene ethnic group [101].

4.1.1 Censuses

The census is the most comprehensive tool for gathering data on language use and records the entire population of a country in regular intervals. In addition to collecting data on the population, it also asks for information on other demographic characteristics such as occupation, household size—and language, starting from the end of the 19th century.[33] The exact questions vary depending on the country. As censuses are regulated by law and the participation is thus obligatory for all residents[34], the result (theoretically) gives information about the language use distribution in the entire country.

Censuses were and are carried out with questionnaires. In contrast to this are the so-called register-based censuses (*Registerzählungen*) in which data is drawn from existing registries without separate surveys. In register-based censuses, it is only possible to record characteristics for which registers already exist (e. g. tax data, resident registers). Languages are usually not recorded in these registers, so information about language use cannot be gained from register-based censuses. Since the census in Austria after 2001 is carried out as a register-based census, no census results on language use exist after 2001.

Censuses are generally not designed to conduct elaborate linguistic studies. They are first and foremost a tool for the state to gain information on its language distribution and possibly create policies or legislation to regulate language use [103, p. 111]. However, they can also be used for demographic studies in conjunction with linguistic questions, i. e. the demography of language (John de Vries [104] calls this "demolinguistics", see also [105]). Additionally, census data can be used simply to identify places of interest for more specific studies [106].

4.1.2 Non-governmental surveys

Non-governmental surveys form a smaller-scale counterpart to the official census. While the census covers all languages and a large territory, non-governmental surveys usually focus on answering questions with regard to only one or more (minority) languages and in a smaller area: How many people use the language *X*? These surveys are often carried out by minority organizations and the results for the number of speakers often differ greatly from those of the census (for reasons for this discrepancy, see section 4.2). They can also be carried out by linguists e. g. the Wenker survey of German dialects started by Georg Wenker. Wenker intended to map

33 The decision to include a question related to language on the census was made at the 1872 meeting of the International Statistical Congress (*Internationaler Statistischer Congress*, today the International Statistical Institute) in Saint Petersburg, where the member states of the congress agreed to ask for the "language spoken" from then on [102, p. 213ff]. Countries interpreted this differently, as can be seen from the fact that the Austrian part of Austria-Hungary asked for the language most commonly spoken in everyday life whereas the Hungarian part asked for a person's mother tongue.

34 Residents participating in the census do not necessarily have to be citizens. Sometimes, the results are even split into results for all residents and results for citizens only.

all German dialects of the German Reich (*Deutsches Kaiserreich*) and sent out questionnaires to schools asking the teachers to translate sentences into the local dialect. These questionnaires included a question on what other (non-German) languages were spoken at the school location.

4.1.3 Other statistics on language use

In addition to surveys counting the number of *people* using a language, there are statistics on the *spaces or occasions* (domains e. g. home, work, ...) in which a language is used. These are again usually geographically limited. They can consider only one language (in how many schools is language Z used as language of instruction?) or more generally register the languages used in certain domains (which languages are used in different parishes in a county? how many newspapers in language T are published and how often?).

4.2 Limits on the informative value of language surveys

[*This section is a translation of section 4 in [3].*]

Language surveys are designed to gather information on language use. However, results of language surveys rarely represent "objective" measurements unlike physical quantities like current or wavelength. The results and thus the information on language use to be gained are dependent on many factors which will be outlined in this section.

4.2.1 What is being counted?

Surveys which ask about a person's language can do this in a number of ways [103]. Mackey and Cartwright [107, p. 69–70] distinguish between two types of questions:

- Questions concerning language *skills*: What was the first language you learned? Which language(s) did you learn at school? How do you rate your skills of language X?
  Examples: mother tongue, first language (L1), foreign language, second language (L2).
- Questions concerning language *use*: Which language do you use in a certain context, at a certain time, with a certain person? Which language do you use in your family? Which language do you use at work to communicate with customers?
  Examples: main language, home language, everyday language, family language, working language.

Strictly speaking, results of language surveys can only be compared if the question is the same. Moreover, it remains open how any question on language use is interpreted by the person answering the question—and if he or she interprets it the same when answering the same question in the next survey.

> Especially in language questions based on a self-estimate of a behavioral trait which is emotionally charged, it is likely that there may be a certain amount of inconsistency in the replies. For example in the 1961 Canada Census, according to one inter-census

> study, the index of inconsistency and deviation was highest in questions referring to mother-tongue, bilingualism and ethnic origin. [107, p. 73]

Not only the question, but also the possible answers can influence the result and in the worst case represent the true language use incorrectly: Are only yes/no answers possible? Are multiple answers possible so that multilingualism remains visible? Are answers (e. g. individual language names) predefined or is there a text entry box? If answers are predefined, by what criteria are these selected? If a language has multiple names, which one is used? If there are many varieties of a language, are they all listed and accepted as answers or are they subsumed under one umbrella term (e. g. the standard language and its name)? As one can imagine, having to consider all of these things makes the development of comprehensive language surveys quite tricky. See e. g. ref. [108] for a discussion of the struggles in (re-)formulating questions concerning language use and defining possible answer options for a survey conducted by the Swiss Federal Statistical Office.

Even if any language can be given as an answer in the way the respondent wants, the answers are often condensed again during data evaluation, see [109; 110]. For example, the user manual for the Austrian census 2001 [111, p. 209–210] offers only four options for all African languages (in the sense of languages used on the African continent): *Arabisch* ("Arabic"), *Suaheli* ("Swahili"), *Westafrikan. Eingeborenensprachen* ("West-African indigenous languages") and *Afrikanische Sprachen sonstige* ("African languages, other"). The manual does not explain how the classification of the answers happens.

4.2.2 Who asks?

Answers to language-related questions do not only depend on the way the question is phrased, but also on who asks the question—in more than one way. Firstly, the answer depends on which consequences the respondent associates with their answer and if they see their anonymity guaranteed. Someone who fears negative consequences for acknowledging they use a certain language will be less inclined to answer truthfully.[35] Conversely, one might claim to speak a language only because they want to gain some benefit associated with using the language.

Secondly, even if the question is answered truthfully, it is not guaranteed that this answer will be registered. Questionnaires can be manipulated which has happened e. g. in the 1910 census of the Cisleithanian states of Austria-Hungary, see [112, p. 235]. If the questionnaire is not filled out by the respondent themselves but the survey is done as an interview, then there is another possibility of manipulation: the way the interviewer asks the question and writes down the answer. Brix [112, p. 235] gives again the example of the census 1910 where Italian enumerators (census takers) did not ask, "What language do you speak?" but rather, "Lei parla italiano?" ("Do

35 This is a possible explanation for the discrepancy between census results and results from surveys done by minority organizations when it comes to speaker numbers of minority languages.

you speak Italian?"). Even if the question is posed as intended, it is up to the census taker how they record the answer.

4.2.3 What do the results tell us?

To interpret results of language surveys, it is not only necessary to know the exact questions (as with all qualitative surveys), but also the circumstances surrounding the survey [104]. Owing to these two factors and the uncertainty of how the respondents interpret the questions, results from language surveys are only informative to an extent, especially when it comes to unspecific questions about one's "everyday language".

> The statistics on the answers to language-related questions may be presented in such a way as to show the number of people in any given area using a given language. Such a presentation means: 'At this place at this time so many people have made this statement about their language use'. [107, p. 76]

Results of language surveys can therefore be seen as primarily a statement about one language. This is supported by Brix [113], according to whom the census results about the everyday language of Austria-Hungary 1880–1910 should be seen as results on nationality rather than actual speaker numbers.[36]

Series of language surveys are often used to track the development of a language. It is tempting to follow the speaker numbers according to the census over time and deduce that a language is "dying out" or that the geographical area where a language is used becomes smaller. This might well be the case, but it could also simply point to mobility in the population as two completely different scenarios can lead to the same census results [107, p. 76,86]: in one case, language shift has actually occurred and in the other a redistribution of the population has happened.

Despite these limitations, census data often offers the only way to track quantitatively the development of languages (or statements about languages). Gathering data about language use with the same temporal and spatial scope solely by linguistic fieldwork is very difficult or sometimes plain impossible (e. g. when it comes to data from the past).

4.3 Legal provisions in connection with language surveys

[*This section is a translation of section 4.4 in [3].*]

As mentioned above, the answers to questions about language heavily depend on who is asking and if (perceived) advantages or disadvantages might result from a truthful answer. Ideally, such questions will be answered truthfully without manipulation by others. However, language does

36 Statements about language are often equated with speaker numbers, which is a simplification in any case. On the other hand, the way the question is posed often does not allow for a more nuanced view.

not exist in an ideological vacuum and very real consequences result from stating that one speaks a certain language on a survey [114]. This is especially true for speakers of so-called minority languages. Minorities[37] are granted the right to use their language in many situations by law.[38] These include the right to use the language as an official language for administrative procedures, the right to school instruction and to topographical signs (place names or street signs) in their language. The legal regulations are often based on the results of language surveys and make the rights dependent partially on a numerical threshold (the right exists only if over x% of the population uses the language).[39] In South Tyrol, resources such as government jobs or the composition of the parliament are determined according to the composition of the three language groups (German, Italian, Ladin) [117, p. 124–127]. Austrian examples include the minority protection provision of the Treaty of Saint-Germain-en-Laye (1919) and the provision on topographic signs in the Ethnic Groups Act (Volksgruppengesetz 1976).

> Was das öffentliche Unterrichtswesen anlangt, wird die österreichische Regierung in den Städten und Bezirken, wo **eine verhältnismäßig beträchtliche Zahl** anderssprachiger als deutscher österreichischer Staatsangehöriger wohnt, angemessene Erleichterungen gewähren [...]
>
> (Treaty of Saint-German-en-Laye, Art. 68, emphasis K. P.: "a relatively considerable number [of Austrian citizens using a language other than German]")

> Durch Verordnungen [...] sind [...] festzulegen: [...] Die Gebietsteile, in denen wegen **der verhältnismäßig beträchtlichen Zahl (ein Viertel)** der dort wohnhaften Volksgruppenangehörigen topographische Bezeichnungen zweisprachig anzubringen sind.
>
> (Ethnic Groups Act §2 (1) 2. (original version[40] BGBl. 396/1976), emphasis K. P.: "relatively considerable number (one quarter) [of members of an ethnic group living there]")

---

37 I am using the terms "minority" and "minority group" here rather than "ethnic groups" because in most cases, these groups are actually disadvantaged in contrast to the majority, i. e. in a position of less power.

38 Provisions about minority *languages* are often an addendum to laws regulating the rights of minority *groups*.

39 Rather than restricting rights to members of a minority group who actively proclaim being a member of that minority group (*Bekenntnisprinzip*, "admission principle"), laws might also grant the rights for a certain geographical area (*Territorialprinzip*, "territory principle"). See also [30, p. 12–14]. For example, there is a long-standing discussion on how Article 7 of the Austrian State Treaty (1955) should be interpreted: Should minority rights be dependent on the numerical size of the minority group or only on the area where the minority is present? (see [115, p. 43–46]; [116, p. 175–177])

40 The 25% limit to granting rights was deemed unconstitutional by the Constitutional Court (*Verfassungsgerichtshof*) in 2001. In 2011, the Austrian Federal Government, the government of the state of Carinthia and representatives of the Slovene minority in Austria agreed to a limit of 17.5% (based on the 2001 census) to identify places where bilingual signs must be provided. In its current version (January 31, 2019), the law explicitly lists villages where

Even if the law does not explicitly mention a number (such as in the Treaty of Saint-German-en-Laye), judicial interpretation will likely set one (e. g. a limit of 20% for minority rights in the Treaty of Saint-Germain [116, p. 176]). Determination of an exact number is a lengthy process and subject of much debate: Who can decide what is the limit to a group being granted rights? If the group is scattered across a large area rather than in one densely populated spot, limits—except very low ones—will inevitably lead to most members of the group not being granted rights.

The basis for judicial interpretation is often taken from results of language surveys. For example, the decision on whether more than one quarter of the population belongs to a minority group (and thus whether bilingual topographic signs have to be installed according to the Ethnic Groups Act, see above) is based on results from the language-related question on the Austrian census.[41] Wherever people in Carinthia stated they use Slovenian as their everyday language, they thus automatically became members of the Slovene minority, whether they wanted to or not. For this reason, language surveys and especially the census are also a political vote for many speakers of minority languages. For them (in contrast to speakers of majority languages), personal rights are dependent on "winning the census" [118, p. 194f] because the interpretation of results of language surveys mixes language and ethnicity. The next section will take a closer look at this mixing.

4.4 Mixing language and ethnicity in surveys

[*This section is a translation of section 4.5 in [3].*]

Language surveys ask for a person's language and count only the language. So much for the theory, but reality sometimes looks quite different. The situation becomes even more delicate if results from language surveys are used to directly deduce the size of a minority group. Not every member of a minority group uses the minority language and vice versa, not everyone who uses the minority language is a part of the minority group. The line between minority and minority language is just as difficult to draw as defining when a language is considered a minority language. This is also reflected in the evaluation criteria of census results. Depending on the source, Carinthian Slovenes are all those who stated one of the following as their everyday

bilingual topographic signs are to be installed.

41 The definition according to the Ethnic Groups Act is as follows: "Ethnic groups in terms of this Federal Act are such groups of Austrian citizens living in parts of the Federal territory and having a language other than German as mother tongue and having traditions of their own." (Ethnic Groups Act/VolksgruppenG §1 (2), own translation) From this definition, the language but not the traditions can be measured "objectively" (through the census). It should be noted that the Ethnic Groups Act defines ethnic groups by their *mother tongue* as a characteristic, but the census asked for the *everyday language* from 1951 to 2001.

language in the census:[42]

- “Slovenian” (and only that)
- “Slovenian” and “Windisch”[43] (no combinations with German)
- “Slovenian” and the combination “Slovenian and German” (and sometimes “German and Slovenian”) or
- “Slovenian”, “Windisch” and the combination “Slovenian and Windisch” (and sometimes “Windisch and Slovenian”) and possibly all combinations of these with German

Nevertheless, language is often a defining characteristic of a minority and is easier to measure than cultural attributes—which is why language is so frequently used to identify the minority group. This argumentation can be found as early as 1923 in Wilhelm Winkler’s theory on the significance of statistics in the protection of minorities [122], see also [123].[44]

In censuses, there are sometimes separate questions about language and ethnicity (such as in the Hungarian Census 2011 where different questions ask for citizenship, spoken language and “nationality”). Even in these cases language can serve as a criterion for ethnicity. The Canadian census 1971 asked for ethnic origin and three more questions about linguistic features ([107, p. 78]). For the question on ethnic origin, the respondent is instructed to use language as a criterion:

---

42 The number also changes depending on if the whole area of Carinthia or only the so-called “bilingual part” of Carinthia is considered. The bilingual part of Carinthia is not a well-defined term; in [119] it is defined as the scope of the *Minderheiten-Schulgesetz* (Minorities School Act) for Carinthia without the city of Villach/Beljak.

“Bilingual” here refers to societal bilingualism/multilingualism, the linguistic situation of a society/speech community. This is to be distinguished from individual multilingualism which refers to a single person’s ability to use multiple languages in communication. Bilingual speakers might be found everywhere in Carinthia, but societal bilingualism is seen as restricted to the southern part. The two language shift situations we will investigate in detail later have societal bilingualism/multilingualism, and the census (if it includes bilingualism as a possible option for answering) tells us about the individual bilingualism/multilingualism of people.

43 “Windisch” is another name for the Slovenian language or one of its dialects [120]. With regard to the Slovenian ethnic group in Carinthia, the term also has a political connotation. In the discourse after the referendum 1920 about whether Carinthia should remain part of Austria, the ethnic group was divided according to their conviction [121]: into the “German-friendly Slovenes” who were committed to Austria (the “Windisch”) and the others which were more drawn to today’s state of Slovenia, the “Slovenes”. Nowadays, the term is partly used pejoratively for the Slovenian ethnic group, but also as a positive self-identification of members of the same ethnic group.

“Windisch” is its own answer category on the census or rather it is a separate evaluation category from 1939 on. It is also the only non-standard name visible in the results from 1939 to 2001. Theoretically it is possible to write any dialect or variety of a language (e. g. varieties of German like Viennese or Tyrolean, or Carinthian Slovenian or its dialects which are all a variety of Slovenian) in the text box asking for a person’s language, but the Austrian census results do not show any of these except for Windisch. We thus cannot say how many people called their variety of German something other than “German” on the questionnaire, or how many people used “Carinthian Slovenian”.

44 Winkler also argued for recording the *Denksprache* (“thought language”) instead of the mother tongue, because the latter represented an earlier state rather than the current one [122, p. 51,53–54]; [123, p. 86].

> Use as a guide if applicable in your case: 1. The language you spoke on this continent, if you were born outside Canada. 2. If born in Canada, the language spoken by your ancestor on the male side when he came here. (Instruction Booklet, 1971 Census of Canada, cited in [107, p. 78])

The Austrian 1934 census did not separately ask for language and ethnicity, but about the language of the culture one feels they belong to [30, p. 69], and once again language and ethnicity were equated. After the annexation of Austria (*Anschluss*), the 1939 census of the German Reich asked separately about language and ethnicity. The result was that about 43,000 people stated they used the Slovenian language (or were bilingual)[45], but only about 7,700 stated they were of Slovenian ethnicity [124, p. 68–69].

On the other hand, in Austria today nobody is obliged to claim to be part of a certain ethnic group (Ethnic Groups Act/VolksgruppenG §1 (3)). The same is the case in the Framework Convention for the Protection of National Minorities of the Council of Europe.[46] The Austrian census therefore cannot and should not be a minority survey. Even so, results of the census, more specifically the results of the question about language are used in jurisdiction. Similarly, in statistical publications the number of persons using a language according to the census is equated with the number of people belonging to a minority group e. g. in [119]. The seemingly objective language surveys are thus used for non-linguistic questions and problems which leads to even more criticism of language surveys.

4.5 Criticism of language surveys by minorities

[*This section is a translation of section 4.6 in [3].*]

Language surveys and their methodology—in addition to all above-mentioned issues—have been criticized especially by members of minority groups. They see themselves underrepresented in surveys, particularly in the census, and argue that the census does not give a true picture of where minority groups are present—a valid point, considering the mixing between language and ethnicity done in the interpretation of the results (see section 4.4). Another criticism against recording language use via the census is that the question is often suggestive: The Austrian census asked about the *Umgangssprache* ("everyday language") from 1951 to 2001, i. e. the language

45 This sum includes "Slovenian" and "Windisch" as well as all combinations with them. For the term "Windisch", see footnote 43 on page 47.

46 "Every person belonging to a national minority shall have the right freely to choose to be treated or not to be treated as such and no disadvantage shall result from this choice or from the exercise of the rights which are connected to that choice" (Framework Convention for the Protection of National Minorities, Art. 3 (1)). The framework convention has been ratified by 39 countries including Austria and Hungary.

that is used in everyday life.[47] However, speakers of minority languages might not have many opportunities to use their language in everyday life: the university does not offer study programs in the minority language, there are no jobs which use the minority language, and friends and acquaintances do not understand the minority language but one still wants to communicate with them. Instead, Carinthian Slovenes have demanded that the census ask about the mother tongue instead, but they have also noted that fewer and fewer Slovenian-speaking people in Carinthia admit to their mother tongue due to the political climate [116, p. 204].

Minorities often completely reject language surveys as they fear that their "official" size might be smaller than the size proclaimed by the minority group itself [125, p. 10]. Special surveys by the state whose only purpose is to gain information on the size of a minority group (even if it is again by proxy by asking about language) are boycotted [116, p. 188–189]; [115, p. 44–46], e. g. the *Geheime Erhebung der Muttersprache* ("secret survey of mother tongue")[48] 1976 in Austria in preparation for the passing of the Ethnic Groups Act. This means that there are large differences between the results of official censuses and surveys conducted by minority organizations. *Enotna Lista*, the Slovenian unity party list in Carinthia, notes that according to the 1991 census, approximately 14,000 people in Carinthia used Slovenian, but a private survey in bilingual parishes resulted in 50,000 people using Slovenian [126].

As the size of a minority has consequences in terms of subsidies and protective rights which depend on quotas, minority organizations are interested in having as many people affiliated with the group as possible. They argue with results of their own surveys which tend to produce higher numbers than the official census. On the other hand they call for minority protection without any quotas at all—but even then, it has to be determined where the minority (language) is present. The need for surveys (in whatever form) therefore remains.

4.6 Data on language use in this thesis

After discussing how language use can be recorded and what the limitations of self-reported data on language use are, a final question remains: What data is used to test the model for language shift developed in this thesis?

Language shift is a process, i. e. we need data about language use for at least two time points to compare them and identify differences. Additionally, we can consider language shift in space in the sense of geolinguistics: In what places does language shift occur first/last/not at all? We therefore ideally want to have data with a good temporal and spatial resolution as the basis for

---

47 The official explanation concerning the intended interpretation of "everyday language" varied slightly between censuses.

48 "Secret" here in the sense of "anonymous", meaning that on the ballot only the language and nothing else was recorded. In contrast to this, censuses ask for all manner of personal data (name, age, language, etc.) on one sheet, meaning that results can theoretically (under disregard of data protection) be traced back to individuals.

our model of language shift. Such detailed data combined with mathematical modeling enable us to gain a complete and detailed picture of the process of changing language use. Good spatial and temporal resolution means in this case:

- covering a large timescale with many data points in-between (e. g. data for every year, not only every 25 years)
- covering a large geographic area, but for very small units (e. g. data for every village, not only for every county or state)

In reality, data conforming to these desires are hard to find because data collection is very labor-intensive and time-consuming. Many linguistic corpora (collections of data) have been recorded only at one point in time or covered only a small geographic area. They might have asked only few speakers per village or focused on variation of one language rather than the use of different languages.

An alternative is using census data. Participation in the census is usually mandatory by law,[49] so the results cover the complete population and the entire area of a state. This means the spatial resolution is very good. Censuses are not held every year, but they are held in regular intervals, so the temporal resolution is also good. For example, the Austrian census data on language use covers the time from 1880 until 2001 in approximately 10-year intervals (figure 3).[50] Data for Austria(-Hungary) is available for most of the years in which the census was conducted[51] on the level of individual population units and there is no minimum unit size, i. e. even for hamlets with four inhabitants, information on their language use according to the census is available. The questionnaires and questions about language are well-documented for each year. The data situation (in the sense of: availability of data) is therefore very good, which is why census data from Austria(-Hungary) was chosen as the primary data source for mathematical modeling, despite the limitations pointed out in earlier sections (section 4.2).

Figure 3 gives an overview of available census data for Austria(-Hungary) and the nature of their questions about language. The graphic shows information on data available on the smallest possible spatial resolution, i. e. data which is available for small geographical units (ideally per settlement) and not only the whole country. Such fine-grained data is needed for modeling language shift over time and space. In addition to census data, other sources of data on language

49 While general participation is mandatory, some questions may not be compulsory to answer, e. g. the four questions on ethnicity and language (mother tongue and languages used with family members or friends) in the Hungarian 2001 census [127, p. 97].

50 Earlier censuses had included only data on the number of residents. A census in 1869 included questions on “cultural” characteristics such as religion, but not on language, see [128–130]. After 2001, the census methodology was changed to register-based and no more data on language use is available from the census.

51 For example, the original data sets of the 1923 Austrian census were lost in a fire and only few preliminary results remain [30, p. 64].

use were integrated where available. These include data on languages taught in school [131] and parish language [132] in Austria. For Hungary, no such additional data was available.

The time periods used for modeling are: 1880–1910 and 1971–2001 from the Austrian census, and 1880–1930 from the Hungarian census. These time frames were chosen because for the Austrian census, the way of asking about language during these periods remain the same (*Umgangssprache*, the language used most commonly in everyday life). Data from 1951 and 1961 where the question was also the same could not be obtained for individual settlements.[52] For the Hungarian census, a period starting from the same year as for Austria was chosen, beginning in 1880 in this case where data is again available per settlement. The question about language remains the same (*anyanyelv*, "mother tongue") throughout until 1930 which marks the end of the modeling period. No further data was included because the 1941 census is likely to be influenced by politics (although no obvious deviation is visible on county level, compare figure 29) and the number of German speakers drastically declined after World War II as German-speaking settlers were forced to leave the country or voluntarily resettled.

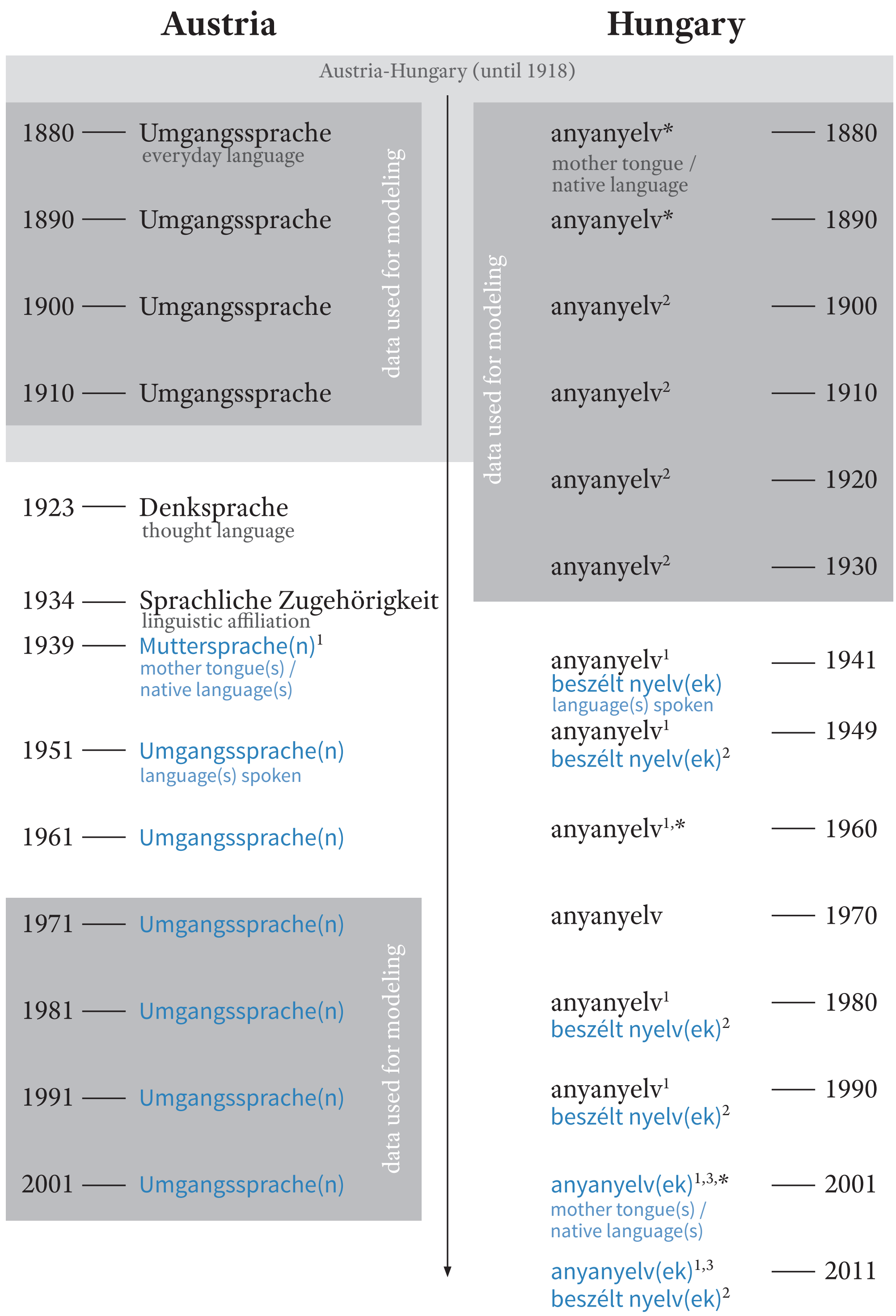
Data used for modeling were either freely available on the internet (the complete Hungarian census[53]) or taken from the official publications (historical censuses of the Austrian part of Austria-Hungary [133–136]). A complete list of original sources can be found on page 123 of this thesis. In these cases, the data had to be typed up manually to bring them into a suitable format for modeling.[54] The Austrian data for 1971–2001 were obtained by a special evaluation commissioned from Statistics Austria (already in a machine-readable format). Data for this time period for individual settlements cannot be published in full due to data protection reasons.

---

52 Here and in the following, "settlement" always refers to the smallest self-contained entity designated in the census, regardless of whether it legally was a village, town or hamlet.

53 Scans of census publications until 1990 are available from the online collection HUNGARICANA: https://library.hungaricana.hu/en/collection/kozponti_statisztikai_hivatal_nepszamlalasi_digitalis_adattar/. Data for the most recent censuses is available from http://www.nepszamlalas2001.hu (2001) and http://www.ksh.hu/nepszamlalas/ (2011).

54 The digitized datasets have been deposited in the online repository *figshare*:
Carinthia, Austria: https://doi.org/10.6084/m9.figshare.4535399
Southern Hungary: https://doi.org/10.6084/m9.figshare.6115634



Notes:

* Some additional data on language(s) spoken available (in a different resolution).

** Some additional data on the mother tongue available (in a different resolution).

1 Separate question on nationality (same categories as mother tongue).

2 Separate indication on whether Hungarian/a language other than the native language is spoken in addition.

3 Separate question on the language used with family members and friends.

Census data which include the possibility of answering more than one language are indicated by blue sans-serif font.

Information shown for data available with the smallest possible spatial resolution (generally per settlement—not included: data only available for the whole country). Not included: other related questions asked on the questionnaire, but for which no data is shown in the results.

Figure 3: Overview of censuses in Austria(-Hungary) with data on language use and the way the language-related question is phrased. Light grey box: Time period of Austria-Hungary. Darker grey boxes: Data used in this thesis for modeling language shift. Sources: [30; 127] as well as the official publications of results (see page 123).

5 The lab: Language contact in Austria(-Hungary)

After obtaining the data, the next step is to choose a sample: one or more specific situation(s) and area(s) in Austria(-Hungary) on which to test the model. These situations are the laboratory where data from an experiment was gathered, in this case data on language use from a very large and long-running experiment (since 1880).[55]

We will be looking at two language contact situations in today's Austria and today's Hungary (shown on the map in figure 4):

(I) Slovenian–German in southern Carinthia (Austria). The first Slavic peoples settled in Carinthia around the 6th century [116], so there has long been language contact in this area—for the most part a peaceful coexistence. This changed with the rise of nationalist tendencies at the end of the 19th century, when the first conflicts between the two language (and ethnic) groups began which to some extent continued until recently.

(II) German–Hungarian in southern Hungary (counties Baranya and Tolna). The German-speaking minority in Hungary is largely the result of historical settlement policies. In the 17th and 18th century after the expulsion of the Ottoman Empire, large parts of Hungary were populated by settlers from Germany and Austria who were motivated by the Habsburg Government to immigrate [137]. From the end of the 19th century, these *Ungarndeutsche* (Hungarian Germans) were increasingly pressured to adopt Hungarian culture and the Hungarian language following the Kingdom of Hungary's assimilation policy (the so-called Magyarization). Nowadays, the majority of the remaining Hungarian Germans lives in the counties Baranya/Branau and Tolna/Tolnau.

As can be seen on the map in figure 4 (and in more detail in figure 14 and figure 30), we restrict ourselves to areas within today's country borders, i. e. no parts of Carinthia or Baranya which were ceded to other countries after 1918.

In order to start modeling language shift in these two situations, it is helpful to have a look at the data first rather than blindly throwing mathematics at them. Looking at the data gives us the opportunity to create data-led hypotheses instead of making random assumptions. Analyzing data is also crucial for model-building itself as the data structure determines modeling possibilities. Coming back to an earlier example: If the data do not contain explicit information on speaker bilingualism, it becomes much harder to test model predictions on bilingualism because in this case, only "secondary expressions" (i. e. other variables on which bilingualism has an effect, e. g.

55 An experiment in which none of the parameters can be changed and which cannot be repeated.

Europe 1880

Europe today (2019)

Figure 4: Areas used for data evaluation and modeling in Carinthia and in southern Hungary, respectively. Top: Location of the two areas within Europe in 1880, bottom: within Europe today (2019). The areas are shown in more detail in figure 14 and figure 30. Map data: [138; 139].

schooling options) are testable—if there is data for these. Such limitations should be considered before creating a model.

In addition to looking at the data for each situation separately, it is also worthwhile to do a comparative data evaluation when we model more than one language contact situation. After all, language shift processes happen differently in different situations. This means that any model for language shift is likely to have situation-dependent parameters as well. To uncover these parameters or even see if there are any differences at all which might affect modeling, comparison is necessary.

We will therefore do both: We will first do a comparative data analysis and see what consequences different circumstances and local conditions in different language contact situations have for mathematical modeling (section 6). Then, we will look at each situation separately (Austria in section 7, Hungary section 8).

6 Comparative data evaluation: What consequences do local conditions have for modeling?

The comparative data evaluation will focus on a time period where the data can indeed be compared: 1880 until 1910, when there was no separate Austria and Hungary but only the Dual Monarchy of Austria-Hungary (the Austro-Hungarian Empire). In this time period, data is available for individual settlements and the phrasing of the language-related question on the census does not change. Additionally, it is a period of (relative) stability insofar as that no drastic political events such as World War I occur.[56]

6.1 Historical background: multilingual Austria-Hungary / Setting the data in context

Austria-Hungary was a historical region and political entity in Central and Eastern Europe which emerged from territories ruled by the Habsburg Empire. From 1804 to 1867, the so-called Empire of Austria existed in which the whole area was run as one monarchist state. In 1867 the Empire of Austria was converted into the Dual Monarchy of Austria-Hungary following the Austro-Hungarian Compromise. The Compromise led to a division of the Habsburg Empire after demands for autonomy had repeatedly arisen within the state. From 1867 on, there were two equal parts of the Empire (figure 5):

- Cisleithania ("Austria"/the Austrian lands, *Die im Reichsrat vertretenen Königreiche und Länder*/"The Kingdoms and Lands Represented in the Imperial Council")
- Transleithania ("Hungary"/the Hungarian lands, *a Szent Korona Országai*/"Lands of the Crown of Saint Stephen")
- In addition, the territory of Bosnia and Herzegovina was under Austro-Hungarian rule (and later in 1908 fully annexed).

Both parts were politically independent (e. g. separate parliaments and no common citizenship), although the military and foreign affairs were still handled jointly and there was a lot of regional diversity in administration [140–142]. In 1918, Austria-Hungary dissolved following the defeat in World War I.

56 It is not a period of complete stability as nationalism was already present and the different ethnic groups began to demand independence which was a factor in the eventual dissolution of Austria-Hungary.

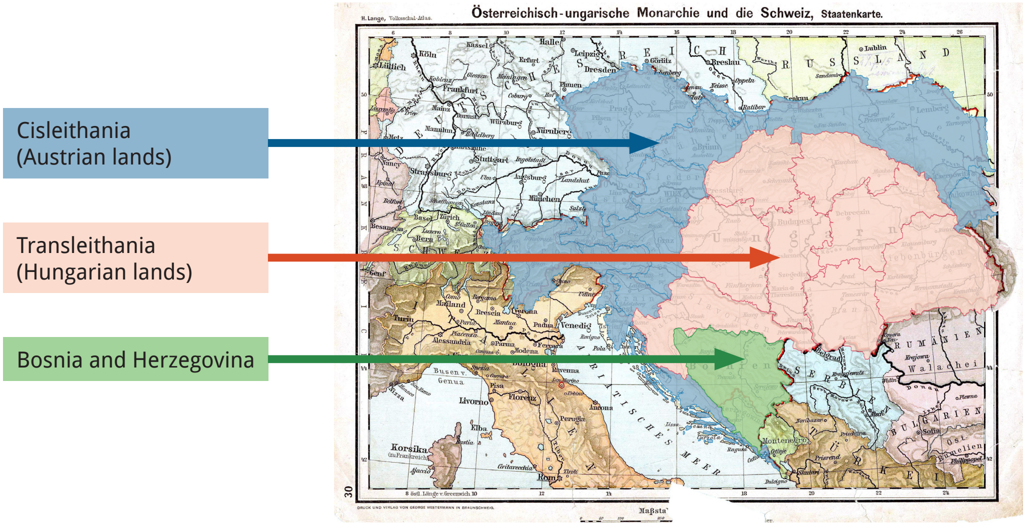


Figure 5: Map of Austria-Hungary and its position in Europe. This figure shows the two halves of the Empire as well as Bosnia and Herzegovina which was ruled by Austria-Hungary. Map data: [138; 139; 143].

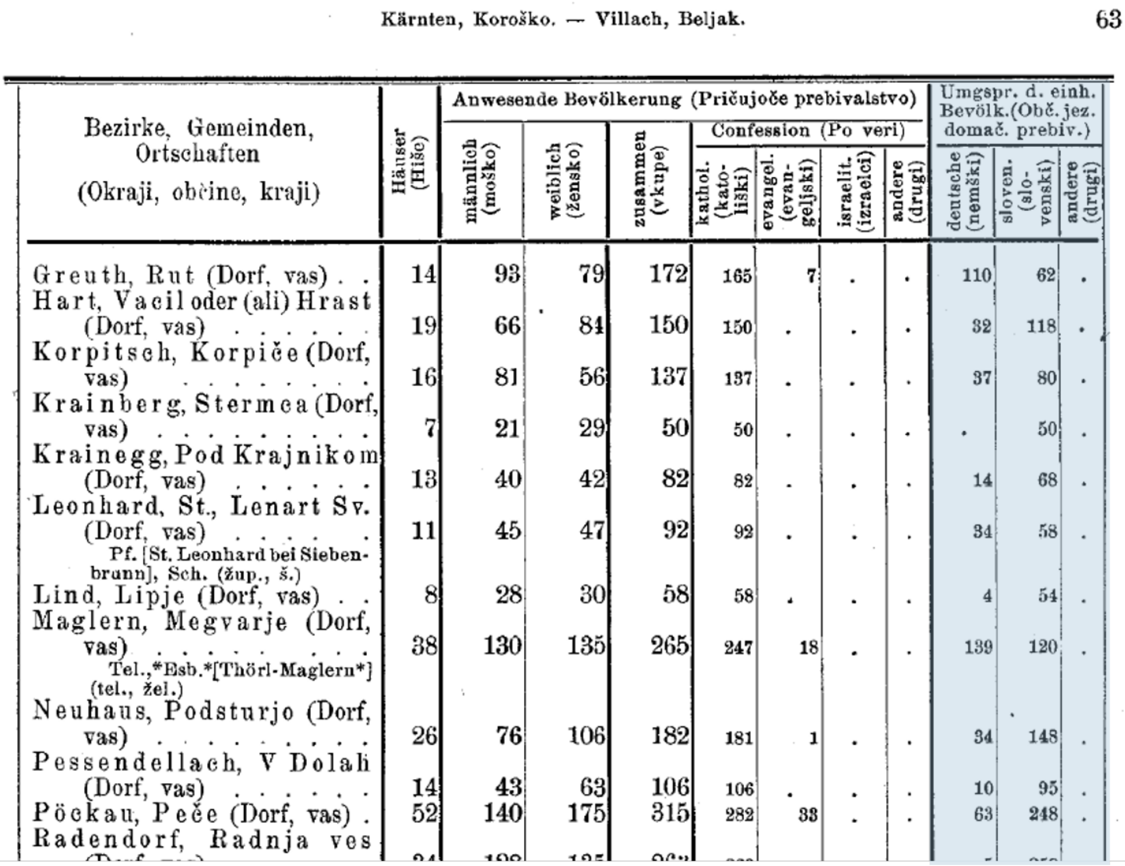
The two parts of Austria-Hungary also conducted separate censuses which differed in some regards. Table 2 gives a comparison of both censuses. Figure 6 and figure 7 show examples of results from the Austrian and the Hungarian census. The most obvious difference concerning language seems to be that the Austrian census asked for *Umgangssprache*, 'everyday language', whereas the Hungarian one asked for *anyanyelv*, 'mother tongue'. According to the official Hungarian interpretation, "mother tongue" was the language "a person sees as their own, speaks best and prefers" as well as the language "which has been acquired in pre-school, at school or through other social contact".[57] The Austrian question for "everyday language" was to be answered with the language "a person usually uses during common interaction".[58]

57 Own translations of the explanations cited in [144]: „[die Sprache] die die betreffende Person als ihre eigene ansieht, und in der sie am besten und am liebsten spricht" und „[die Sprache, die sie sich] im Kindergarten, in der Schule oder durch sonstige gesellschaftliche Kontakte angeeignet hat"

58 Own translation of the explanation cited in [144]: „[die Sprache] deren sich die Person im gewöhnlichen Umgange bedient"

	Austria	**Hungary**
years	1880, 1890, 1900, 1910	1880, 1890, 1900, 1910
spatial resolution	per settlement	per settlement
people covered	present citizens	present citizens
language-related question	*Umgangssprache* (‘everyday language’)	*anyanyelv* (‘mother tongue’)
number of possible answers	only one language 9 available choices	only one language free-text field
results categories	German + 1 (rarely 2) language(s) + “other”	Hungarian + multiple languages + “other”
publications	bilingual (in crown lands with non-German language groups)	mostly monolingual*

Table 2: Comparison of the Austrian and Hungarian census in Austria-Hungary during the time period 1880–1910. * Some results were published in Hungarian and German.



Kärnten, Koroško. — Villach, Beljak. 63

Bezirke, Gemeinden, Ortschaften (Okraji, občine, kraji)	Häuser (Hiše)	Anwesende Bevölkerung (Pričujoče prebivalstvo)							Umgspr. d. einh. Bevölk. (Obč. jez. domač. prebiv.)		
					Confession (Po veri)						
		männlich (moško)	weiblich (žensko)	zusammen (vkupe)	kathol. (katoliški)	evangel. (evangeljski)	israelit. (izraelci)	andere (drugi)	deutsche (nemški)	sloven. (slovenski)	andere (drugi)
Greuth, Rut (Dorf, vas) . .	14	93	79	172	165	7	.	.	110	62	.
Hart, Vacil oder (ali) Hrast (Dorf, vas)	19	66	84	150	150	.	.	.	32	118	.
Korpitsch, Korpiče (Dorf, vas)	16	81	56	137	137	.	.	.	37	80	.
Krainberg, Stermca (Dorf, vas)	7	21	29	50	50	.	.	.	.	50	.
Krainegg, Pod Krajnikom (Dorf, vas)	13	40	42	82	82	.	.	.	14	68	.
Leonhard, St., Lenart Sv. (Dorf, vas) . . . . . Pf. [St. Leonhard bei Siebenbrunn], Sch. (žup., š.)	11	45	47	92	92	.	.	.	34	58	.
Lind, Lipje (Dorf, vas) . .	8	28	30	58	58	.	.	.	4	54	.
Maglern, Megvarje (Dorf, vas) . . . . Tel.,*Esb.*[Thörl-Maglern*] (tel., žel.)	38	130	135	265	247	18	.	.	139	120	.
Neuhaus, Podsturjo (Dorf, vas)	26	76	106	182	181	1	.	.	34	148	.
Pessendellach, V Dolah (Dorf, vas)	14	43	63	106	106	.	.	.	10	95	.
Pöckau, Peče (Dorf, vas) .	52	140	175	315	282	33	.	.	63	248	.
Radendorf, Radnja ves [illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]				[illegible]	[illegible]	

Figure 6: A page showing results of the Austrian census 1890 [134, p. 63]. Columns containing data on language use have been colored.

30

Baranya megye.

Folyó szám	A község neve	A házak száma	A jelenlevő polg. népesség száma	A jelenlevő polgári népesség: anyanyelve szerint: magyar	német	tót	oláh	ruthén	horvát-szerb	egyéb hazai nyelvű	külföldi nyelvű	beszélni nem tudó	hitfelekezete szerint: katholikus: római	katholikus: görög	gör.-keleti	evangelikus: ágostai	evangelikus: helvét	unitárius	izraelita	egyéb hit-felekezetű	olvasni irni tudása szerint
	6. Baranyamegye.																				
	I. Pécsi járás.																				
1	Aranyos	94	544	530	6	.	.	.	.	.	.	8	536	.	.	.	.	.	8	.	138
2	Árpád	81	468	76	360	.	.	.	5	.	1	26	466	.	.	.	2	.	.	.	160
3	Áta	52	257	119	22	.	.	.	101	.	.	15	239	.	.	5	7	.	6	.	38
4	Bánya (Uj-)....	58	337	8	313	1	.	.	.	.	.	15	336	.	.	.	1	.	.	.	152
5	Belvárd	142	723	428	43	.	.	.	226	2	.	24	370	.	.	.	343	.	10	.	321
6	Birján.........	100	524	95	81	11	.	.	308	.	.	29	507	.	.	.	6	.	11	.	145
7	Bogád (Püspök-)	142	730	694	15	.	.	.	2	.	.	19	721	.	.	.	.	.	9	.	282
8	Bosta..........	66	327	313	.	.	.	.	.	.	.	14	318	.	.	.	9	.	.	.	66
9	Budmér (Kis-)..	53	326	5	305	.	.	.	1	.	2	13	321	.	1	.	.	.	4	.	192

Figure 7: A page showing results of the Hungarian census 1880 [134, p. 63]. Columns containing data on language use have been colored.

From its beginning, Austria-Hungary (and the former Habsburg Empire) was a multinational state extending over multiple national groups and cultures[59]. We therefore also have many different languages used in Austria-Hungary, as is visible in figure 8. Figure 8 is an ethno-linguistic map, meaning that it shows the ethnic groups of Austria-Hungary by equating ethnic group with the language given in the census (see section 4.4 for further discussion of this practice). If we consider this image as a purely linguistic map (each color a language), then we see a plethora of language contact situations wherever different colors meet. In many of these language contact situations, language shift occurred and is visible in the census data.

As mentioned above, the goal of this thesis is to take a closer look at two of these situations: Slovenian–German in Carinthia and German–Hungarian in Hungary. These two are situated in different halves of Austria-Hungary, meaning we can compare the different circumstances in Austria and Hungary as well as the consequences of these differences for quantitative mathematical modeling. I will focus here on comparison with regard to two circumstances: language policy, and population and settlement structure.

6.2 Influence of language policy

Language policy (or politics)[60] refers to the sum of all measures and mechanisms which affect (the status of) one or multiple language(s) [146, p. 3ff]; [147]. These measures regulate how languages are used in practice, especially in multilingual societies or organizations. Examples of language policies are laws which determine the official language(s) of a state, or the language(s) used for certain administrative processes. Language policies can also be used to assign a higher social value (prestige) to one language or variety of a language by actively promoting it or forbidding the use of other languages. Interventions and efforts to "save" minority languages also fall under this term. Language policies can be explicit (as laws) or implicit (the value assigned by the speakers to a language or variety).

59 The terms "nation" and "culture" are used here as categories to distinguish groups of people from each other, to illustrate that the people living on the territory covered by the political entity Austria-Hungary were diverse. They assume that people in the group share certain characteristics such as language, ethnicity, territory or traditions. However, it should be kept in mind that human identity is complex and assigning people to a group based on their outward characteristics is not always in line with how they identify. We cannot assume that "a nation" or "a culture" is one big homogeneous group and that anyone who e. g. uses a certain language is automatically part of that group. As such, these terms (and similar ones e. g. "ethnic group") should be regarded critically.

60 Language politics as a term is usually used strictly for political/legal measures, i. e. intentional measures from an authoritative or regulatory body—coming from "top-down". More generally, language policy can also begin from the "bottom-up", i. e. from the speakers themselves, and it does not always have to be deliberate. The field is also often subsumed as "language planning and policy (LPP)" and has been existing as an academic discipline since the 1960s, although interventions to influence language can be found earlier [145].

Figure 8: Distribution of ethnic groups of Austria-Hungary in 1910, based on census results on language. Shaded blue and red areas denote the two language contact situations studied in this thesis. Map: based on [148].

In German, a further distinction is made between *Sprachenpolitik* and *Sprachpolitik* [149, s.v. "Sprachenpolitik", "Sprachpolitik"]: *Sprachenpolitik* concerns the interplay of multiple languages, whereas *Sprachpolitik* is targeted at only one language e. g. languge planning or language regulation to maintain a certain standardization of the language. However, the line between the two is sometimes not an easy one to draw.

In Austria-Hungary, the two parts of the Empire pursued different language policies [137; 142; 150–152]. In the Austrian part, German was the prestige language and had a high social status. German therefore held a special position and was also the primarily used language, e. g. in the military (which was jointly handled with the Hungarian part, i. e. there was only one military with German as the official language used for drill commands—which led to conflicts with the Hungarians [153]). Nevertheless, by law all languages in the Austrian part were equal as established in the December Constitution (*Dezemberverfassung*) of 1867, even though in some areas German was the de-facto official language as other languages were simply not recognized by governing bodies. Overall, there was a relatively liberal acceptance of multilingualism in the sense that it was not actively discouraged by the government. In some cases, this liberal approach paradoxically led to linguistic segregation: as no citizen could be forced to learn the language of another ethnic group, this meant that in bilingual crown-lands, it was forbidden to make both languages mandatory subjects in school [154].

Circumstances were different in the Hungarian part of the Empire, which comprised the Kingdom of Hungary and the Kingdom of Croatia-Slavonia.[61] In contrast to the Austrian part, Hungarian was legally established as the only official language of the Kingdom of Hungary by the Hungarian Nationalities Law 1868 (Act XLIV/1868), although the law granted rights to use other languages in education, local administration and justice.[62] The situation is perhaps best described as a "language-led" patriotism—only those who used Hungarian were "true Hungarians". At the same time, unlike in the Austrian part, there were also large-scale governmental efforts to enforce Hungarian as the only language. As part of this so-called Magyarization, teachers for example offered prizes if children performed well in Hungarian language-classes and showed patriotism [155, p. 26]. In general, there was a much stronger pressure towards a monolingual society in which one language was hierarchically at the top, namely Hungarian.

The question is now if these two different language policies in the Austrian part and the Hungarian part are reflected in the census results. One would probably expect that in the more liberal Austrian half the minority language Slovenian will decrease less than the minority language German in the Hungarian part. However, when we look at the census results this does not seem to be the case. Figure 9 shows the percentage of the two language groups (majority and minority) from 1880 to 1910 for both language contact situations. In Carinthia the percentage of Slovenian decreases from roughly 46% to about 30%, a decrease of 35% of the starting value. In southern Hungary on the other hand, the percentage of German decreases only from roughly 35% to about 32%, a decrease of 10% of the starting value. That means as a consequence for our model that the Magyarization is not directly visible in census results and therefore does not need to be taken into account as an extra parameter. The model can only (at best) imitate the results given by the census, and if these results do not show an influence of Magyarization, then the model does not need to include one. This is an important point concerning the limits of mathematical modeling: no matter what we *know* about the process, about what *should* be there, in the end we try to replicate the empirical data and are bound by them and any effects visible in them.

61 Croatia-Slavonia had been given special status within the Hungarian part of Austria-Hungary by the Croatian–Hungarian Settlement (*nagodba* in Croatian, *kiegyezés* in Hungarian) in 1868.

62 The Hungarian Nationalities Law only applied to the Kingdom of Hungary, but did not extend to the Kingdom of Croatia-Slavonia. In Croatia-Slavonia, Croatian was established as the official language for legislation, administration and judicature by the Croatian-Hungarian Compromise. Representatives of Croatia-Slavonia were also permitted to use Croatian in the Hungarian parliament.

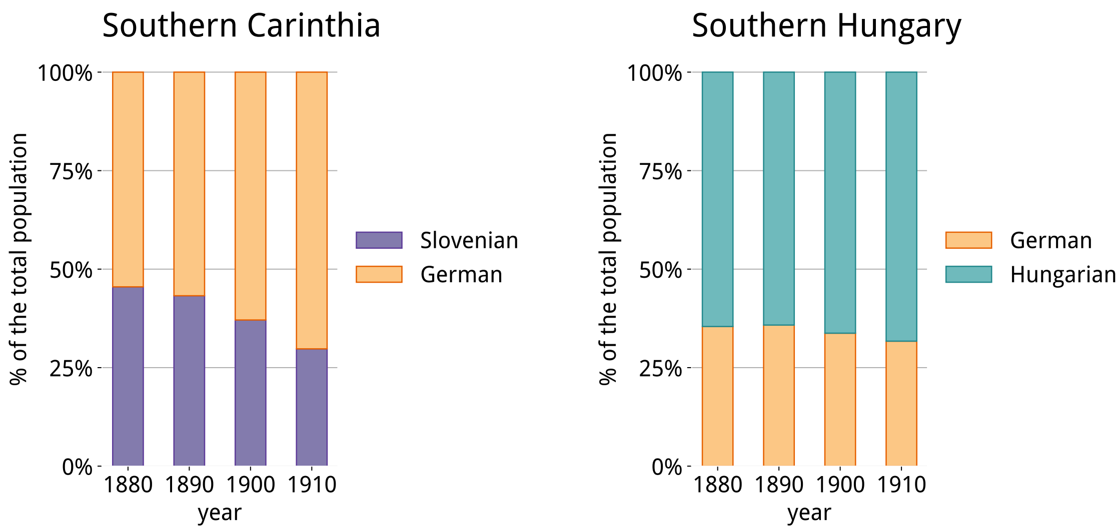


Figure 9: Percentage of the two language groups (majority and minority language) in the population from 1880 to 1910, calculated from census results. Only Slovenian and German, or German and Hungarian are taken into account.

It is surprising that the census results do not reflect the large pressure to adopt Hungarian in the Hungarian half. There is a possible explanation which again shows the importance of setting the data into context: The Austrian census asked for Germ. *Umgangssprache* ("everyday language", the language used most commonly in everyday life) whereas the Hungarian census asked for Hung. *anyanyelv* ("mother tongue"). It is likely that a person's mother tongue (or the statement about it) does not change, no matter what language the person uses later in life. Mother tongue is a characteristic fixed in the first few months and years of a person's life and therefore cannot be altered except by a conscious decision to "re-write history" by giving a different statement on the census. Everyday language on the other hand is more fluid and can change based on different circumstances. Even though we cannot know how much the people answering the questionnaires were influenced by the state's explanation of these terms (see section 6.1), at face value (and even considering the intended interpretation) the two categories represent different aspects of language use. We can therefore assume that the nature of the question does play some role in explaining the smaller decrease of German in Hungary compared to Slovenian in Carinthia.

6.3 Influence of settlement and population structure

After looking at the influence of different language politics, we will now turn to the settlement and population structure of Austria and Hungary to see (a) where the differences lie and (b) if these differences have consequences for modeling.

Table 3 shows a summary of data on geography and population for the two language contact situations. As noted before, “settlement” designates the smallest administrative population unit shown in the census, disregarding any further distinctions (whether it is administratively speaking a village or a hamlet, town privileges, city status, etc.). From this table, we see that in Carinthia, there are more settlements in a smaller area than in Hungary, meaning there is a difference in settlement structure. There is also a difference in settlement size. Figure 10 shows the distribution of settlement sizes for Carinthia and southern Hungary, i. e. how often settlements with 100, 200, etc. inhabitants occur. Here, settlement size is calculated only by taking the two most frequent languages (Slovenian/German or German/Hungarian). We can see that the distributions for the two areas have their maximum at different points: the maximum for Carinthia is at a smaller range than the one for southern Hungary. This means that in Carinthia, smaller settlements occur more frequently than in southern Hungary. As visible on the inset, in Carinthia, most settlements have 100 or fewer inhabitants, in Hungary, most settlements have 300–400 inhabitants.

	Slovenian–German in Carinthia	**German–Hungarian in southern Hungary**
number of settlements	$\approx$ 1500	$\approx$ 450
geographic area	$\approx$ 3100 km^2	$\approx$ 8100 km^2
language distribution in the beginning (1880)	$\approx$ 85,500 Slovenian + $\approx$ 102,300 German	$\approx$ 157,000 German + $\approx$ 285,000 Hungarian

Table 3: Summary of data on the two language contact situations investigated in Austria-Hungary. Numbers have been rounded.

The difference in settlement sizes in Carinthia and southern Hungary is reflected in geography and spatial distribution. Figure 11 shows two maps depicting areas of Carinthia and southern Hungary on the same scale. The maps show that Carinthia has rather small settlements which are closer together (sometimes only the settlement name is visible and not even the settlement area itself on this scale) while Hungary has larger settlements which cover a larger area and are further apart. What does this difference in geography mean for modeling? For modeling language shift or spread over time and *space* we need an abstraction of the real geographical space for a mathematical description. Let us assume that we want to create an abstract representation of the area, e. g. through a grid of square cells, where on average each settlement should be covered by only one cell. Then the consequence is that this grid should have a smaller cell size in Carinthia, where settlements are smaller, than in Hungary, where settlements are larger (see figure 11). In other words, different settlement structures enable different abstract representations.

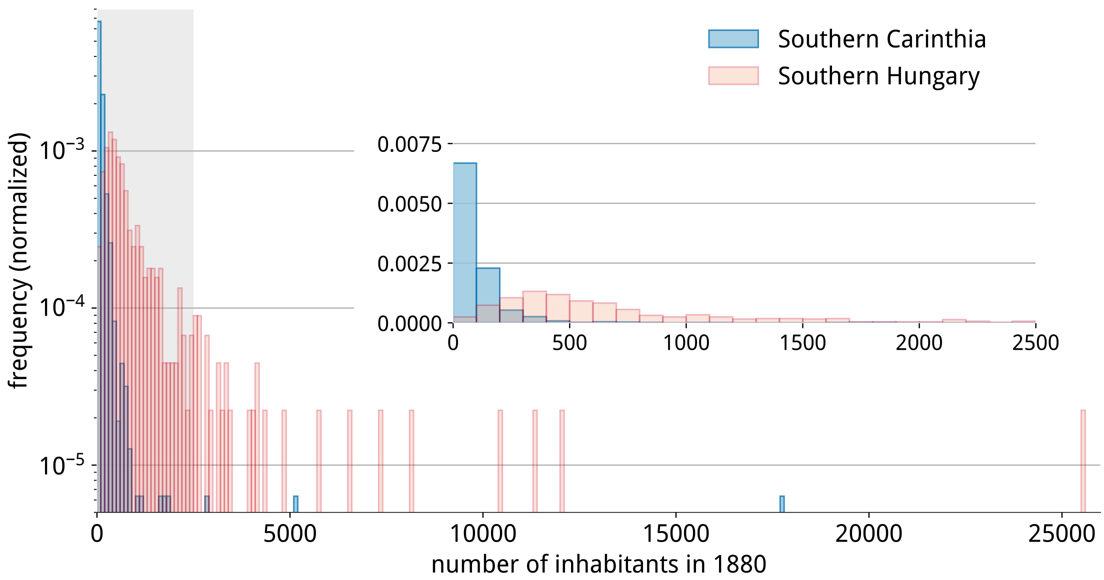


Figure 10: Distribution of settlement sizes (number of inhabitants) in Carinthia (blue) and southern Hungary (red) in 1880. Bin size of the histogram: 100 inhabitants. The inset shows the section marked in grey on the complete histogram. The y-axis on the main plot is logarithmic in order to make small frequencies visible.

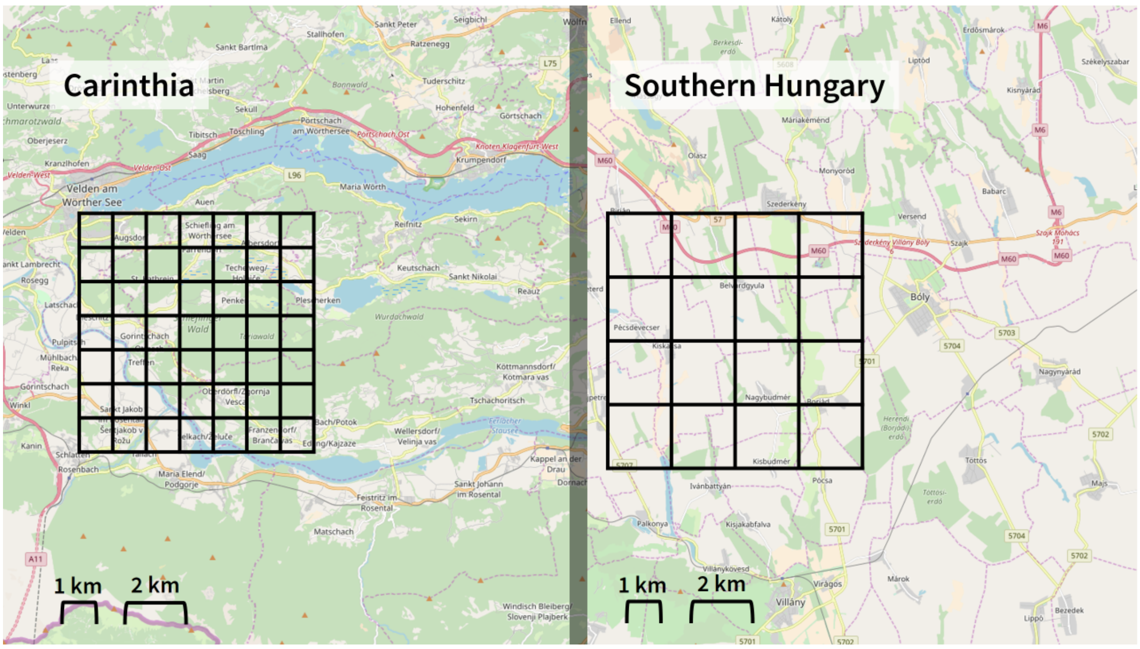


Figure 11: Maps of areas of Carinthia (left) and southern Hungary (right) on the same scale with a grid overlay, showing that different grid sizes are to be used for the different areas. Maps: OpenStreetMap.org / © OpenStreetMap contributors.

The differences in geographical structure and settlement distribution also have consequences for language contact. Figure 12 shows a graphical representation of these differences. In Carinthia, we have one cohesive Slovenian language area due to the proximity of the individual settlements. In Hungary, we have more distant “language islands” which are separated and not interconnected. The contrast becomes even starker when taking topography into account: For Carinthia, the white areas in the figure are mostly mountainous regions, and people tend to settle in valleys and flat areas. For Hungary, the whole area is mostly flat plains, meaning that the scattered settlement structure in this region is not necessarily due to geographic constraints. The different distributions do not mean settlements in Hungary are completely isolated from each other (as the term “island” might imply), but simply that it is less likely for them to directly influence each other than in Carinthia. As a consequence for our mathematical model, this means interaction probably should be incorporated as a parameter in a model for language shift in Carinthia. In Hungary, interaction should be tested as a model parameter as well, but seems less likely to have an effect based on the geographic structure as we see it here.

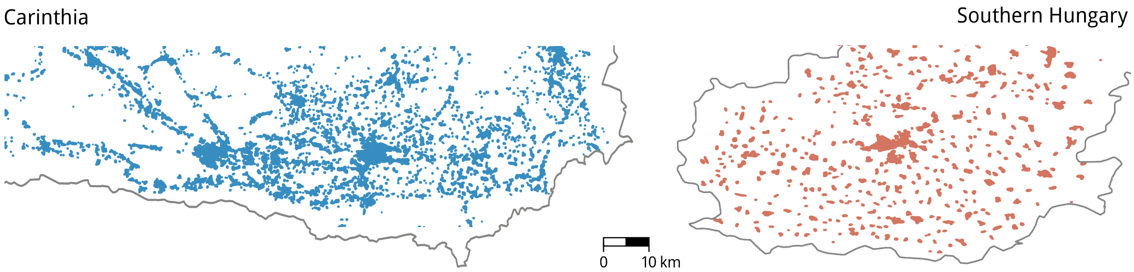


Figure 12: Schematic representation of the settlement structure in Carinthia and southern Hungary. Maps show all settlements regardless of linguistic make-up on the same scale. In Carinthia (left), settlement structure is more cohesive and the settlements are grouped closer together. In southern Hungary (right), the settlements are more distant and not interconnected. Geographic data from Kärnten Atlas (https://gis.ktn.gv.at) and OpenStreetMap.org / © OpenStreetMap contributors.

Apart from the general difference in geographical structure in both regions (which does not discriminate between languages, i. e. one language group does not settle only on mountains and the other only in valleys), there are also other more cultural differences in the population structure of the minority groups. Concerning religious affiliation, the Slovenian-speaking population in Carinthia is more homogeneous than the German-speaking population in southern Hungary: While almost all Carinthian Slovenes have been and are Catholic, there have always been both Catholic and Protestant (and Jewish) Germans in Hungary and interfaith marriages were unacceptable [137, p. 85]; [156].[63] In addition, the two minority groups in both countries are distinct in their linguistic varieties. There are several Slovenian dialects in Carinthia [14], but they are closely related and classified as one dialect group (the Carinthian dialect group of Slovenian). In contrast, the dialects of Germans in Hungary are very heterogeneous, not least owing to the history of German in Hungary: German-speaking people settling in Hungary from the 17th century on came from many areas of Germany and Austria. This led to a mixture of dialects from various German dialect groups, eventually evolving into today's dialects of German in Hungary which can still be distinguished by their original provenance [158, p. 262f]; [159].

As a final point, let us have a brief look at the linguistic structure of the whole population. We have noted (table 2 and figures 6 and 7) that the Austrian census publications show language-related data on German plus 1 (or rarely 2) language(s) and everything else is grouped into the category "other". The Hungarian census publications show data for many different language groups (plus a category "other" in one way or another). Neither census in this time period has results on bilingualism. In Carinthia, the "other" column is empty most of the time, meaning there are no languages represented other than Slovenian and German. In Hungary on the other hand, there are often considerable numbers of other languages, especially Croatian/Serbo-Croatian in Baranya county, the south of which is part of Croatia nowadays. This means that the situation in Hungary is a lot more complex because we have more than two language groups. A model therefore also has to be more complex if it is to take into account all language groups.

We can of course focus solely on the two groups of German and Hungarian (which we have done so far and will do in the following for reasons of simplicity), but the question remains: Where do we draw the limit on when to ignore "other" languages? Put differently: Can we ignore 29 people using Croatian, but not 30? I am afraid there is no clear answer to be found within this thesis (or anywhere else). Any mathematical model is necessarily a simplification of the "real world" and this is a tradeoff we make when creating an abstraction. Still, even a simplified model can be helpful to gain insight into a complex process such as language shift. To create such a model, we will now return to the two individual language contact situations.

63 Religion played an even larger role than language in marriage patterns—Catholic Germans would rather marry Catholic Hungarians than Protestant Germans [157, p. 323].

7 Situation I: Slovenian–German in Carinthia

The first language contact situation we consider for modeling is Slovenian–German in Carinthia, Austria. In this region, use of Slovenian is declining in favor of the majority language German and a language shift is taking place (and has been for quite a while). After giving a short overview of the historical background, I will outline the data preparation and finally describe the modeling process. Lastly, results will be discussed and influence factors on language shift will be identified.

7.1 Historical background

[*This section is largely a translation of section 4.1 in [4].*]

In the southern part of Carinthia, Austria, there has long existed a language contact situation between a Slovenian-speaking autochthonous[64] minority and the German-speaking population group [116; 160–162]. The first Slavic peoples settled in Carinthia in the 6th century during the Migration Period (*Völkerwanderung*). For a long time, the area was multilingual. It was only in the middle of the 19th century that the rise of nationalism led to the first major conflicts between the two population groups, which were to continue for a long time. The question of which language was spoken by whom and what that meant in conjunction with belonging to an ethnic group became more and more important. Ethnicity and language were often equated during this time, a practice which continued later on (section 4.4).

At the end of the 19th century, the first large-scale surveys on the distribution of Slovenian speakers began. The first language statistics for Carinthia were issued in 1846 by Karl Freiherr von Czörnig. Subsequently, until 2001, the census also included a language-related question in some form [30, p. 45ff]; [124]. However, these results on language use were used not only for linguistic investigations (as in the present case) but also as a political instrument.

Slovenian is a recognized minority language in today's Austria, i. e. there is a legal right to using Slovenian as official language for administrative processes, the right to Slovenian-language instruction in schools as well as a right to bilingual topographical signs (village signs with both the Slovenian and German name on it). These rights were established by the Austrian State

64 "Autochthonous" means "native to the land, indigenous". Autochthonous minorities (historical/indigenous minorities) are groups which have been living in an area for a very long time, usually for centuries, before today's states and state borders were established. The term is used in contrast to minority groups which have been "newly" created through recent immigration. Once again, the line between the two types is not straightforward to draw. The distinction can also become problematic as minority rights are usually restricted to autochthonous minorities.

Treaty (*Staatsvertrag*) in 1955 and regulated in more detail by the Ethnic Groups Act 1976 (*Volksgruppengesetz*).[65] For the most part, these rights are somehow linked to the presence of minority members—the rights are intended to benefit the Slovenian ethnic group in areas where they actually live. However, in Austria no-one can be forced to declare membership to an ethnic group, therefore results from the census on language use are often used as a criterion for the granting of rights. For example, in its original version (BGBl. 396/1976), the Ethnic Groups Act calls for a 25% threshold of ethnic group members in the population for the provision of bilingual topographic signs. To determine where the 25% threshold was met, census results on language were used.

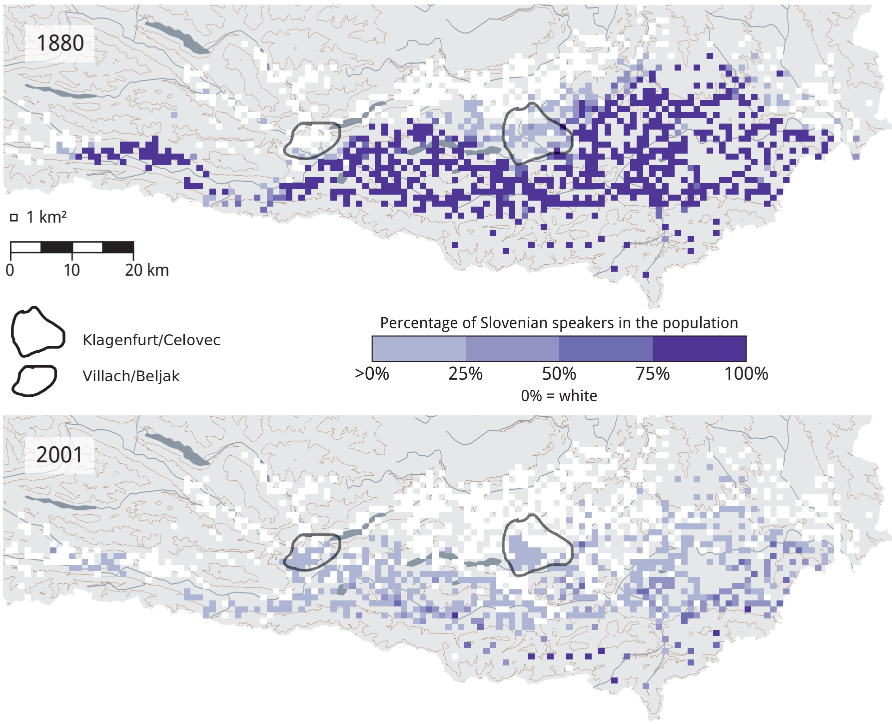


Figure 13: Percentage of "Slovenian" in census results in southern Carinthia, Austria. Top: census results for 1880, bottom: census results for 2001. The two biggest Carinthian towns, Villach/Beljak and the capital Klagenfurt/Celovec, are encircled according to present administration.

Figure 13 shows a picture familiar from the introduction of this thesis: the percentage of "Slovenian" as everyday language in southern Carinthia in 1880 compared to 2001, according to census results. If we compare the two figures, the area in which mainly Slovenian is used becomes

65 Minority language rights in Austria had actually been already formalized after World War I in the Treaty of Saint-Germain-en-Laye 1919, before the annexation of Austria 1938 (*Anschluss*) and its re-establishment by the Austrian State Treaty in 1955.

smaller over time. We can therefore assume that language shift, i. e. a change in language use in some form is happening, even if the census does not necessarily reflect actual language use (section 4.2). This notion that language shift is happening is supported by many smaller-scale qualitative studies of Slovenian in Carinthian villages e. g. [163–167]. With the help of census data, we can now examine this language shift quantitatively in the whole area of Southern Carinthia.

7.2 Preparing the data

[The first part of this section is largely a translation of section 4.2 in [4].]

As mentioned in section 4.6, census data provide—despite their limitations—the best possible starting point for building a quantitative model of language shift in Austria(-Hungary). The Slovenian ethnic group is located mostly in the south of Carinthia and this is where language shift is happening. For this reason, it makes sense to restrict ourselves geographically to that area. Therefore, only data from the districts of Völkermarkt, Klagenfurt/Klagenfurt-Land, Villach-Land and parts of Hermagor was used. Data from municipalities bordering the area in the North was included as well (see figure 14). For this region, the following two data sets were selected for modeling:

– 1880 to 1910 from the census of the Austrian half of Austria-Hungary
– 1971 to 2001 from the census of the Federal Republic of Austria

These periods were chosen because during them, the census consistently asked for *Umgangssprache* (defined as “the language used most commonly in everyday life”), i. e. the phrasing of the language-related question did not change. A certain degree of comparability is therefore given, while the phrasing differed during the years in-between (figure 3). For 1951 and 1961, where the census also asked for *Umgangssprache*, no data for individual settlements could be obtained; data are available only at the level of municipalities in the official publications. After 2001, no more data on language use is available through the Austrian census.

A large part of the data we are working with are historical data from the Austro-Hungarian Empire. It is generally not digitized other than possibly existing as scans. The first step in preparing these data for mathematical modeling was therefore digitization.[66] The second data set for 1971 until 2001 was already available in digital form (special data evaluation commissioned from Statistics Austria).[67] The Austro-Hungarian census forces a binary situation as only one language could be answered: according to the data, people in Carinthia use either German or

---

66 The digitized census data for southern Carinthia 1880–1910 have been made available online (https://doi.org/10.6084/m9.figshare.4535399).

67 Due to privacy protection laws, data on language use for individual settlements cannot be published for 1971 to 2001.

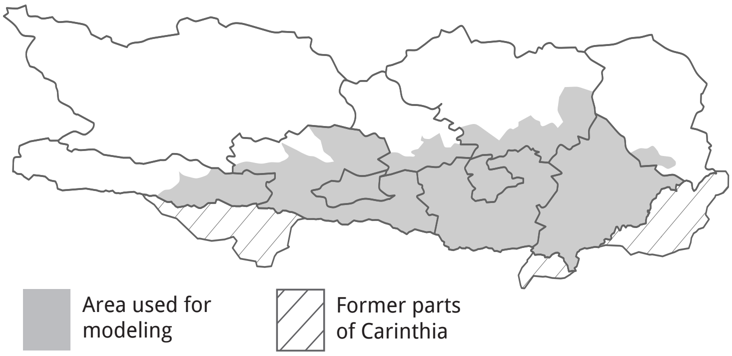


Figure 14: Map of Carinthia with its districts (today's borders). Grey shading roughly indicates the area used for modeling, based on today's borders of municipalities. Map data: [138; 139]

Slovenian as everyday language, but not both. Later census results from 1939 and afterward (and thus the second time period from 1971 on) allow for bilingualism as an answer category, e. g. "Slovenian, German" or "German, Slovenian". For the sake of consistency, these bilingual answers had to be converted to either German only or Slovenian only. To do so, the first language given in the answer was counted, i. e. "Slovenian, German" was reduced to "Slovenian" and "German, Slovenian" to "German". Similarly, statements including "Windisch" were counted as "Slovenian". "Windisch" is a politically-charged term for members of the Slovenian ethnic group in Carinthia who feel they belong to German(-speaking) culture and often do not want to claim minority rights [115, p. 42]; [120; 121], see also footnote 43 on page 47. From this, we once again see that the census data do not necessarily offer "objective" data on language use.

The data give us information on the number of speakers per settlement, in 10-year intervals. As we want to build a model for language shift over time and space (as in physical diffusion), we now need to connect these speaker data to their spatial position. To do so, we first create an abstract representation of Carinthia itself by subdividing the area into a regular grid with cells sized 1 km $\times$ 1 km (figure 15). Carinthia is roughly 84 km high (north border to south border, approx. 0.8° of latitude) and roughly 184 km wide (west border to east border, approx. 2.4° of latitude), the grid covering the total area of Carinthia therefore has $84 \times 184 = 15{,}456$ cells. As Carinthia is not a perfect rectangle shape, only 9,549 of the cells actually cover the area inside the Carinthian borders, the rest covers the area outside state borders which was not considered for modeling. We also do not consider periodic boundary conditions of any sort for simplicity, meaning that the number of people in the area is not conserved (as it should not be—after all, the number of people in the world is not conserved).

The data on language use (speaker numbers) from individual settlements are assigned to grid cells based on the geographic coordinates of the settlement (figure 16). Geographic coordinates (WGS 84) for the position of each settlement were manually retrieved from the KAGIS Kärnten

Figure 15: Creating an abstraction of the geographic area by overlaying a regular grid

Atlas (https://gis.ktn.gv.at). The capital of Carinthia, Klagenfurt/Celovec, was split into multiple cells for the second time period due to its size. Speaker numbers for Klagenfurt/Celovec were equally distributed to all cells.[68] If two or more settlements are covered by the same grid cell, the speaker numbers are added in that cell. In the end, we obtain a set of matrices for each year of the census, for both Slovenian and German (8 matrices per time period, 4 per language). These matrices have the speaker number per language and year as entries, with speaker numbers in positions where settlements are, i. e. they are a regular abstract representation of census data distribution in space to make mathematical modeling easier.

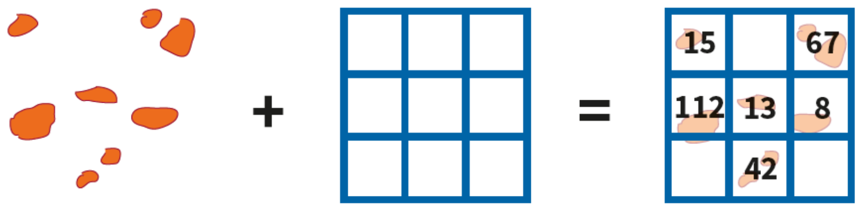


Figure 16: Assigning speaker numbers from individual settlements (orange) to grid cells

7.3 Macroscopic view: The language border

[*This section is based on [1].*]

Now that we have digitized data in a format we can do calculations with, it is time to set up the model and simulation. As mentioned in section 3.2, there are two ways of looking at physical diffusion, and similarly two ways of looking at linguistic diffusion (language shift in this case): macroscopically, and microscopically. In this section, we will first consider the macroscopic view.

68 As the census gives no information on the distribution of speakers within a settlement (whether they cluster together or are spread evenly), we must assume they are distributed equally over the whole settlement area.

In the macroscopic view of language shift, we can look at two things: (1) the "concentration" (proportion/fraction) of speakers of a language in the population and its change over time and space and (2) the area where a language is used, i. e. the border around the area where Slovenian is used and its change over time and space. These two aspects are actually intertwined: if people change their language and the fraction of speakers of a language decreases, then the area where a language is used becomes smaller and the border moves (not considering here a scenario in which "holes" develop in the language area and it dissolves from within).

Mathematically, the existence of such a moving border between two groups which differ in one characteristic follows from Fisher's equation [168], which is a reaction–diffusion equation.[69] Fisher's equation describes the spread of genes in biology where one gene is more advantageous than the other:

$$\frac{\partial c}{\partial t} = D \cdot \frac{\partial^2 c}{\partial x^2} + k \cdot c(1-c) \tag{8}$$

Here, c is the concentration/fraction of the advantageous gene in the population, D is its diffusivity (a measure for its spread) and k is the conversion rate from non-advantageous to advantageous. The fraction of the non-advantageous genes is given by $c' = 1 - c$.

A similar situation applies in Carinthia, where German has a higher status than Slovenian in many parts, i. e. it is more "advantageous" and people generally "convert" from Slovenian to German. We can therefore write the development of the fraction of German speakers as follows:

$$\frac{\partial c_{\mathrm{G}}}{\partial t} = D_{\mathrm{G}} \cdot \frac{\partial^2 c_{\mathrm{G}}}{\partial x^2} + k \cdot c_{\mathrm{G}}(1-c_{\mathrm{G}}) \tag{9}$$

Here, c_{G} is the concentration/fraction of German speakers in the population, D_{G} is the diffusivity of the German language and k is the conversion rate from Slovenian to German. The fraction of Slovenian speakers is given by $c_{\mathrm{S}} = 1 - c_{\mathrm{G}}$. The one-dimensional approach is sufficient in this case as the census data show that the language border movement is basically one-dimensional (from north to south).

Equation (8) (and thus also equation (9)) has solutions which describe the moving border (or traveling front) as the "wave" of the more advantageous, higher status group spreads. The velocity v of this moving front is given by:[70]

$$v = 2\sqrt{D \cdot k} \tag{10}$$

69 This equation is also known as the Kolmogorov–Petrovsky–Piskounov (KPP) equation.

70 A derivation of the wavefront velocity can be found e. g. in [48] or [168].

Linguistically, language borders, language boundaries or language fronts delineate the border between two areas where people use different languages.[71] However, language borders are often influenced by political boundaries. They are also hard to draw because in most cases, people in a country do not use the same language everywhere, even if it is subsumed under one umbrella term. For example, the German spoken in Lower Austria, Vienna and Burgenland (the easternmost states of Austria) is wildly different than the German spoken in Vorarlberg (the westernmost state). However, they are still accepted as the same language, namely German. Many societies are also multilingual and people are mobile, making it hard to draw clear borders on a map as there are no definitive boundary criteria [169, p. 134]. Any boundary between languages drawn on a map is therefore deceptive: it is unlikely that people abruptly stop using a language once we take one step over the border, or that everyone within the language area uses that language and those outside do not. Even the term “border” itself creates a false illusion; that of a clean cut rather than the transmission zone it represents in reality.

Coming back to our mathematical model which is supposed to follow the movement of a language border, this means we must first define a language border between Slovenian and German in some way. As this border should be based on language use rather than other factors like ecological factors (mountain ranges …), we turn to the speaker numbers provided by the census. From these results, it is evident that we do not have an “ideal” clearly delimited Slovenian language area where 100% of the population uses Slovenian on one side and 100% of the population uses German on the other. We have seen in figure 13 that there is an area where Slovenian is predominant (in the south), but there is no clear-cut border to the rest of Carinthia but rather a buffer zone along which language use changes gradually.

To draw a language border between Slovenian and German, we must therefore set a limit on when a cell counts as “Slovenian” and when it counts as “German” for the purpose of modeling. In other words, how many people in the population must use Slovenian for this cell to be within the Slovenian language area. Figures 17 to 19 show the results for different limits on the percentage of people using Slovenian: 50%, 25% and 17.5%.[72] From these figures, we can see that the language border cannot be drawn as a straight line which would be a gross oversimplification. It is also evident that with none of these limits it is possible to draw a continuous language border between German and Slovenian for the second time period (1971–

71 A similar concept would be that of isoglosses, which separate areas where one linguistic feature of the same language is different, e. g. the pronunciation of a phoneme or a syntactic construction. Bunches of isoglosses (many isoglosses in the same place) create dialect boundaries. Dialect boundaries separate dialects whereas language borders separate languages. In reality sets of isoglosses generally do not neatly line up, so dialect boundaries are a theoretical idealization of complex reality.

72 The latter two were chosen as the lower limit for the provision of bilingual topographic signs in the Ethnic Groups Act 1976 in the original (25%) and revised version (17.5%).

2001) because there are only few dispersed cells with a significant percentage of Slovenian speakers. As the traveling wave fronts resulting from Fisher's equation are continuous,[73] we cannot apply it in the second time period where there is no continuous language border which could move.

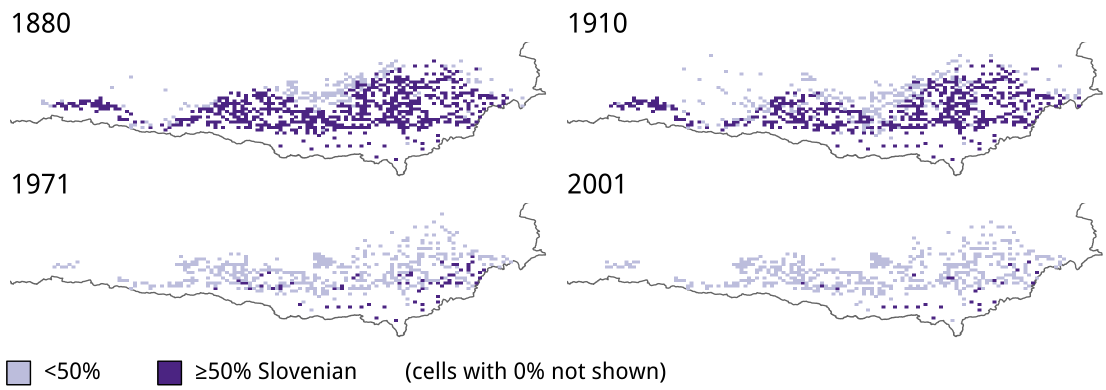


Figure 17: Slovenian language area in southern Carinthia for four different years (start and end of each modeling period). Cells with $\geq$ 50% Slovenian speakers are shown in dark purple, cells with $<$ 50% Slovenian speakers in light purple. Cells with 0% Slovenian are not shown.

For the first time period (1880–1910), it is possible to draw a language border and measure its velocity from the data. This velocity was calculated as follows: The language border was defined as the line bordering all cells with $\geq$ 50% Slovenian speakers, minus outliers not connected to the contiguous language area. The language border was then horizontally divided into n points (corresponding to the number of cells comprising the language border horizontally, i. e. one point per kilometer). For each of these points P_i, the vertical change in position (north-south difference due to movement) between the two time points t_1 (1880) and t_2 (1910) was calculated. The difference is divided by 30 (the number of year between t_1 and t_2 to obtain the average velocity per year:

$$v = \sum_{i=1}^{n} \frac{P_i(t_2) - P_i(t_1)}{30n} \tag{11}$$

73 It is also assumed that the concentration c is a continuous function, i. e. there are no sudden jumps. Concerning language shift, such jumps might occur in cases where e. g. all people speaking one language are killed or deported.

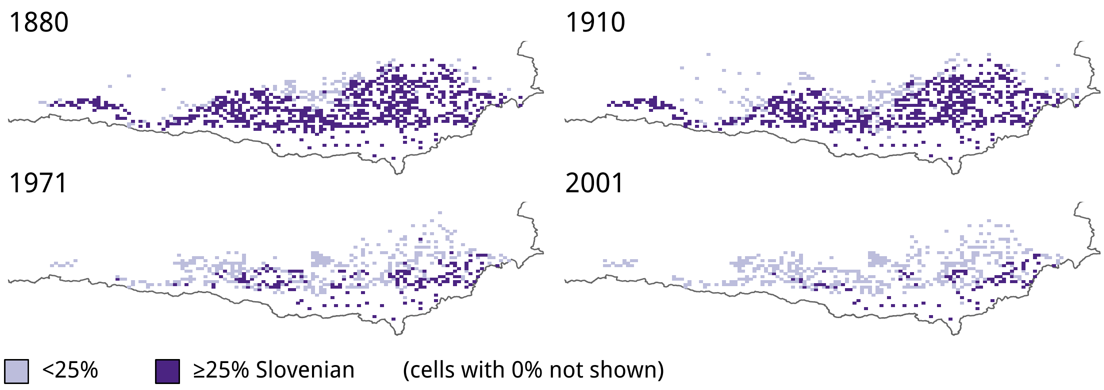


Figure 18: Slovenian language area in southern Carinthia for four different years (start and end of each modeling period). Cells with $\geq$ 25% Slovenian speakers are shown in dark purple, cells with $<$ 25% Slovenian speakers in light purple. Cells with 0% Slovenian are not shown.

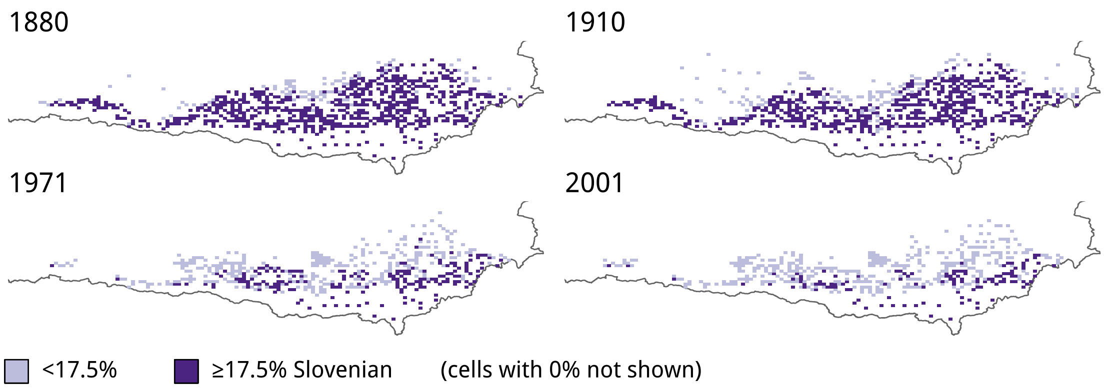


Figure 19: Slovenian language area in southern Carinthia for four different years (start and end of each modeling period). Cells with $\geq$ 17.5% Slovenian speakers are shown in dark purple, cells with $<$ 17.5% Slovenian speakers in light purple. Cells with 0% Slovenian are not shown.

For a limit of 50%, we obtain a velocity of $v = 0.034 \pm 0.017$ km/year. Using equation (10) we can also calculate the product of diffusivity and conversion rate $D_G \cdot k$, or if the conversion rate is known, the diffusivity of the German language. The conversion rate can be calculated from the census data by calculating the fraction of German speakers c_G at the beginning ($t_1 = 1880$) and end ($t_2 = 1910$) of the time period. If we neglect the diffusion term in equation (9) (pure growth of German speakers), then we obtain for k:

$$k = \frac{c_G(t_2) - c_G(t_1)}{\overline{c_G}(1 - \overline{c_G})/30} \tag{12}$$

where $\overline{c_G}$ is the average fraction of German speakers between 1880 and 1910. Using the census data for southern Carinthia[74] and assuming their error to be 10%, we calculate a conversion rate of $k = (0.0224 \pm 0.0065)$ year^{-1}. From k and the measured velocity v, we obtain a diffusivity of $D_G = (0.013 \pm 0.014)$ km^2 year^{-1}.

It is important to note that the velocity (and hence the conversion rate and diffusivity) given above is an average (a mean-field approximation). Due to this reason, we will see later in section 7.4 that the diffusivity calculated from the velocity of the language border is much smaller than the diffusivity obtained from a microscopic model—compare table 5 with the above value for D_G. (Conversely, the velocity calculated using the diffusivity of the microscopic model is considerably higher than the velocity obtained from census data: $v = 0.1101 \pm 0.0034$ km/year compared to $v = 0.034 \pm 0.017$ km/year.) However, this averaging masks important differences along the movement of the language border.

Figure 20 shows the language border for both 1880 and 1910, so the movement along the border can be examined. It is apparent that there are parts along the border where the distance moved is large (black arrows) while in other parts, there is no movement at all (green arrows). The two areas without border movement indicated in figure 20 are rural areas behind mountain ranges, i. e. next to unpopulated areas, so there is no German language area on the other side. This absence of "linguistic pressure" from the other side might explain the lack of border movement. In any case, it is another reason why border velocity is not very insightful in this case: equation (8) and equation (10) presume the existence of a moving border everywhere (a traveling front which does not change its shape while moving) and can thus not be employed to describe the "shrinking" of the Slovenian-language area. We will therefore take a different approach, which is described in the next section: a microscopic view which is much more detailed, making use of the grid cells previously mentioned.

74 1880: 102,314 German and 85,369 Slovenian speakers; 1910: 154,361 German and 65,352 Slovenian speakers

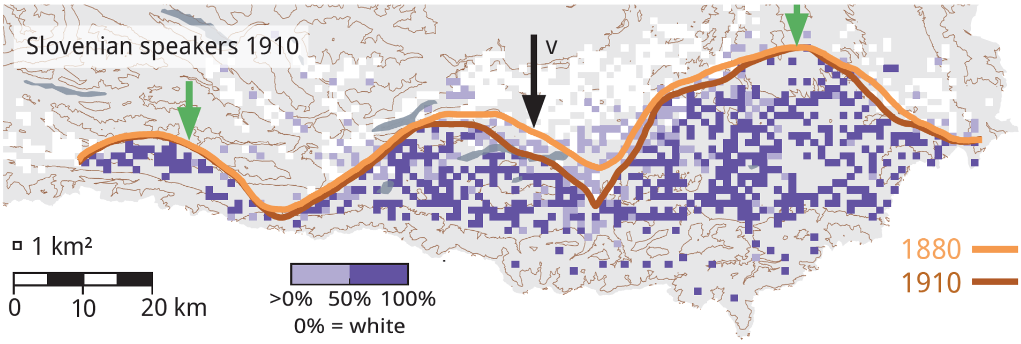


Figure 20: Percentage of Slovenian speakers in 1910 and schematic of the language border for 1880 (orange line) and 1910 (brown line). The border is defined as the line framing all cells with $\geq 50\%$ Slovenian speakers and moves southwards in one direction (black arrow). Green arrows indicate areas without border movement. Reprinted from [1].

7.4 Microscopic view: Grid cells and cellular automata

We have seen that the "macroscale view" of tracking the movement of the language border has its limitations because it essentially averages over the whole area. A microscale view and model can give us more insight into the local processes happening in different areas by calculating each point individually. In doing so, we can follow the development over time and space globally while being able to identify deviating processes locally, such as the different dynamics in urban centers compared to rural areas. A microscale model means we can also use more detailed data—if such detailed data are available, meaning sufficiently high-resolution data are a prerequisite for microscopic models.

As described in section 3.1, the general principle in (mathematical) modeling is to first establish hypotheses (rules) to describe the change for each data point. These rules are then applied to all data points and the result is compared to the empirical data. If the goodness of fit between model results and empirical data is not satisfactory, then the rules are improved and the process is repeated until suitable agreement between the model results and the data has been achieved.

In a microscale model, we calculate the change in each data point individually. Calculating each point individually in our case means turning back to the grid cells and calculating what happens inside each grid cell. In the terminology of cellular automata (see section 3.2.2): Each cell has an initial state and its state will be updated according to a fixed mathematical rule. By applying the rule to all cells simultaneously and equally, a new generation of cells is created after one time-step, with new data/states. The process is then repeated.

We apply this procedure here as follows: the cells in our matrix (the grid cells) include the speakers numbers, hence we need to create rules for how the speaker number of a language changes over time and then apply these rules to all cells. In other words, we have a certain number of speakers now and we want to predict by some rule how many speakers we will have per cell after one time-step. To create hypotheses on what these rules look like before formulating them mathematically,[75] two things are important:

– *What makes sense?*
  Any hypothesis should be a reasonable assumption based on previous studies.
– *What can we measure?*
  We want to eventually compare the model to empirical data, so ideally the parameters should at least be correlated to something which can be measured and for which data are available.

In addition to these, the model should be kept as simple as possible at first rather than starting with a hypothesis which already includes 50 different variables. With that many parameters, it becomes difficult to say which one plays a substantial role.

Keeping this in mind, the next step after mapping the data (section 7.2) is to create hypotheses on how the number of speakers changes per cell and year. Although census data are available only every ten years, we divide the periods in-between into yearly steps to simulate the language distribution for each year. This lets us interpolate missing data—an advantage of mathematical models—to follow the gradual nature of language shift.

We create our hypothesis based on factors influencing language use and language shift mentioned in the literature (section 2.1). As a first basic hypothesis, we assume that the number of speakers in the next year (after one time-step) is dependent only on two factors:

(1) the current number of speakers
(2) the (potential) interaction with other speakers of the same language in the surrounding cells[76]

A schematic of the application of this hypothesis to cells is shown in figure 21. We assume that the hypothesis we have created applies to all cells simultaneously and equally. The hypothesis can now be written as a formula by combining two underlying concepts: the probability of using a language in a given cell, and the calculation of speaker numbers from that probability.

Following our hypothesis, we assume the probability $p_\alpha(\mathbf{r},t+1)$ of using a language α (with $\alpha = \mathrm{S}$ or G, Slovenian or German) after one year (one time step) is directly proportional to:

(1) the absolute number of speakers of that language in the previous year $n_\alpha(\mathbf{r},t)$
(2) the (potential) interaction with other speakers of the same language in the neighborhood $F_\alpha(\mathbf{r},t)$

---

75 Technically speaking, hypotheses *become* rules by formulating them mathematically.

76 There is no reason to assume people will automatically speak to each other solely based on the fact that they use the same language, hence the term “potential interaction”. However, even if interaction is not realized, the mere possibility of using the language influences language shift.

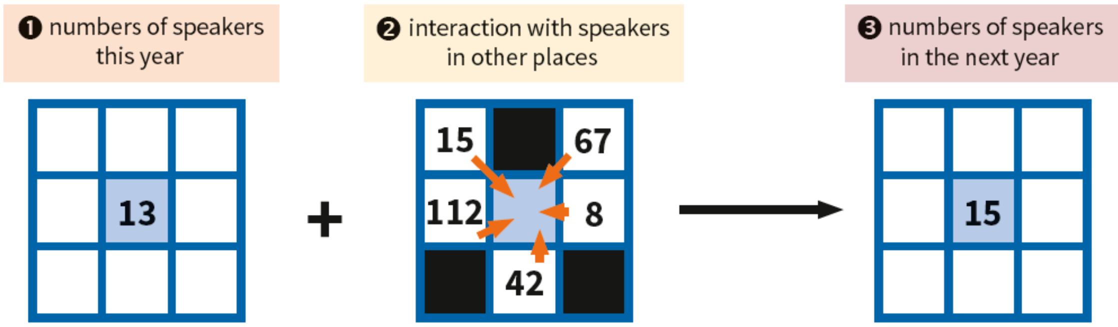


Figure 21: Schematic representation of how speaker numbers in the following year are calculated. Each cell represents an area of 1 km $\times$ 1 km, the number inside gives the number of speakers of a language in that area. In this case, we assume that the number of speakers in the next year depends on two factors: the current number of speakers plus interaction with other speakers of the same language in all surrounding cells (not only the nearest neighbors as depicted here).

The probability $p_\alpha(\mathbf{r},t+1)$ in a cell at position $\mathbf{r}$ after one time-step (one year) is then given by:

$$p_\alpha(\mathbf{r},t+1) = \frac{n_\alpha(\mathbf{r},t) + F_\alpha(\mathbf{r},t)}{n_\mathrm{S}(\mathbf{r},t) + F_\mathrm{S}(\mathbf{r},t) + n_\mathrm{G}(\mathbf{r},t) + F_\mathrm{G}(\mathbf{r},t)} \tag{13}$$

This probability is normalized to the total number of speakers and the total interaction in a cell, i. e. it is given as as the ratio of a language-specific influence per cell compared to the total influence on language use (for both languages) in that cell.

To compute the (potential) interaction F_α between speakers of the same language α with the neighboring cells, we sum up the contribution (interaction) of all cells surrounding a cell at position $\mathbf{r}$ (as shown in figure 21, but including all matrix cells, not only the nearest neighbors):

$$F_\alpha(\mathbf{r},t) := F_\alpha(\mathbf{r},n_\alpha,t) = \sum_{\mathbf{r}_j \neq \mathbf{r}} c_\alpha(\mathbf{r},\mathbf{r}_j,n_\alpha,t) \cdot \mathbf{dA} \tag{14}$$

The element $\mathbf{dA}$ is equal to the size of one cell, in this case 1 km $\times$ 1 km. The contribution (interaction) arising from each individual cell is modeled as a Gaussian function:

$$c_\alpha(\mathbf{r},\mathbf{r}_j,n_\alpha,t) = \frac{n_\alpha(\mathbf{r}_j,t)}{4\pi D_\alpha \Delta t} \cdot \exp\left(-\frac{|\mathbf{r}-\mathbf{r}_j|^2}{4D_\alpha \Delta t}\right) \tag{15}$$

where D_α is the diffusion constant, a measure for the diffusivity and the sphere of influence of a language, i. e. how far and fast a language "diffuses" from its outset. Δt is set to 1 year as the interaction is calculated in one-year time steps from the result of the year before.

Modeling interaction in this way is motivated by both linguistics and physics. In linguistics, following Trudgill [34] interaction between settlements (or in this case, two cells) is proportional to the number of speakers in the settlements n_a and decreases with the distance $|\mathbf{r} - \mathbf{r}_j|$ between the settlements (compare equation (2)). In physics and chemistry, Gaussian distributions are used to describe the diffusion of particles, i. e. they are a solution to the diffusion equation equation (4) under certain boundary conditions. The form of equation (15) is based on the solution of the diffusion equation in two dimensions, for a single non-continuous source (i. e. particles are released once at one point in space but not continuously or at multiple points)[77]:

$$c(\mathbf{r},t) = \frac{N}{4\pi Dt} \cdot \exp\left(-\frac{\mathbf{r}^2}{4Dt}\right) \tag{16}$$

77 A complete derivation with the exact mathematical boundary conditions (e. g. concentration vanishing far away from the source) and an overview of other possible solutions can be found e. g. in [170–173]. This solution was chosen because in our cellular automaton, we only have one source per cell (as we are looking at each language separately), the cell is treated as a point in space, and we are only looking at one timestep before starting with new initial conditions (new initial distributions of speakers) so there is no continuous release of the same language.

It would of course be possible to use solutions for other boundary conditions, e. g. ones where the total number of particles is not constant (continuous source) or where release of the substance occurs over a finite area rather than at one point, and arguments can be made for either. However, in these cases solutions to the diffusion equation are given by the error function $\mathsf{erf}(x) = \frac{1}{\sqrt{\pi}} \int_{-x}^{x} e^{-t^2} \mathsf{d}t$ whose value can only be approximated numerically. It does not appear meaningful to choose a non-analytical function for modeling language shift with the simple model presented here, especially when there are arguments for using an analytical solution (the Gaussian distribution) and the analytical solution provides a good fit to the empirical data, as we will see later. Similarly, one might expand the Gaussian distribution by introducing further parameters (e. g. different parameters for the width of the Gaussian and the height/amplitude rather than only one D_a), but again we strive for the simplest model possible which can describe our data reasonably well and can be motivated by existing research.

One could also ask if diffusion is even the right theory/concept to be applied here because in physical diffusion, the total number of particles is a conserved quantity and the total number of speakers of a language obviously is not. Speaker numbers can change through birth and death, but there is also no natural limit to how many people can learn a certain language—and people can learn more than one language. (This is also part of the reason why reaction–diffusion equations are popular for modeling language shift, as they generally do not assume a conserved quantity of speakers but rather that speakers of one language transform into speakers of another language, as in chemical reactions.)

We are not dealing with a closed system here (there are no periodic boundary conditions—speakers/languages which disappear to the right side do not enter the system again on the left side) so we are not modeling what happens to the speakers who leave the system. However, we are looking at discrete time steps and calculating the probability of using a language iteratively, with completely new initial conditions after each time step. This means that within one time step, the amount of language (expressed by proxy through the speaker numbers n_a) which spreads is conserved. This can be seen by turning equation (15) into a continuous function (by replacing Δt with t and integrating over space (from $-\mathsf{inf}$ to $+\mathsf{inf}$ in both directions). The integral yields the number of speakers $n_a(\mathbf{r}_j,t)$ for all t, i. e. the amount of language which spreads is equal to the number of speakers at the beginning of the time step, not more and not less (as it should be, and as it is in physics—the integral of equation (16) over space yields the number of particles N released in the beginning, no matter how much time has elapsed). Therefore there is no conflict with applying diffusion theory to language shift in the way done here.

where $c(\mathbf{r},t)$ is the concentration of a substance at position $\mathbf{r}$ (with $\mathbf{r}^2 = x^2 + y^2$ for two-dimensional Cartesian coordinates x,y) and time t. N is the total number of particles released initially at $t = 0$ and $\mathbf{r} = 0$, which then spread, and D is the diffusion constant which determines how fast the substance spreads. We assume here that D is the same in both x and y direction, i. e. $D_x = D_y = D$.

The model we have introduced thus directly incorporates inspiration from both disciplines: through the mathematical formulation and the inclusion of factors on language shift identified through previous studies and their dependencies.

The number of speakers of a language in a cell is calculated by multiplying the total number of people in the cell by the probability of using a language:

$$n_\alpha(\mathbf{r},t) = n_{\text{total}}(\mathbf{r},t) \cdot p_\alpha(\mathbf{r},t) \tag{17}$$

The total number of people is not predicted by this model, it is an external input (census data). As the census is only available every ten years, $n_{\text{total}}(\mathbf{r},t)$ is given by linear interpolation between censuses for a first simple approximation.

Equation (13) and equation (17) can be combined and we obtain:

$$n_\alpha(\mathbf{r},t+1) = n_{\text{total}}(\mathbf{r},t+1) \cdot \frac{n_\alpha(\mathbf{r},t) + F_\alpha(\mathbf{r},t)}{n_{\text{S}}(\mathbf{r},t) + F_{\text{S}}(\mathbf{r},t) + n_{\text{G}}(\mathbf{r},t) + F_{\text{G}}(\mathbf{r},t)} \tag{18}$$

This rule is iteratively applied to all cells to calculate new speaker numbers for any number of years, from the previous year's results. The rule we have created here is, of course, only a simple first model which will be expanded later (section 7.4.2). We will now check the validity of the underlying hypothesis (that language shift depends on the two factors included in the model) and the corresponding mathematical formulation by turning the model into a computer simulation.

7.4.1 Judging the model: Simulation and evaluation procedure

[*This section, apart from the first paragraph, is a direct quotation of the sections "Evaluation procedure" and "Evaluating model performance" (Supporting Information) in [1].*]

After creating a model, we now want to implement this model as a computer simulation and more importantly, judge its performance compared to the empirical data. Ideally, the model recreates the data perfectly and gives us the same results as the data. This is rarely the case, so we need some benchmarks to see just how well the model reproduces the given data set. We are also interested in whether the model performs better than a so-called baseline model. A baseline model might be a very simple model, or one which essentially assumes that everything stays as it is. We hope that the model we have created performs better than the baseline model (i. e. is able to reproduce the data more accurately), but again we need some value on which the two models can be compared. This section will therefore describe the simulation and evaluation procedure. Simulations were performed using GNU Octave 4.0.0. The data from the first census in each period (1880 and 1971) were set as the initial state from which the number of speakers in each cell changes according to equation (18), assuming a linear population development between censuses. To evaluate the goodness of fit between simulated data and census data, a great number of metrics are available which each have advantages and disadvantages. It is therefore helpful to use multiple metrics, especially in cases of large data scattering such as with census data on language use. We use ordinary least squares to minimize the squared sum of errors (SSE):

$$\mathrm{SSE} = \sum_{k=2}^{m} \sum_{i=1}^{n} (O_{i,k} - E_{i,k})^2 \tag{19}$$

where $O_{i,k}$ is an observed data point (census data) and $E_{i,k}$ is an estimated data point (simulated data). k is the number of times the observed data can be compared to the estimated data. k runs from 2 to m, where m is the number of censuses within the period. The data from the first census in each period ($k = 1$) are excluded as they are equivalent to the initial state of the system, hence there is no error for the initial state and we sum only over the remaining censuses. Optimization was done using the Nelder-Mead method [174]. Additionally, we used least absolute errors where the sum of absolute errors (SAE) is minimized to check the reliability of the fit:

$$\mathrm{SAE} = \sum_{k=2}^{m} \sum_{i=1}^{n} |O_{i,k} - E_{i,k}| \tag{20}$$

To evaluate model performance, a baseline for comparison is helpful. As a baseline we use an interaction free model ($F_\alpha = 0$) which means that the fraction of speakers of either language remains constant, speakers being lost or gained only through changing population size. To check if our model is better than the baseline, we use three metrics:

(1) the total number of Slovenian speakers in the last year of each period as calculated by the model, which should be close to the real number.

(2) root mean square error which is related to OLS (equation (19)). The root mean square error (RMSE) gives the mean error per cell per 30 years in speakers, which should be low:

$$\mathrm{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (O_i - E_i)^2} \tag{21}$$

where O_i is an observed data point (census data), E_i is an estimated data point (simulated data) and n is the number of populated cells.

(3) mean absolute error which is related to LAE (equation (20)). The mean absolute error (MAE) is the sum of absolute errors divided by the number of cells n, which should also be low:

$$\text{MAE} = \frac{1}{n}\sum_{i=1}^{n}|O_i - E_i| \tag{22}$$

Both errors are calculated from the result of the simulation and the census data at the end of each 30-year period. Results are given in table 4, indicating that the model with interaction (and optionally with habitat parameters) consistently leads to a better fit than the baseline. Note that RMSE and MAE average over all cells. A more detailed look into the model's error per category/number of speakers in a cell is given below.

Reliability of the model per category

Figure 22 shows two measures of the model's reliability: the mean absolute error (MAE, equation (22)) per category and grid cell, and the relative error per category (per initial number of speakers). To gain insight into where the model works best, we show the error per category to differentiate between cells with different numbers of Slovenian speakers. Both errors are given per 30 years, i. e. the error in the result of the simulation after 30 years compared to the census data at the end of each 30-year period.

The relative error is given by the sum of absolute errors divided by the sum of the number of Slovenian speakers in this category

$$\text{relative error} = \sum_{i=1}^{n}|O_i - E_i| \Big/ \sum_{i=1}^{n} S_i \tag{23}$$

where S_i is the number of Slovenian speakers per grid cell, summed over the n grid cells in this category.

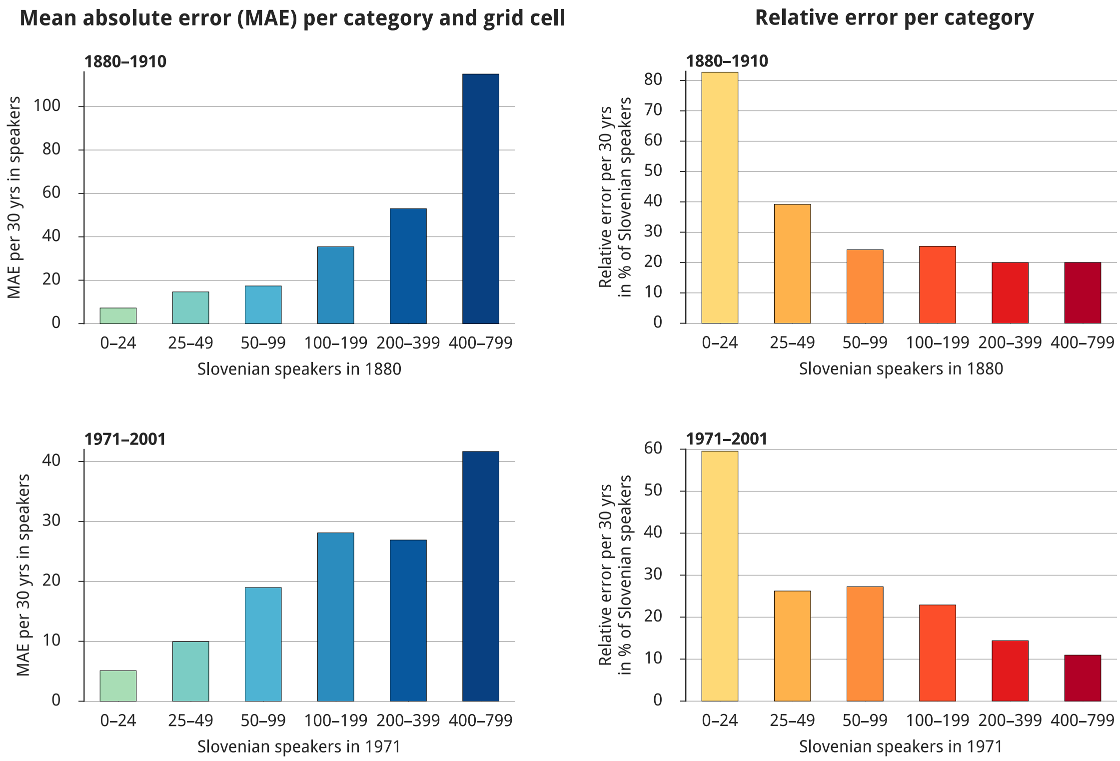


Figure 22: Measures of the model's reliability compared with census data: mean absolute error per category in speakers (left) and relative error per category in percentage of speakers (right). Both errors shown are per 30 years. Categories are set by the number of Slovenian speakers in the initial state of the system (census data from 1880 and 1971, respectively). The simulated data for the period from 1880 and 1910 is calculated including the habitat parameter for bilingual schools; the data for the period from 1971 to 2001 are calculated including the habitat parameter for urban areas. Reprinted from [1].

	Period					
	1880–1910			1971–2001		
Model	Total no. of Slovenian speakers in 1910	RMSE per 30 years (speakers)	MAE per 30 years (speakers)	Total no. of Slovenian speakers in 2001	RMSE per 30 years (speakers)	MAE per 30 years (speakers)
Baseline model	85,233	52.41	20.32	16,336	17.86	9.35
Interaction model	67,727	44.11	18.41	11,260	15.06	8.01
Interaction model with habitat	64,092	41.75	18.41	12,052	12.94	6.93
Census data	65,352	—	—	12,056	—	—

Table 4: Comparison values for the goodness of fit of the baseline model (constant fraction of speakers of either language), the interaction model (equation (13), and the interaction model with habitat (equation (25)), for Slovenian–German language shift in Carinthia. Three metrics are shown: the total number of Slovenian speakers (closer to census data is better), the root-mean-square error (RMSE) (equation (21); lower is better) and the mean absolute error per cell (MAE) (equation (22); lower is better). All results given for best fits (table 5). The model with habitat includes the bilingual schools habitat parameter for the period 1880–1910 and the urban habitat parameter for the period 1971–2001.

Total number of Slovenian speakers

Figure 23 shows the total number of Slovenian speakers according to all eight censuses in both periods (1880, 1890, 1900, 1910 and 1971, 1981, 1991, 2001) in comparison with the simulated data. Figure 24 shows the fraction of Slovenian speakers for the same periods, again comparing simulated data with census data. The agreement is satisfactory.

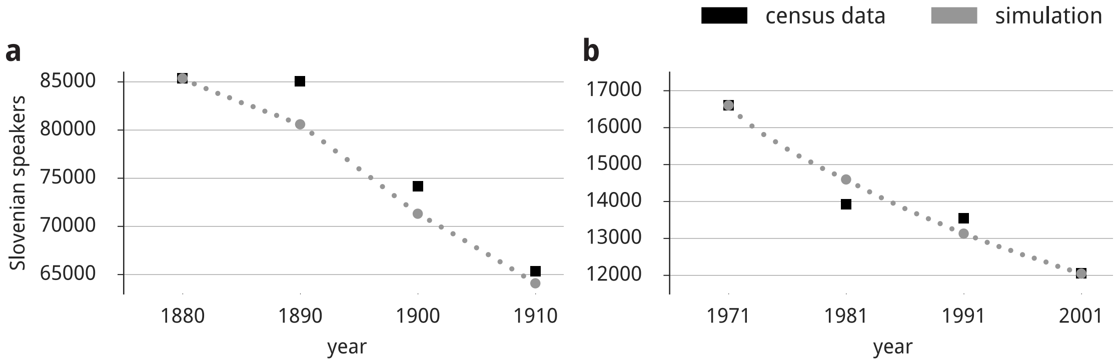


Figure 23: Total number of Slovenian speakers in southern Carinthia as estimated by the simulation (gray dots) and according to census data (black squares). (a) Period from 1880 to 1910. (b) Period from 1971 to 2001. The simulated data for the period from 1880 and 1910 are calculated including the habitat parameter for bilingual schools; the data for the period from 1971 to 2001 are calculated including the habitat parameter for urban areas. Reprinted from [1].

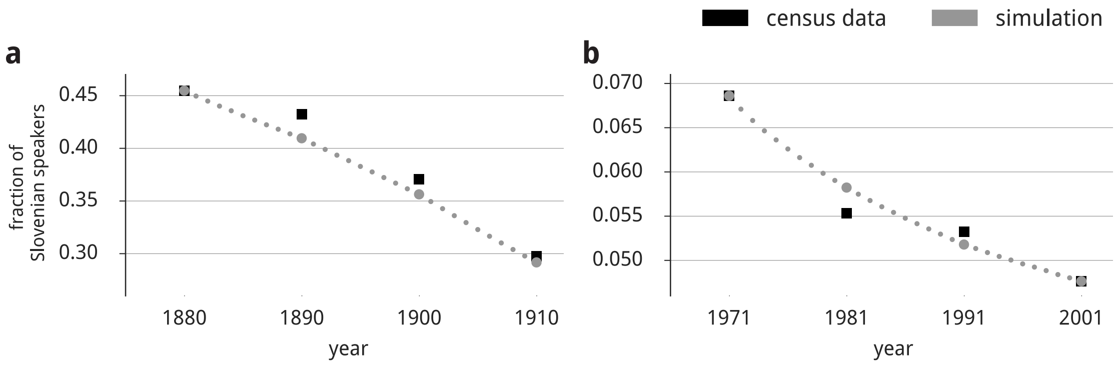


Figure 24: Fraction of Slovenian speakers in southern Carinthia as estimated by the simulation (gray dots) and according to census data (black squares). (a) Period from 1880 to 1910. (b) Period from 1971 to 2001. The simulated data for the period from 1880 and 1910 are calculated including the habitat parameter for bilingual schools; the data for the period from 1971 to 2001 are calculated including the habitat parameter for urban areas. Data for total population used in the calculation of the fraction are given by linear interpolation of census data.

Deviation of simulated data from census data over space

Figure 25 shows the residuals for the two periods (census data minus simulated data). Evident deviations in period 1 find their explanation in extraordinary outliers in the census data: some settlements switched from a strong German-speaking majority to a strong Slovenian-speaking majority. The same also happened in the opposite direction. Both of these developments are very different from the average trend in southern Carinthia which was a moderate transformation from Slovenian speaking to German speaking. In addition, several settlements "flip-flopped" from one census to the next, changing from a Slovenian-speaking majority to a German-speaking majority and then back to a Slovenian-speaking majority and back again to a German-speaking majority in the last census of the period [112, p. 238f]. This behavior, which seemed to be influenced by (local) politics rather than actual language use changes, cannot be captured by our model. The residuals thus show where language spread and retreat deviates from "average" development and open up possibilities for further research (far beyond the scope of this thesis, meaning that answers unfortunately cannot be given here): What were the reasons for these deviations? Can these reasons—which might be identified only by other research approaches (sociological, historical, …)—be integrated into the model as a habitat factor?

7.4.2 Results

[*Apart from the introductory paragraph, this section is a direct quotation of the results section in [1] and the corresponding sections from the Supporting Information of [1].*]

In the previous section we have established a way to evaluate the performance of our model and shown that it performs better than the baseline, i. e. the model is appropriate for describing language shift in Carinthia. We have also identified the regions where the model does not appropriately reproduce the data, thereby indirectly giving a preview of the results. These results are outlined in full in the following: simulation of language shift in Carinthia for 1880 to 1910 (period 1) and 1971 to 2001 (period 2). We start with the results for a basic model containing only two influence factors (number of speakers + interaction) which is then extended through the introduction of so-called "habitat conditions". "Habitat conditions" are, similar to the term of habitat in biology and in the spirit of ecolinguistics, environmental conditions which influence language use. "Environmental" is here to be taken loosely and does not necessarily refer to the natural environment (mountains, rivers, …) but rather to things like schools using a language. The extended version of the model allows us to capture the influence of these additional factors on language use and hence language shift.

Language shift in the period 1880–1910

The widths of both Gaussians, and hence the diffusivities for German and Slovenian, are fitted to the number of speakers in each cell as given by the census data. Fits to the census data were

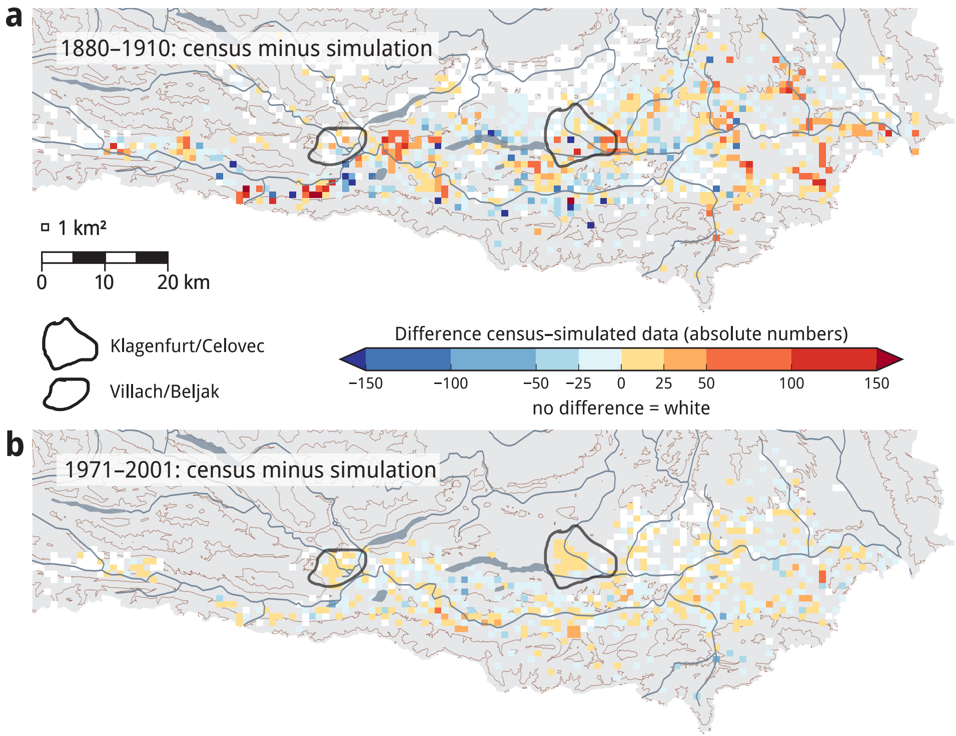


Figure 25: Residuals for the number of Slovenian speakers: census data for the last year of each period minus simulated result at the end of each period. (a) Period from 1880 to 1910. (b) Period from 1971 to 2001. Shades of blue: model produces too many Slovenian speakers, shades of red: model produces too few Slovenian speakers. The two biggest Carinthian towns, Villach/Beljak and the capital Klagenfurt/Celovec, are encircled according to present administration. As in figure 23, the simulated data for the period from 1880 and 1910 are calculated including the habitat parameter for bilingual schools; the data for the period from 1971 to 2001 are calculated including the habitat parameter for urban areas. Reprinted from [1].

performed for the period from 1880 to 1910. The best solution was achieved with the values given in table 5.

Figure 26 shows the increase (red) and decrease (blue) of the number of Slovenian speakers in southern Carinthia for census data and simulated data. We obtain satisfactory agreement between the empirical data and the predicted data on a microscopic scale. This can be seen in detail in figure 22 where the model's errors are shown for cells with different numbers of Slovenian speakers. Figure 25 shows the residuals (census data minus simulated data). The total number of Slovenian speakers as predicted by the simulation also agrees with the census data (see figure 23). Thus our model is able to follow how either language has spread and retreated in the time period 1880 to 1910.

Diffusivity	Period 1880–1910	1971–2001
D_G (km^2 year^{-1})	0.1356	0.1078
D_S (km^2 year^{-1})	0.0939	0.0926

Table 5: Fit values for the diffusivities D_α for both time periods. Changing both diffusivities in the same direction and by the same amount changes little in the quality of the fit, whereas even a small change to the ratio of diffusivities has a big impact on fit quality. A 10% change of both diffusivities in the same direction leads to a 1% change in the sum of absolute errors. On the other hand, if only one diffusivity is changed by 10% (i. e., the ratio between diffusivities changes as well), this leads to a 10% change in the sum of absolute errors.

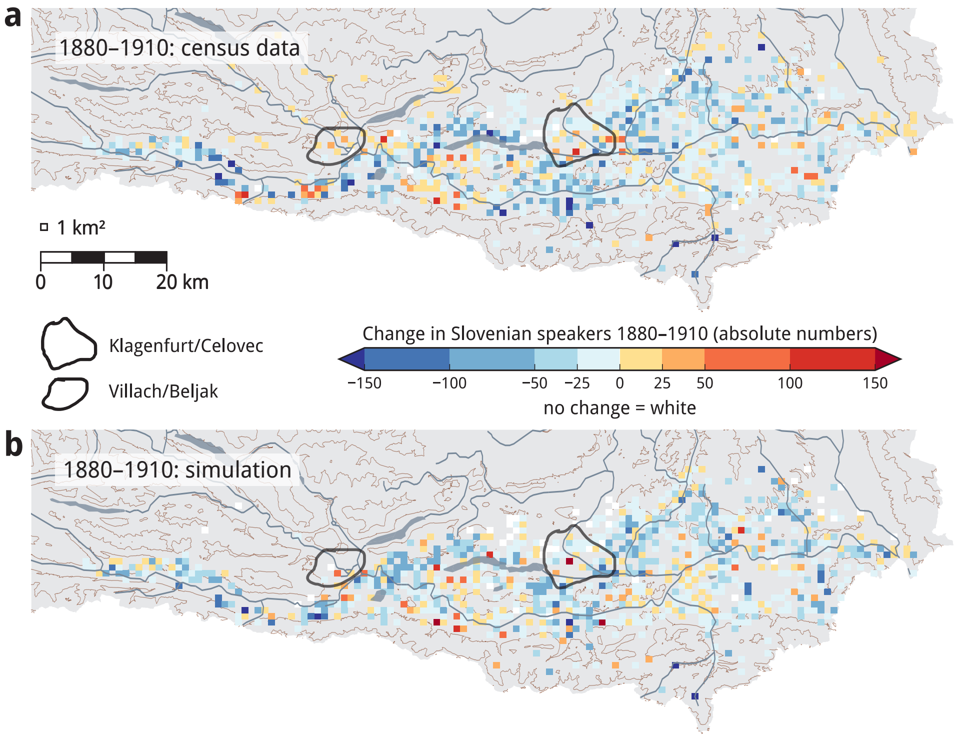


Figure 26: Increase and decrease in the number of Slovenian speakers in southern Carinthia between 1880 and 1910. (a) Census data. (b) Optimum simulation without habitat parameters. Increase is shown in shades of red, and decrease, in shades of blue. Numbers shown are absolute numbers. The two biggest Carinthian towns, Villach/Beljak and the capital Klagenfurt/Celovec, are encircled according to present administration. The optimum simulation with the bilingual schools habitat parameter is not shown because there is no visible difference compared with (b). Reprinted from [1].

Extension of the model through habitat parameters

In a second step the influence of habitat conditions, such as the influence of urban areas, i. e. major towns, the language of schools and language in parishes was investigated for the period from 1880 to 1910. To this end, we introduced a habitat parameter h_i into equation (13) which modifies the effect of local speakers by an exponential function with the argument $(\pm H h_i)$:

$$n_\alpha(\mathbf{r},t) \mapsto n_\alpha(\mathbf{r},t) \cdot \exp(\pm H h_i) \tag{24}$$

We assume that the effect is symmetrical, i. e. if the effect on Slovenian speakers is given by $n_\mathrm{S} \cdot \exp(+Hh_i)$, then the effect on German speakers is given by $n_\mathrm{G} \cdot \exp(-Hh_i)$. In the presence of an external influence i, H is set to 1 and the coefficient h_i gives the strength of influence. In cells without an external influence $H = 0$ and equation (13) is recovered. The exponential function was chosen as a modifier because it is a simple function which for small h_i adds $n \cdot h_i$ to the speaker effect if $H = 1$, while recovering the basic model (equation (13)) if $H = 0$.

The multiplicative factor H indicates the presence ($H = 1$) or absence ($H = 0$) of a local habitat condition i. e. $H = 1$ for the two largest towns Klagenfurt and Villach or if a bilingual school or Slovenian parish existed. Otherwise H is set to zero. We obtain the following equation for the probability $p_\alpha(\mathbf{r},t+1)$ of speaking a language α at position $\mathbf{r}$ and time $t+1$:

$$p_\alpha(\mathbf{r},t+1) = \frac{n_\alpha(\mathbf{r},t) \cdot \exp(\pm Hh_i) + F_\alpha(\mathbf{r},t)}{n_\mathrm{S}(\mathbf{r},t) \cdot \exp(+Hh_i) + F_\mathrm{S}(\mathbf{r},t) + n_\mathrm{G}(\mathbf{r},t) \cdot \exp(-Hh_i) + F_\mathrm{G}(\mathbf{r},t)} \tag{25}$$

Optimization was performed as before. We investigated three parameters which were chosen because data were available: urban areas, bilingual schools and parish language. Of these three investigated parameters, only the bilingual schools showed a small influence. However, the influence is so small that figure 26b does not visibly change with the introduction of the bilingual schools habitat parameter.

Urban centers

Language change patterns differ depending on whether the environment is rural or urban. Fishman [175] argues that speakers in an urban area are typically more likely to shift from the minority to the majority language while the inhabitants of isolated rural areas resist language shift. Movement to larger urban agglomerations therefore increases the chance of giving up the minority language in favor of the majority language.

This development is marginally noticeable in the period from 1880 to 1910 for the two largest towns Klagenfurt and Villach. In these two towns, the number of Slovenian speakers decreases slightly faster than the basic model (equation (18)) predicts. An interesting phenomenon appears between 1971 and 2001 when loss of minority language by moving to urban centers is reversed: the number of Slovenian speakers now definitely increases faster in urban centers than predicted by the basic model. These localized developments can be captured by our model by introducing a parameter h_1 and setting $H = 1$ in the largest towns for each period (Klagenfurt and Villach).

Best fit for the period from 1971 to 2001 is provided by $h_1 = 0.025 \pm 0.005$. h_1 is positive for this period which means that Slovenian speakers have more impact (equation (24)). The model with this urban habitat parameter better describes the actual data in these urban centers in the sense that it better reproduces the direction of change, that is, decrease or increase.

Bilingual schools

Between 1880 and 1910 so-called utraquistic elementary schools were meant to teach pupils in both languages [176]. In 1880, these schools existed in 83 settlements in the bilingual region of southern Carinthia [131]. We examined whether in these settlements ($H = 1$) the Slovenian language was preferentially preserved compared to localities where no such school existed. Best fit for the period between 1880 and 1910 was achieved with $h_2 = -0.0224 \pm 0.0050$. From equation (24) it follows that the presence of an utraquistic school has a negative impact on the influence of existing Slovenian speakers. After World War II the bilingual instruction system in elementary schools was repeatedly changed and unfortunately no detailed data are available on how many pupils attended classes in Slovenian language.

Parishes

In settlements with a Slovenian majority from 1880 to 1910 mostly Slovenian native-language speakers were hired as priests [177]. They read the mass in the Slovenian language. Altogether, there were 98 Slovenian-language parishes in the bilingual region in southern Carinthia in 1880 [132]. We examined the influence of these parishes on the development of Slovenian by applying the same procedure as for the schools: $H = 1$ in settlements or towns with masses in Slovenian, else $H = 0$. Neither in the first nor in the second period we could find a substantial influence of Slovenian-language parishes on the probability of speaking Slovenian.

Language shift in the period 1971–2001

After simulating the language dynamics during the Austro-Hungarian Empire, we now turn to language development in the second period after the two World Wars. The development from 1971 to 2001 was first pursued using the same basic model (equation (18)). Table 5 shows the numerical results and figure 27 shows the increase (red) and decrease (blue) of the number of Slovenian speakers in southern Carinthia for census data and simulation. We also investigated the influence of two habitat parameters for this period: urban areas and parish language. Only the urban habitat parameter resulted in a noticeable difference in goodness of fit (figure 27c). Errors depending on the number of Slovenian speakers present are shown as before in figure 22.

The difference in figure 27c between the outskirts and inner cells of the city of Klagenfurt is a result of the model dynamics: The outer cells have populated neighbor cells only on one side whereas the inner cells are completely surrounded by populated cells. Introducing a habitat

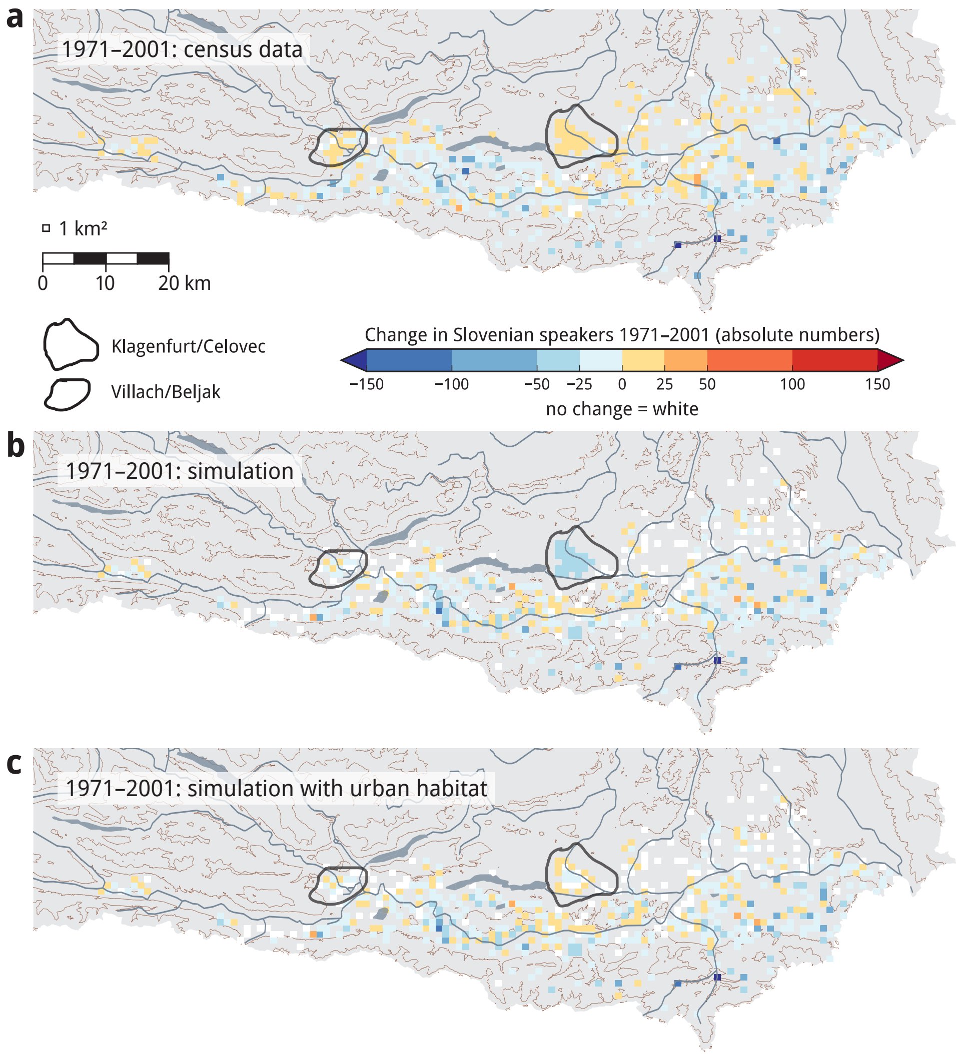


Figure 27: Increase and decrease in the number of Slovenian speakers in southern Carinthia between 1971 and 2001. (a) Census data. (b) Optimum simulation without habitat parameters. (c) Optimum simulation with urban habitat parameter. Increase is shown in shades of red, and decrease, in shades of blue. Numbers shown are absolute numbers. The two biggest Carinthian towns, Villach/Beljak and the capital Klagenfurt/Celovec, are encircled according to present administration. In (b), no additional habitat parameter was introduced, and a difference between census data and simulation in the two urban centers Klagenfurt and Villach is particularly visible in this period. This difference indicates a deviating development in urban areas, which requires the introduction of an additional habitat parameter. Reprinted from [1].

condition h increases the probability of speaking Slovenian in the outer cells compared to the model without habitat. On the contrary, in the inner cells the effect of h is compensated by interaction (F) with German speakers in the neighboring cells and the increase in Slovenian speakers is not as strong. This color difference would vanish for larger values of h.

Local differences in the language diffusivities

We have cut out three regions in the districts of Völkermarkt, Klagenfurt and Villach in order to search for local differences in the diffusion behavior: is there a difference between rural and urban regions?

In all three regions in the districts of Völkermarkt, Klagenfurt and Villach the diffusivity of the Slovenian language D_S is between 25 and 50 percent lower than the diffusivity of the German language D_G, largest difference (factor of two) for the urban region of Klagenfurt. We suppose that the discrepancies between the different regions are due to local differences in language spread and retreat because of differences in geography and population distribution: faster diffusion in urban areas, German diffusing particularly faster than Slovenian in urban region of Klagenfurt.

7.4.3 Discussion

[*Parts of this section are a translation of sections 4.4 and 5 in [4].*]

Modeling means creating hypotheses and testing them on empirical data. After this testing, hypotheses can also be rejected when the simulation results do not adequately match the empirical data, i. e. if the model cannot reproduce the data with sufficient precision. In this case, we have seen in the previous section that the basic model (where speaker number is only dependent on the current speaker number and interaction with others) is able to reproduce the census data with satisfactory agreement. This model was then extended through the introduction of three habitat parameters: utraquistic (bilingual) schools, parish language and the influence of urban centers. Surprisingly, no significant influence of parish language and a small negative influence of school instruction language was found. The lack of influence of parish language might be explained by the data quality: The listing of bilingual parishes used for modeling gives no indication on how much Slovenian is actually used during church activities and whether that amount is the same in all parishes marked as bilingual or Slovenian. One wagers that it is probably not. As the negative influence of bilingual schools seems especially surprising, it will be discussed here in more detail because it shows once again the necessity of seeing data on language use in context.

During the time of the Dual Monarchy (Austria-Hungary), the Slovenian language was taught in Carinthia in so-called "utraquistic" elementary schools. These were nominally bilingual, although Slovenian was used as a language of instruction only in selected subjects [176]. Completely Slovenian-language schools existed only in very few places (at the beginning of World War I there were three in all of Carinthia [178, p. 156]), as they had been replaced by utraquistic schools [116, p. 43]. The introduction of utraquistic elementary school led to many conflicts.

The Slovenian ethnic group repeatedly accused the German-speaking side of unduly favoring the German language and having the “Germanization” of the population as a goal [178, p. 156], [176, p. 297f]. In some parts, there was also a lack of teaching staff proficient in Slovenian [178, p. 156f], [176, p. 297f], so it remains unclear to what extent teaching actually took place in Slovenian and how much Slovenian was taught.

The locations of the schools and their languages of instruction (utraquistic, German only or Slovenian only) are listed in the Carinthian teacher calendars (for example [131]) and were used for modeling. There was a slight negative (!) influence of the utraquistic schools on the language shift from Slovenian to German (in the sense that an utraquistic school weakens the local influence of the Slovenian language). Such a result can be seen as an indication that the utraquistic schools actually served to “Germanize” the population rather than to preserve the Slovenian language. Thus, this intention of “Germanization” as suggested by other sources is also reflected in the result of the simulation—the simulation reveals a large-scale tendency and puts it into numbers.

As noted in section 7.4.2, the system of utraquistic schools does not remain in place today. After the collapse of the Austro-Hungarian monarchy in 1918 and two World Wars, the Slovene minority in Carinthia was granted the right to elementary education in Slovenian by the Austrian State Treaty 1955. With the *Minderheiten-Schulgesetz* (“Minorities School Act”) for Carinthia in 1959, the use of the Slovenian language in primary and secondary schools was regulated by law. Compulsory bilingual education was abolished [179]. Pupils were now able to voluntarily sign up for Slovenian-language instruction, regardless of prior knowledge or ethnicity, but they had to be registered by their parents every school year. Thus, in the bilingual area of Carinthia (as defined by the Minority Schools Act) nominally Slovenian-language instruction takes place everywhere. However, the registration system also means that the number of pupils attending Slovenian-language classes in a school fluctuates constantly, which is why individual schools cannot be classified as bilingual (or not). It is impossible to draw a line on how many percent of children at each grade level need to attend Slovenian-language classes in order for a school to be considered bilingual. A purely Slovenian-speaking high school is only available in Klagenfurt/Celovec. The fundamental reforms of school education in Carinthia after the two World Wars therefore make modeling for the second period from 1971 to 2001 impossible.

The model also revealed a difference in the development in urban vs. rural areas. In the first period (1880 to 1910), the number of Slovenian speakers in the two larger cities declined more rapidly than predicted by the model for the rest of Carinthia. The urban environment probably favored the shift to the German language. In the second period (1971 to 2001), however, the trend was reversed: especially in Klagenfurt/Celovec, the number of Slovenian speakers increased more than predicted by the model. In recent times, a change of heart seems to be taking place and additional language skills in a non-majority language are perceived as an asset [180]. A similar development was observed in members of the Carinthian Slovenes who live in Vienna

[181]. This is also reflected in the increasing number of applications for bilingual education in Carinthia [182]. Another factor could be that people with a Slovenian-speaking background in the cities are again more committed to keeping Slovenian "alive" for socio-symbolic or political reasons—or a nostalgic romanticization, especially those with a higher level of education [183]. Migration probably also plays a role: it is possible that Slovenian speakers from more rural areas increasingly move to the larger cities and thus increase the number of speakers there. However, this hypothesis cannot be examined using the census data, as they are anonymous. We unfortunately cannot compare whether the Slovenian speakers in Klagenfurt/Celovec in 2001 are the same who lived there in 1971 and whether and where these speakers from rural areas came to Klagenfurt/Celovec.

Compared to other minority languages, Slovenian has the distinct advantage of having a homeland, i. e. a state (the Republic of Slovenia) where it is not a minority language and where all sorts of media and linguistic material are being produced in the language daily. This state today borders Carinthia (as did its predecessor state in the times of Austria-Hungary), so one would assume that Carinthian Slovenes benefit from the geographical proximity to Slovenia. However, Carinthia and Slovenia are separated by a large mountain range, the Karawanks, which prevented easy exchange between the two places and even hindered reception of Slovenian TV and radio across the border in Carinthia in former times. Nowadays, particularly with the advent of the internet, it has obviously become much easier to access Slovenian-language media from anywhere. It remains to be seen how this influences the development of Slovenian in Carinthia, especially considering that Slovenian dialects in Carinthia are different from the ones in Slovenia.

7.5 Preliminary conclusion

From everything said/written so far, we can now draw preliminary conclusions before going on to the second language contact situation of German–Hungarian in Hungary. To briefly sum up: We have seen from the example of Slovenian–German in Carinthia that quantitative research on language shift using modeling and simulations is a suitable tool for identifying and describing trends over a large area. More specifically, quantitative models of language shift based on models of physical diffusion can be used to identify factors influencing language shift in language contact situations. These fall into two major categories: "macroscopic" models which look at a larger area, and "microscopic" models which look at individual data points.

Macroscopic models can be used to follow the movement of a border separating two language areas. However, defining such a language border is not straightforward and difficult to do objectively. In some cases, such as in Carinthia for the second period (1971 to 2001), it does not seem possible at all: when the speaker number has been drastically reduced through language shift, no contiguous language area exists any more. In this case, microscopic models such as our model based on grid cells offer an alternative for a more detailed and complete description of language shift. The microscopic model developed in this thesis also takes into account the interaction with all neighboring cells, instead of interaction from only one direction as in the language border model. More generally, we have demonstrated that a microscopic model based on models for physical diffusion can be successfully used to describe language shift.

However, microscale models are only possible where data are available in sufficient resolution (with high temporal / spatial accuracy). This point is also the biggest challenge because such data are not available for many language contact situations. Census data represent a possible data source, as argued in section 4.6, even if the goal of the census is not the direct survey of linguistic competence or if the results do not necessarily reflect the actual use of language. However, as language shift is a multifactorial problem, we need data on many factors for an appropriate model which need to have the same resolution as the census. The census also provides other data (for example, socioeconomic indicators such as commuting) alongside data on language use. However, this data cannot be directly linked to the language use data due to the anonymity of the census: If in a settlement at a time t 30 people speak German and commute to Klagenfurt and 10 years later 30 people in the same place speak German and commute to Klagenfurt, then we cannot assume that they are the same 30 people. Due to this problem of not being able to track groups of people, factors like the influence of commuting to large cities was not investigated, even though data on the commuter numbers are available through Statistics Austria (albeit only at the level of the municipalities and not the settlements, that is to say in a significantly worse spatial resolution—hence it would have only been possible to take commuting into account statistically, e. g. by estimating the fraction of commuters in a municipality and applying this same number to all settlements in the municipality).

Our basic model uses only two influence factors, both of which can be calculated from census data: the number of speakers, and interaction with other speakers. This makes it possible to follow language shift using only one data set. Another advantage: by including only parameters whose value can be directly derived from data, the model can be applied in situations where data on other factors cannot be obtained or are simply not available.

Application of our model has helped identify the primary factors influencing language use and lets us answer the question: What drives language shift in southern Carinthia? Or, conversely, what prevents language shift? The number (and indirectly the fraction) of speakers (directly in each cell and indirectly through the interaction with surrounding cells) seem to be a good predictor of language shift: if there are many possibilities to use the language, then language shift occurs less frequently. Once we remove interaction from equation (13), the probability of speaking a language remains constant.[78] Interaction (and accordingly the number of speakers

78 As this is a very simple model, we have not modeled the interaction *within* a cell. Such interaction in the same cell can change the overall probability of speaking a language even in the case of an isolated settlement without interaction with neighboring cells. Birth and death rates were also not considered. However, assuming these rates are the same for both language groups (i. e. Slovenian speakers do not die faster or have more children than German speakers) and

in the neighborhood of a cell) is therefore essential for changes in language use (in whichever direction), which has been argued by previous linguistic studies [33; 34] and is confirmed by our simulation.

We have also seen that while the mathematical description and modeling of complex processes can help us understand the underlying relationships, they do not necessarily provide us with waterproof evidence out of the box. This is exemplified in the influence of the school language in southern Carinthia as described above: According to the data, the simulation shows a negative influence of a utraquistic (bilingual) school on the preservation of the Slovene language. From this, we must not conclude that bilingual schools generally harm language preservation and argue for their abolition.

With the background knowledge of the actual nature of the utraquistic school, the result is better understood: In its execution, the utraquistic school may not have served the preservation of the Slovenian language because it was instrumentalized by politics to push the interests of one side (the German-speaking side). The lack of data on the language of instruction in the second period also indicates opportunities for future research. Quantitative description of language shift can therefore never replace the qualitative approach—but it can complement it excellently.

the transmission of each language to children is equal to the probability of speaking a language within a cell (i. e. not only the dominant language is taught to babies), the probability is not changed.

8 Situation II: German–Hungarian in southern Hungary

In the previous chapter we have presented a model for language shift in Carinthia, a part of Austria. We would obviously like this to be a general model for any possible language contact situation, but so far we have only shown that it works for one situation. In this section we will therefore have a look at another test case for our language shift model. I will introduce the case of German–Hungarian in southern Hungary, a language contact situation in the Hungarian half of Austria-Hungary (and later Hungary). Here, language shift is occurring, though the circumstances are different from the situation in the Austrian half and later Austria (see section 6 for a more in-depth look at these differences). As with Carinthia, a short overview of data preparation is given before applying the microscopic model to the data. From the results, conclusions can be drawn for the modeling of different language shift situations. These "lessons learned" will be discussed in the last chapter of this thesis.

8.1 Historical background

The first German settlers on the territory of today's Hungary (and the much larger Kingdom of Hungary in Austria-Hungary) arrived in the 10th century after the establishment of the Kingdom of Hungary with the coronation of Stephen I (Stephan/István) and his German wife, Gisela. Later waves of immigration followed during the Middle Ages, but were interrupted by the war against the Ottoman Empire in the 16th century. The larger German-speaking groups and areas which remain today arose from an active recruitment policy of the new Habsburgian rulers and the residing great landowners after the expulsion of the Turks [137; 184; 185].

When the Ottoman Turks left after their defeat, the region, now in the Habsburg Empire, was very quickly re-populated with people from the territory of the Habsburg Empire and the Holy Roman Empire starting around 1720.[79] Large numbers of the new settlers coming to the area of today's Hungary were German-speaking, many from Swabia, leading to the term "Danube Swabians", which was applied to settlers regardless of their actual origin. As people came from many different German-speaking areas, the result was a linguistically diverse group and different dialects of Hungarian German [158; 159].

---

79 Previously, at the end of the 17th century, the area was briefly re-populated with South Slavic peoples. These were asked to assimilate or leave after conflicts with the lords of manor arose, and the overwhelming majority left [184, p. 85–100].

Speakers of German[80] were distributed all over the Hungarian part of the Austro-Hungarian Empire, both in regions which nowadays are no longer part of Hungary but also in today's Hungary, around the capital Budapest and in the counties Baranya/Branau and Tolna/Tolnau south of Budapest [186]. The latter two and the neighboring county of Somogy/Schomodei are often referred to as "Swabian Turkey", deriving from the previous rulers of this area, the Turks. It is also the area where most of the remaining Hungarian Germans in today's Hungary live, and the area which we will use for modeling (excluding the county Somogy).

As noted before (section 6.2), Hungarian Germans who tended to use German as their primary language, like all other non-Hungarian speaking groups, were subject to a strong pressure for assimilation in the late 19th and early 20th century, as the Kingdom of Hungary pursued its so-called Magyarization policy—language was regarded as the basis for patriotism; accordingly, all "real Hungarians" should also speak Hungarian. Consequently, we would expect this to lead to a drastic language shift, even more than in the "usual cases" of minority languages without additional overt political pressure. However, no drastic effects can be seen in the data for the first 50 years (figure 28) although the percentage of Germans in the population declines over time and local effects are visible when taking a closer look at individual settlements. This means we may be able to apply the same model as in Carinthia because the dynamics *reflected in the data* do not seem different at this stage.

There is an important exception to the point of "no drastic effect in the data": After World War II, the attitude towards Hungarian Germans changed. They were forced to leave on a large scale (expulsion) and many voluntarily emigrated under the changed political conditions. This led to a drastic drop in the percentage of German speakers in the population which is apparent e. g. in census data (figure 29).[81] In recent times, the number of Hungarian Germans seems to be on the rise again according to the census—though we have seen that caution must be

80 Here things get a bit complicated: As I have noted, someone being a speaker of a language does not equal being part of a certain ethnic group, or the other way around. Hence we cannot assume that German speakers in today's Hungary also have ethnic ties to the German-speaking settlers who immigrated in the 18th century. However, we can assume that in the beginning, most people immigrating continued to use German as their primary language, this being their first language back in the area they had left to settle in Hungary. In addition, many German speakers in Hungary are descended from these settlers and today self-identify as Hungarian Germans, *Ungarndeutsche*, regardless of their level of use of or proficiency in German. The term "Hungarian Germans" is a bit ambiguous in that it could suggest these people have ties to today's Germany which did not exist as a country when the settlers arrived (only as part of the larger Holy Roman Empire), so Hungarian Germans are technically not the descendants of "German settlers" but rather of settlers from German-speaking areas.

81 In physics, such an abrupt change would correspond to a phase transformation. These occur under changing environmental conditions, as is the case also for the Germans in Hungary.

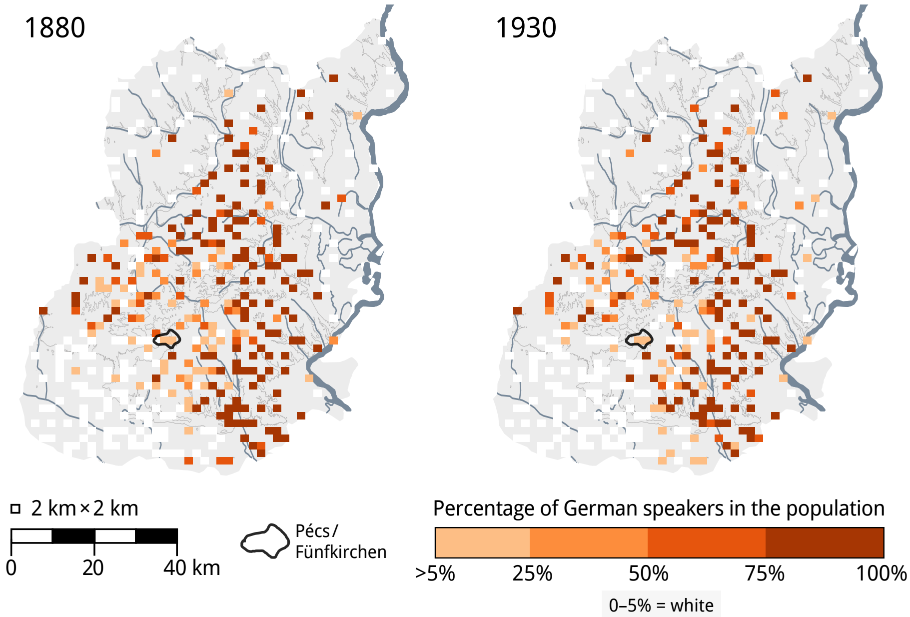


Figure 28: Percentage of “German” in census results in southern Hungary (counties Baranya and Tolna). Left: census results for 1880, right: census results for 1930. The biggest town, Pécs/Fünfkirchen, is encircled according to its 1880 borders.

taken when using numbers from the census to describe the development of a minority.[82] A recognized minority by the 2011 Act on the Rights of Nationalities (*2011. évi CLXXIX. törvény a nemzetiségek jogairól*), Hungarian Germans are also granted several rights e. g. minority self-government (*Landesselbstverwaltung der Ungarndeutschen*) or minority schools (whose status is again tied to a percentage, 25%, of children who take part in the minority education).[83]

The recent history of Hungarian Germans means that no meaningful continuous model for language shift covering the time from 1880 until 1950, 1960 and beyond can be applied in this case. This is different from the situation in Carinthia, where the number of Slovenian speakers declined over time, but there never was a drop as drastic as in Hungary, allowing us to use the same model for two time periods 60 years apart. This cannot be done for Hungary. What we can do, however, is model the time period from 1880 until 1930 (before World War II) in the same

82 The Hungarian census is especially tricky in this regard as recent ones asked multiple questions related to nationality/ethnicity and language [127] (the Hungarian census is still conducted with questionnaires unlike the Austrian one), and it is sometimes unclear which answers to which questions are taken as arguments, and whether this is justified. On the other hand, this allows for a granularity of answers pertaining to language and identity which is not possible in the Austrian census.

83 The Act on the Rights of Nationalities is the most recent legislation concerning the treatment of minorities in Hungary. See [187] for a brief history of minority rights in Hungary.

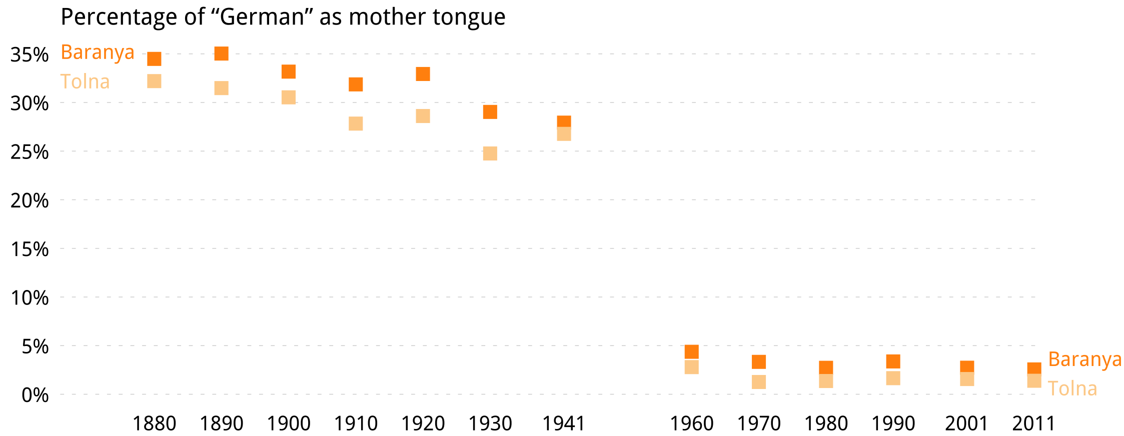


Figure 29: Percentage of "German" in census results (mother tongue) for the counties Baranya and Tolna from 1880 until 2011. Numbers take into account all languages in the population (not including people who did not answer or who could not speak) and also reflect the county borders at census time (no re-transformation to current borders). For these reasons, the fractions for 1880 to 1910 are slightly different from the ones in figure 9. Sources: [188–198].

way we did in Carinthia, to see if our microscopic model holds, and to identify factors influencing language shift in the case of Hungary. To do so, we will again turn to the census.

8.2 Preparing the data

Although the reservations about census data on language discussed earlier (section 4.2) remain, we use census data, simply because they provide the necessary amount of data in sufficient spatial and temporal resolution, and they are the only source available to do so. Censuses were conducted in the Hungarian part of Austria-Hungary with a similar bundle of questions as in the Austrian part. These censuses were continued in Hungary after the collapse of the Austro-Hungarian Empire. Concerning language, the questionnaire asked for the mother tongue (rather than for the everyday language as in the Austrian part), although there was no possibility to indicate bilingualism.

As mentioned in the previous section, speakers of German were (and are) distributed all over the area of the former Kingdom of Hungary. To study language shift from German to Hungarian, here we choose two counties (*vármegye/megye* in Hungarian, or *comitatus* in Latin) within the borders of today's Hungary: Baranya and Tolna (figure 30). As we do not extend our model to today, we use the county borders of 1880[84] for simplicity, not including the area in the south

84 These are slightly different from today's borders as established in the administrative reform of 1950 (after the end of the chosen time period, meaning the reform does not affect modeling).

of Baranya which bordered the Kingdom of Croatia-Slavonia and was ceded to Croatia after the dissolution of Austria-Hungary. The time period we use runs from 1880 (the first Hungarian census to include data on language) to 1930 (the last census before World War II). The next census in 1941 is not used because it is likely even more politically influenced than other censuses, being conducted during World War II. An extension to later times is not meaningful for the reasons outlined above: an artificial drop in the percentage of Hungarian Germans owing to deportations and emigration.

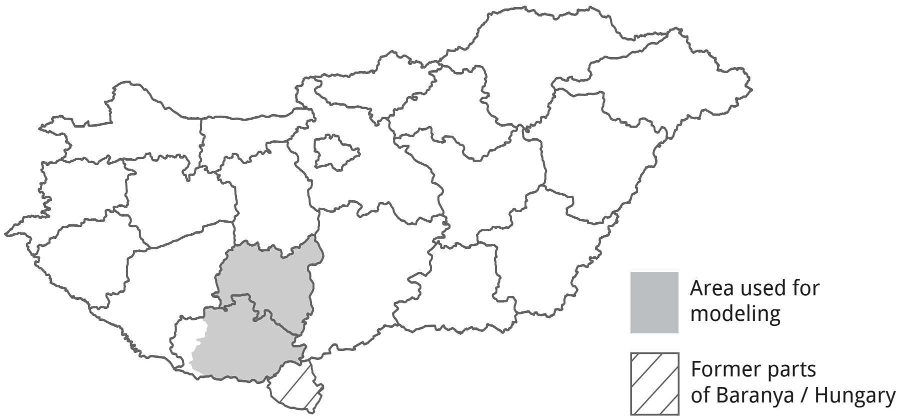


Figure 30: Map of Hungary with its counties (today's borders). Grey shading roughly indicates the area used for modeling, counties Baranya and Tolna, based on the 1880 county borders. Map data: [138; 139].

The Hungarian census data are available on the Internet as scans (processed with optical character recognition and thus searchable), however the speaker numbers still had to be typed up so they could be processed further.[85] Throughout, only one language could be given as mother tongue, so the data did not need to be consolidated (mapping bilinguals to monolinguals) as with the Austrian data. For simplicity, we use only information on the number of German and Hungarian speakers, disregarding other languages. While there are speakers of other languages listed in the census, their number is rather small compared to the number of German speakers: In only 53 of the 449 total settlements (11.8%) there are more than 100 speakers of other languages in 1880.[86] The total population of each settlement is therefore calculated by adding the numbers of German and Hungarian speakers.

85 The digitized census data for southern Hungary 1880–1930 have been made available online (https://doi.org/10.6084/m9.figshare.6115634).

86 While a limit of 100 speakers of other languages may seem high, keep in mind that settlement size in Hungary is rather large, generally $>$ 300 as seen in figure 10 (only counting Germans and Hungarians), so percentage-wise the fraction of other languages is still rather small. Plus as noted, a substantial presence of speakers of other languages occurs only in one-eighth of all settlements, so we neglect it here.

After digitization, we have data for the number of speakers (to be precise: the number of people with a certain mother tongue/first language/L1) per settlement, in 10-year intervals. These are then assigned to matrices based on their geographic coordinates the same way as in Carinthia (section 7.2): the area comprising the two counties is divided into grid cells. The total area is approximately 98 km north to south (≈1.12° of latitude) and 124 km wide (east to west, ≈1.27° of longitude). In this case, a larger cell size of 2 km×2 km was chosen because the settlements in Hungary are larger and further apart than the ones in Austria. This means it is very likely that with a smaller cell size (e. g. again 1 km × 1 km), a settlement would be distributed over multiple cells and total speaker numbers would have to be divided and assigned to multiple cells. To avoid this, a more coarse-grained abstract representation with larger cells was used (except for the largest city Pécs/Fünfkirchen which due to its size had to be divided into two cells, as with Klagenfurt/Celovec in Carinthia). Overall, the result is a grid of (124:2)×(98:2) = 62×49 = 3,038 cells. Of these, 1,956 cover the area of the two counties with the rest being outside county borders and hence not used for modeling.

The language data from the census (number of people giving a language as their mother tongue) is mapped to the grid cells based on the geographic coordinates of each settlement. These coordinates were retrieved from the GeoNames database (geonames.org). The result is a set of matrices (one for each census year and language, resulting in 12 in total, 6 per language) which represent the geographic distribution of speakers per year in an abstract way.

8.3 Applying the microscopic language shift model

Using the matrices of speaker numbers we have obtained, we can calculate the changes in speakers per cell in the same way that we did previously in Carinthia. The full procedure is described in detail in section 7.4, but a shortened version is given here for a comprehensive picture (without being too repetitive).

In our model, we calculate the probability p_α of speaking a language α (in this case German G or Hungarian H) for each grid cell $\mathbf{r}$ and each year t as follows:

$$p_\alpha(\mathbf{r},t+1) = \frac{n_\alpha(\mathbf{r},t) + F_\alpha(\mathbf{r},t)}{n_\mathrm{G}(\mathbf{r},t) + F_\mathrm{G}(\mathbf{r},t) + n_\mathrm{H}(\mathbf{r},t) + F_\mathrm{H}(\mathbf{r},t)} \tag{26}$$

where n_α is the number of speakers of language α in a cell and F_α is the interaction between speakers of the same language α. The interaction F_α effective on one grid cell at position $\mathbf{r}$ is given by the sum of the contributions of all other cells at $\mathbf{r}_j$:

$$F_\alpha(\mathbf{r},t) = \sum_{\mathbf{r}_j \neq \mathbf{r}} \frac{n_\alpha(\mathbf{r}_j,t)}{4\pi D_\alpha \Delta t} \cdot \exp\left(-\frac{|\mathbf{r}-\mathbf{r}_j|^2}{4D_\alpha \Delta t}\right) \tag{27}$$

where D_α is the diffusivity of a language, a measure for its spread. Δt is set to one year because the model calculates the interaction for each year from the result of the preceding year.

The number of speakers n_α for each grid cell is then calculated by multiplication of the probability p_α with the total number of speakers n_{total}:

$$n_\alpha(\mathbf{r},t) = n_{\text{total}}(\mathbf{r},t) \cdot p_\alpha(\mathbf{r},t) \tag{28}$$

The total number of speakers per cell and year is given by linear interpolation between censuses.

The model is implemented as a simulation using GNU Octave 4.2.0. The data from the first census (1880) are set as the initial state of the system to which the model as described here and in section 7.4 is applied. From this, the number of speakers is iteratively calculated for each year which gives us an estimated result. The model parameters (diffusivities D_α in equation (27)) are optimized by a Nelder-Mead algorithm [174] using Ordinary Least Squares to check the goodness of fit to the census data. This means that the algorithm varies parameters to minimise the sum of squared errors (SSE) between simulation results and empirical data:

$$\text{SSE} = \sum_{k=2}^{6} \sum_{i=1}^{n} (O_{i,k} - E_{i,k})^2 \tag{29}$$

Here, $O_{i,k}$ is an observed data point (census data) and $E_{i,k}$ is an estimated data point (simulated data). The squared difference between $O_{i,k}$ and $E_{i,k}$ is summed for all populated cells n and all census dates except the first one $k = 2$ to 6 (1890, 1900, 1910, 1920, 1930). The first census ($k = 1$) is excluded because it is set as the initial state of the system and hence has no error.

To see if the hypothesis behind this model (interaction as the driving factor of language shift) applies to the situation in Hungary, we compare the results of the interaction model with the results of a baseline model (constant fraction of speakers of either language, i. e. $F_\alpha = 0$ in equation (26)). We check if the interaction model performs better than the baseline by comparing the following at the end of the simulation period in 1930 (table 6):

(1) the total number of German speakers as calculated by each model. This should be close to the real number according to census data.

(2) root mean square error (RMSE) per cell:

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (O_i - E_i)^2} \tag{30}$$

where O_i is an observed data point (census data 1930) and E_i is an estimated data point (simulated data 1930). RMSE should be low.

(3) mean absolute error (MAE) per cell, which should also be low:

$$\text{MAE} = \frac{1}{n}\sum_{i=1}^{n}|O_i - E_i| \tag{31}$$

where O_i is an observed data point (census data 1930) and E_i is an estimated data point (simulated data 1930).

As no data on other habitat factors (schools, parish language) were available, the model was only used in the basic form given above. Judging from pictures (figure 31), the agreement between model and census data is adequate, although we might notice that the model tends to predict no change in German speakers (white cells) rather than reproducing the slight decrease which is taking place (light blue cells in the bottom left part). From the numerical results (table 6), we see that in the case of language shift in Hungary, our model did perform moderately better than the baseline model. The error metrics RMSE and MAE are rather high in both cases, especially compared to Carinthia (see table 4). Part of this can be ascribed to the fact that the average settlement size in Hungary (and thus the number of speakers per populated cell) is rather large. (Compare the error per category for Carinthia in figure 22—the mean absolute error rises with the initial cell value/number of speakers.) Even so, a good model would obviously result in small errors regardless of the actual value of a cell.

The total number of German speakers in 1930 is more closely approximated by the interaction model than by the baseline model, although the difference in MAE between models is small which implies that errors in speaker numbers cancel out when adding everything up: in some parts, the interaction model produces too few German speakers and in others too many (figure 32). Overall, it seems to even out somewhat, hence the total number being close to the real one. For the time-steps between 1880 and 1920 the interaction model is able to satisfactorily reproduce the total number of German speakers according to census data (see figure 33), although both models (baseline and interaction) miss the mark in 1930.[87]

Taken together, these results indicate that the interaction model is not a good fit for the Hungarian data and suggests that other models (based on other hypotheses/processes) may be superior. The point to be made here is not that our model is equally bad as the baseline (because it is not), but rather that the interaction model is not a good choice for this situation because errors remain high. In contrast to Carinthia, the interaction with other settlements which the

87 In contrast to Slovenian in Carinthia, the absolute number of German speakers barely decreases in southern Hungary over 50 years. Even so, the percentage in the population decreases owing to a likewise increase of Hungarian speakers. We also see that there is a spike in the absolute number in the first ten years owing to a general population increase. Both models are able to reproduce this spike because from 1880 to 1890, the percentage of German speakers remains nearly constant.

model is based on does not seem to be the primary driving factor for language shift in Hungary. While this may seem sobering at first glance (the model is a tolerable, but not a good fit), in the next (and final) section we will look at how this can help us better understand language shift and suggest ways to approach the Hungarian data with an improved model.

	Period		
	1880–1930		
Model	Total no. of German speakers in 1930	RMSE per 50 years (speakers)	MAE per 50 years (speakers)
Baseline model	197,810	449.35	131.14
Interaction model	176,714	344.12	122.10
Census data	156,893	—	—

Table 6: Comparison values for the goodness of fit of the baseline model (constant fraction of speakers of either language) and the interaction model (equation (13)), for German–Hungarian language shift in southern Hungary. Three metrics are shown: the total number of German speakers (closer to census data is better), the root-mean-square error (RMSE) (equation (21); lower is better) and the mean absolute error per cell (MAE) (equation (22); lower is better). All results given for best fits.

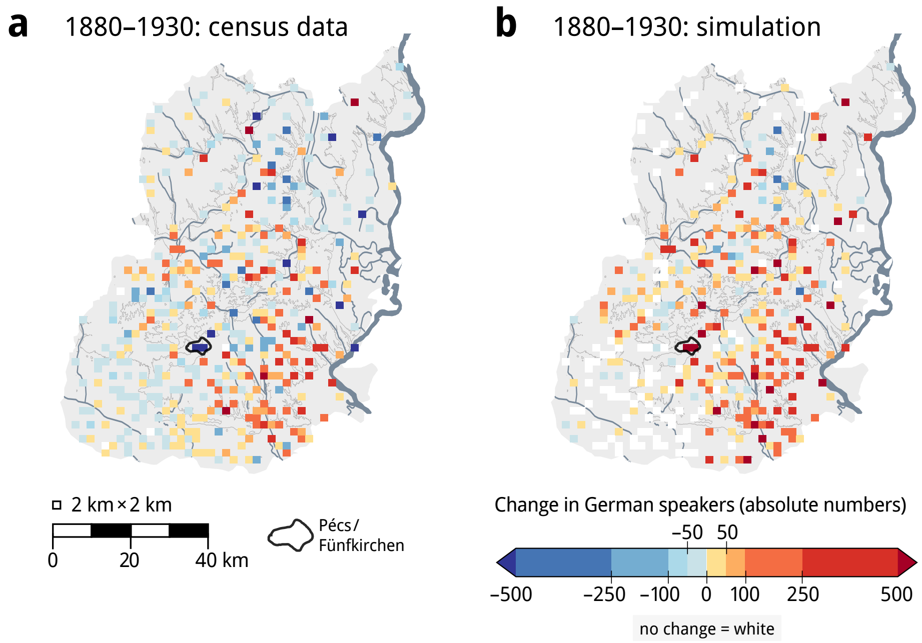


Figure 31: Increase and decrease in the number of German speakers in southern Hungary between 1880 and 1930. (a) Census data. (b) Optimum simulation. Increase is shown in shades of red, and decrease in shades of blue. The biggest town, Pécs/Fünfkirchen, is encircled according to its 1880 borders.

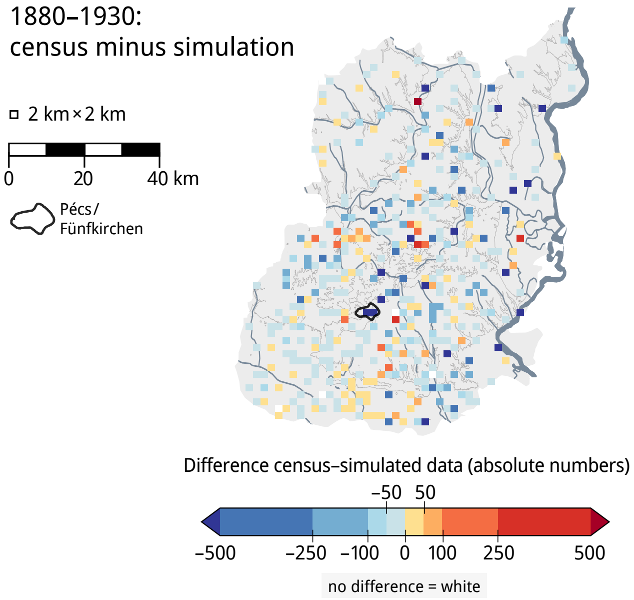


Figure 32: Residuals for the number of German speakers: census data for 1930 minus simulated result for 1930. Period from 1880 to 1930. Shades of blue: model produces too many German speakers, shades of red: model produces too few German speakers. The biggest town, Pécs/Fünfkirchen, is encircled according to its 1880 borders.

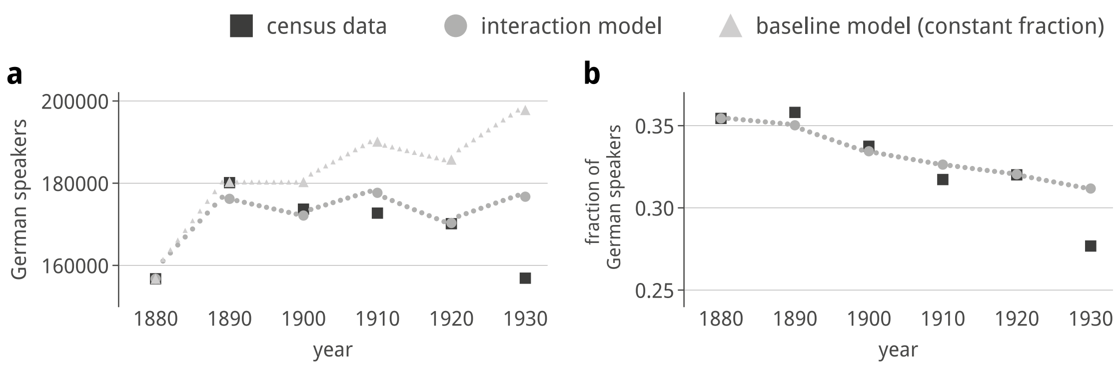


Figure 33: Development of German speakers in southern Hungary (sum of counties Baranya and Tolna) as estimated by the simulation (gray triangles and dots) and according to census data (black squares). Period from 1880 to 1930. (a) Absolute number of German speakers. (b) Fraction of German speakers in the population. In (b), the baseline model is not included because it assumes a constant fraction of German speakers.

9 Discussion: On different language shift situations and models thereof

As we have seen in the preceding section, there are unfortunately no guarantees in life: even if you have a model which works satisfactorily for one language shift situation, this does not mean it will provide a good fit for another. In this case, the model which could be used to recreate our Carinthian data worked only to a lesser extent when applied to the Hungarian data. Given the complexity of language shift and the differences between the two situations (section 6), this is not surprising, although it is obviously disappointing.

In an earlier section (section 3.1), I have written that mathematical models are tools to help us uncover the rules behind a process. They can also do this when they do not work as expected. Model failure means that either the hypotheses behind the model do not hold, or that they are put into an unsuitable mathematical form. As an analogy: it means we have either chosen the wrong tool for a process (a light microscope instead of an electron microscope to look at atoms), or the wrong configuration for our tool (an excitation energy too low on the electron microscope). In the first case, we will do more research to choose another, more appropriate tool; in the second case, we will try and try again until we have found the right configuration (or give up when we see that it is not possible given the technical boundaries). In either case, we will learn something new about the object we are trying to measure and model, as has happened here.

Coming back to language shift in southern Carinthia and southern Hungary: We have a working model for Carinthia, which can help us explain the processes underlying language shift and the factors involved as well as the (mathematical) form of their influence *in this situation*. We see that the situation in Hungary is different and cannot be explained to its full extent in the same way. However, we cannot proclaim with certainty that in Hungary, the processes underlying language shift are fundamentally different from those in Carinthia, not with the data we have—and remember that census data are not linguistic measurements, they are only statements made by people about their language use.

The point therefore is this: If the aim is to create models which are to be compared to empirical data, we are somewhat limited by the very same data. This is not a value judgment, it is how it is, and in some cases (both in physics and linguistics) it is simply really difficult to obtain data in sufficiently large resolution. Sometimes, it also simply does not make sense to do large-scale modeling, as for the German minority in present-day Hungary after the 1940s. Owing to the policy pursued by the Hungarian government towards Hungarian Germans at the end of World War II, continuity was lost as the majority of German speakers emigrated or was forced to leave. For this reason, we can create a mathematical model for language shift in Hungary, but it will likely remain a historical investigation—unless things drastically change, in which case we might witness language shifts from Hungarian to German (!). As it stands, the current number of German speakers in Hungary is very small and mathematical modeling runs the risk of creating artifacts (effects which are not there) due to a tiny sample size. However, we know a lot about language use and shift in the area from smaller-scale (socio-)linguistic studies e. g. [199–204].

There are many ways to go from here, and many possible approaches to take (several of which have been taken in previous mathematical models). We have seen that the Carinthian model is not applicable, but then what is? We have also seen that close attention must be paid to the data when creating a model. The Hungarian data represent or should represent people's mother tongue, and we can think about what processes cause changes in the population's mother tongue. One possibility has been mentioned before and is a "risk factor" for the endangerment of languages: declining transmission from generation to generation [25–27], meaning that the language is no longer taught to infants and the speaker community becomes older and older with no new speakers being added.[88] Such a situation might occur for reasons of prestige (parents expect better socioeconomic opportunities for their children with a language different to their own, as was the case in Hungary where jobs in administration and government (and likely education) were tied to knowledge of Hungarian [137]), political pressure (also present in Hungary during Magyarization) or in mixed-language marriages/inter-marriages = exogamy (especially in situations where one language has become a minority very small in number and only inter-marriages are possible, unless other factors preclude them e. g. strong religious reasons which necessitate in-group marriages = endogamy).[89]

The different variables encoded in the census data (mother tongue vs. everyday language) are likely not the only (or perhaps not even the main) reason why the interaction model does not work well for the Hungarian data, given the uncertainty of how people interpret the question on the census and that the Hungarian "mother tongue" was also intended to be a language the person knows best (compare the official instructive explanations cited in section 6.1). We have already noted the differences in settlement structure (section 6.3), that in southern Hungary settlements are further apart than in Carinthia. For this reason, the Gaussian distribution might simply not be a good approximation for interaction because it falls off too quickly. In the case of Hungary

88 In this case, "the language dying out" is for once a literal statement as people still using the language keep passing away.

89 Migration to a different country — for whatever reason — is also a situation where children might get taught only the language of the "new" country but not the language of their parents. This does not factor into the situation in southern Hungary, where we can assume that the overwhelming majority of Hungarian Germans which show up in the census had been living there for a long time and was likely descended from German-speaking people arriving in the 17th and 18th century.

with its settlements spaced far apart, it might be more advantageous to use a distribution with a longer tail.

Another factor is probably something which does not even have to do with language *per se*, although it is a source for changes in language distribution: migration. People moving and taking their language with them—or changing it once arrived in a new village—introduce an external movement which is not accounted for in the interaction model. The interaction model, while not explicit in its interpretation, aims to model the transfer of languages rather than that of people. These "free-floating" languages are then mapped to the population of a settlement. This is obviously not perfect, it should rather be seen as a first approximation. In reality, both languages and people diffuse; sometimes at the same time, sometimes not.

Considering human migration is important particularly when there are effects in the data which cannot be accounted for by changes in language transmission. In the interaction model, there is no specific model for population changes. The implicit assumption is that changes are mainly due to births and deaths within a settlement. This works for Carinthia, because even for the largest town the population rises rather slowly compared to southern Hungary: Klagenfurt/Celovec goes from around 18,000 inhabitants in 1880 to around 25,000 in 1910, and from around 80,500 in 1971 to around 82,000 in 2001.[90] In contrast, the largest city in the two Hungarian counties we studied, Pécs/Fünfkirchen, grows from about 28,000 inhabitants in 1880 to almost 50,000 in 1910. By 1930, the population had risen to over 61,000. So while even for Carinthia attributing population growth purely to a net difference in births and deaths might be a stretch, this is completely unrealistic for southern Hungary.

In any case, migration should be included in a better model for language shift. The obvious problem is because of its anonymity, census data cannot be used to follow individual people and see where they go. This would therefore have to be estimated by seeing where the population decreases or increases more than predicted by birth/death rates. In addition, each village can be assigned an "attractiveness" which depends e. g. on the number of people speaking the same language there. This attractiveness (a mathematical function of variables, ideally ones for which data is available) would then govern the migration between villages and be able to reproduce the flow of people and the changes in population—similar to physical diffusion which is driven by a gradient in chemical potential, i. e. particles essentially flow according to local changes in chemical potential.

An improved model for the Hungarian data could therefore consider both factors: migration and processes changing the use of a language (language shift in the narrow sense). The second, language shift, can still be modeled mathematically similar to the interaction model: the probability of changing your language—or not—depends on the number of people using the same

90 These numbers represent Austrian citizens as the census only gives data on language for citizens.

language around you, and their impact decreases with distance.[91] As hinted above, this does not have to be in the form of a Gaussian distribution; other distributions can be chosen. For modeling, we would calculate in each time step the population changes by births and deaths and migration (governed by a function giving the attractiveness of places, see above). Then, the new population would shift their language—obviously not everyone but a fraction. This shift would be driven by local language distributions: If one is in a predominantly Hungarian area, then that person is more likely to shift to Hungarian.

This model would retain the basic principles of the interaction model presented in this thesis, but would (hopefully) provide a better fit for language shift in Hungary. It would also make the interpretation more explicit as it does not create immaterial language amounts which are then mapped to people in order to obtain the number of speakers. Instead, it directly models human behavior—people might shift their language, but this is done individually for each person (and only for a subset of the population) rather than creating a probability for the whole cell and re-calculating everyone's language according to that probability in each time step.

In the end, it might also be the case that we need even more variables to adequately describe the Hungarian data apart from migration and language shift in the narrow sense. There might be multiple divergences between Carinthia and southern Hungary beside the different census variables and the role of migration, which contribute to differences in model fit. Not all of these divergent factors are available to us, and some might never be revealed.

We have seen that the interaction model as presented here does work, but also that it does not tell the full story of language shift in southern Hungary. However, there are many avenues to be investigated further, some of which are sketched above (in words and not yet in definite mathematical formulas, because this would go beyond the scope of this thesis). For the moment, we know that language shift can be followed by models inspired from physics and replicate a large-scale, multi-decade empirical data set (this is sometimes called a "proof of concept")—and that there is lots of space for refinement. After all, any model is a simplification of a complex process, an abstraction from the real world, and the interaction model is no different.

91 The use of cellular automata would not even be necessary, given that distance can be calculated solely from geographic coordinates.

10 At the end: Summary, lessons learned and outlook

To round off this thesis, I would like to summarize the ground we have covered and the main points to take away from this work.

We have started out by looking at language shift, a situation where people gradually change the language they use, either completely (for all of their communicative purposes) or for certain domains (e. g. communication with their children, or at work). After an overview of the multiple viewpoints and terminology concerning changes in language use, we have seen that language shift is a complex phenomenon influenced by many factors. However, it is generally hard to determine the specific factors contributing to a certain language shift situation. In other words, we would like to know what exactly drives people to shift their language.

I have proposed a new tool to help us in this regard: mathematical models inspired by physics, which can help identify trends on a large scale and give us a better insight into how language shift works on a general, society level rather than the level of the individual (the latter of which can be investigated by qualitative case studies in more detail). This approach is motivated by analogies between physics and linguistics, in this case the analogy between the movement of atoms and the spread of linguistic behavior, i. e. spread of the use of a new language. More specifically, I have shown how to go from diffusion in physics to diffusion in linguistics, and discussed similarities and differences along with existing models of language shift. As physics makes heavy use of mathematics, I have introduced the principle of mathematical modeling as a tool to look at empirical data and processes leading to changes in data (as in language shift, where the language(s) used by a person changes).

To create a model based on empirical data, we have looked into possible sources of data on language use and the pitfalls involved in creating these data in the first place (how to measure a person's language?) and working with them (what do the numbers tell us?). We have seen how language surveys are rarely "objective" and heavily depend on the circumstances. We have also seen that they are used for political decisions and to answer questions they have not asked—especially census data on language which is often invoked for granting minority rights. However, as pointed out in the same section, census data are often the only large-scale source we have covering both a large area and a large time-frame in regular intervals. Accordingly, the census was chosen as the data source in this work, while keeping in mind that any data on language use must be critically put into context rather than taken as absolute fact, and that when we are modeling such data, we are only replicating the data, the statements people make on the census, not their actual language use (though the two hopefully do overlap to some extent).

In the following part, we have combined mathematical modeling with empirical data on a scale and in a spatio-temporal resolution never done before: over periods of 30+ years and covering an area of over 3000 km^2. We have chosen a historical language contact scenario and its aftermath as our laboratory for this experiment: multilingual and multinational Austria-Hungary, and its successor states Austria and Hungary today. We have taken a closer look at two language shifts in this area: (1) from Slovenian to German in southern Carinthia, Austria and (2) from German to Hungarian in southern Hungary. Before beginning the actual modeling procedure, we took a step back and examined how local circumstances influence possible mathematical models of language shift. To do so, we have compared our language contact situations with regard to two specific factors: language policy (the legal status of the languages involved and the attitudes towards them), and settlement and population structure. From the analysis, we were able to uncover the consequences these factors have for models of language shift and show how different data sets enable different models.

After the comparative data evaluation, we undertook the actual modeling. Starting out with Slovenian in Carinthia, we have first looked into a macroscopic (larger-scale) approach to language shift: trying to quantify the movement of an abstract "language border" (a term problematic in itself), the line separating the area where mostly Slovenian is used and the area where mostly German is used. Unsurprisingly, this was unsatisfactory. We were able to derive a movement speed of the "language front". However, this speed averages over the whole area and does not consider local differences such as the fact that there is little movement/change in some areas and substantial movement in others. Applying different thresholds for defining the area where mostly Slovenian or German is used also showed that in some cases no continuous language region can be defined at all and the concept of "border movement" is void.

As we found the macroscopic approach to be inadequate, we turned to a microscopic (smaller-scale) approach where the development in individual settlements (or rather cells) is modeled. We first gave an overview of how to prepare the data by creating an abstract representation of the area as grid cells. The next step was to create rules (hypotheses) for the change in each of these cells. The validity of our rules was then checked against the census data. For Carinthia, we were able to identify one main factor for determining the number of people using a language (or more accurately: reporting it on the census):[92] the (potential) interaction with other people using the same language. The (potential) interaction was split into two terms: (1) interaction with people in the same grid cell, given by the number of people previously using a language in the same grid cell, and (2) interaction with other people using the same language in surrounding cells, quantified by the number of people in surrounding grid cells weighed by their distance.

92 As language shift is measured among others by the number of people using a language and its change over time, this can be seen as a driver of language shift (or rather a deterrent of shift).

The mathematical formula deduced for interaction is motivated both by linguistics (interaction decreases with distance and increases with the number of people, as in [34]) and physics (it is given as a Gaussian distribution which is used to describe the diffusion of particles, i. e. the concentration of a substance over time). This means we could successfully describe language shift in Carinthia, for two 30-year periods, both historically and in recent times.

In addition, we have created an extension of the model to examine the influence of so-called "habitat factors", i. e. factors which act locally. We then investigated two factors for which data were available (bilingual schools and parish language) as possible other drivers of language shift. Surprisingly, neither factor showed a significant influence, with bilingual schools even showing a slight negative influence on the persistence of Slovenian. However, this result becomes clearer against the historical background: these schools in Carinthia were never intended to promote actual bilingualism, but were rather meant to guide non-German-speaking children towards German. It also demonstrates once again the importance of setting empirical data in their proper context. We could also identify a deviating development in the two largest cities Klagenfurt/Celovec and Villach/Beljak, where the use of Slovenian declined faster (first historical period) or slower (second recent period) than in the rest of the area, showing an urban-rural divide in processes of language shift.

The last part of this thesis looked at the second language shift situation in southern Hungary. This served as our test case for the general applicability of the Carinthian interaction model—after all, simply because something works for one example does not mean it is universal. Using the same procedure as in Carinthia for data preparation, we then tried to model the data in the same way, by the same influence factors (number of people using a language in a cell and in surrounding cells). In this case the model did provide a better result than a baseline model chosen for comparison, but it still was not a good choice for this situation with the error remaining high. This tells us a number of things, mainly that different language shift situations are different (which should come as no surprise after having seen the wealth of factors influencing language shift given in the literature) and that the interaction model cannot explain the data to their full extent.

What does the interaction model miss? In the case of Austria vs. Hungary, there might be differences in the data material at play: the Austrian data at face represent statements about everyday language, while the Hungarian data should represent statements about mother tongue. Even though both are self-reported data and not actual linguistic observations, we can assume that the statements people make on the census are to some extent influenced by the question, hence the answers being different. So it is entirely possible that even *statements* about different aspects of language use (no matter how much they overlap with actual language use) cannot be represented by a single process in a mathematical model. (This is even more likely for actual linguistic data on everyday language and mother tongue, both of which are affected differently by language shift.) The two censuses with different measurement variables concerning language might result in the same numbers by chance, but this does not seem to be happening here.

Instead, it seems that there are additional factors influencing language shift which could be neglected in Carinthia but are clearly visible in the data for southern Hungary and thus cannot be ignored. One such additional factor could be differences in settlement structure: because Hungarian settlements are spaced further apart, the used Gaussian distribution is perhaps not a good mathematical fit to describe interaction between settlements.

As another example, I mentioned the role of migration to explain the large population growth of the largest town Pécs/Fünfkirchen in southern Hungary. Between 1880 and 1930 the population more than doubles and this cannot be attributed solely to births and deaths. As a last point, I outlined a new way to approach the Hungarian data—an extension of the interaction model, if you will—which combines migration and language shift. Migration would be driven by the attractiveness of a settlement, a function governing the flow of people from one place to the other. This is loosely inspired by physical diffusion, which is driven by something called the chemical potential and its local differences—only in this case it is the people who move rather than the languages. In this model, we would first calculate the population of a settlement by taking into account births, deaths and migration. Then, a subset of the population would shift their language based on local conditions such as the local language distribution and/or the potential for interaction with others using the same language. The second part can be mathematically modeled similar to the microscopic interaction model presented previously, in that the (potential) interaction increases with the number of speakers and decreases with distance. Such a model combining migration and language shift would explicitly only model people and not languages, making the interpretation more straightforward than in the interaction model which has to go through the additional step of transforming the virtual amount of language back into numbers of speakers.

So, after all is said and done what lessons have we learned in these pages, and how do they relate back to the research questions posed at the beginning (section 1.2)?

- **How can language shift be described by mathematical-physical models?**

  We have seen that there are ways to use mathematical models inspired by physics to describe and follow language shift—I have demonstrated one such way in this thesis, by developing a model motivated by physics but also incorporating factors backed by linguistic research. I hope to have demystified the procedure enough by giving a complete walk-through so that it can and will be replicated for other language shift situations—the latter part being especially important because we have seen that the model presented for Carinthia does not reveal all to the same degree in another situation, with another data set.

Throughout this work I have argued (and hopefully shown) that models based on empirical data are worthwhile.[93] Modeling with data goes both ways: the model tries to emulate the data and is checked against data, but is also *a priori* informed by the data. If a model does not work, it means it is not appropriate either for the data or for the processes being modeled (or both). The takeaway here is this: Pay close attention to the data, and be certain of what you are actually modeling. Mathematical modeling is not a magic tool which gives you an answer as long as you put data into a box. You have to know the inner workings of the box and what fits into there, and whether your data are appropriate. Mathematical modeling is a tool, but it needs someone who knows how to work it. Otherwise, as with most tools (in the strict sense, e. g. machine tools), the danger of injury is high.

– **Why do people change their language, what are the factors influencing language shift?** Given that mathematical models work for specific language shift situations, this can be answered for one situation only. For Hungary, we cannot definitely say at this point (from mathematical modeling). For Carinthia, the biggest factor playing a role in language shift was—broadly speaking—the number of people using the same language, either in the same city or in the surrounding area. In detail, it was the interaction with other speakers of the same language which in turn depends on the number of people and their proximity, for modeling quantified as a formula. These findings can be used to draw conclusions on how to foster the use of a certain language: in this case, by increasing the interacting or the potential for interaction. Here, geographic structure and distribution of settlements also have to be taken into account, as these influence interaction patterns in a speaker's network. However, while mathematical modeling can tell us that language shift is influenced by interaction (and how, e. g. is there a threshold for interaction so that a language will never "die out"? what is it? how does the process change if the proportion between interactions with the two different languages changes?), it cannot tell us what exactly to do to increase interaction.

  We have also seen in Carinthia that in larger cities, the dynamics are different from the rest of the area. This indicates that for the promotion of minority languages, attention should be paid to urban areas, although the direction of influence is not clear-cut: While larger cities generally seem to have a detrimental effect on minority language use (especially historically), they also offer more opportunities for preservation. Many language and culture clubs as well as minority language schools are started in larger cities owing to the better infrastructure and the

93 This thesis did not set out to bash abstract models without data or "prove" that models based on empirical data are superior. The idea of this work was simply to think big and study language shift on a truly large scale, and to show what can be done if one is lucky enough to have an appropriate data set. Sometimes we are limited by the data available (or not, as it were) and in this case, it is legitimate to resort to abstract models. It is obviously harder to interpret the results of abstract models if they cannot be directly mapped to an equivalent in "the real world", but this does not mean they cannot be helpful.

increased reach. These opportunities could also lead to an increase in people re-claiming their "minority heritage" by starting to learn and use the minority language again if they have family ties to it. Here, the same holds as in the paragraph above: Mathematical modeling cannot show us *what* to do, but it can show us *where* to do it.

This thesis has forged a path, but it must also be walked upon. To this end, the approach introduced in the preceding pages can (and should) be applied to further language shift situations—and it will not work for all of them. Failure along the way (though shown less frequently than the success stories) is part of the process, and also the reason I have included the chapter on language shift in Hungary where the main result seems to be that the microscopic model developed works, but is not a good fit because it does not tell the whole story. (Section 9 gives suggestions on how to improve our approach in this case.)

Theses and other pieces of scientific writing usually conclude with an outlook or suggestions of specific things to be explored still, and this one is no different: There are many avenues to take from here, from creating a model for the Hungarian data—perhaps with the approach I have outlined in section 9 which separates people and languages more cleanly than the basic interaction model—to modeling other language contact situations (of which there are many in the world) in Austria-Hungary, Europe and beyond, both historical and contemporary. In these cases, mathematical modeling could be combined with other methods, with data sources other than the census, for cross-validation or to identify discrepancies between data sources.

A simple(r) option would be to run the Carinthian model again and play around with interaction as a parameter, to investigate e. g. how low interaction can be for a language to still persist or how the ratio between interactions with different languages affects language shift. One could also improve the model I have presented, to refine it even further, for the Carinthian data or other data sets—perhaps by going into a more ecolinguistic direction and seeing how geographic factors such as proximity to the border to Slovenia (and conversely the increasing "pressure" of the majority language the closer a settlement is to the primarily German-speaking area) or the isolation of settlements affect language shift. Distances between settlements could be estimated in more sophisticated ways, e. g. by taking into account the road and rail network. Similarly, more complex models for population development and/or migration could be included. Or one could look into additional habitat factors which have not been treated here, e. g. the influence of cultural institutions and national activism (see [205–207] for a discussion on the construction of nationalist identities in the population of Austria-Hungary; see also [208] for a more general examination of the strategies employed in the making of a nation as well as the role of economic, cultural and social conditions for state-building and ethnic homogenization).

One could (and should) think at length about the separation between people and languages and how this is represented in mathematical models, an issue I have touched on only very briefly in the context of the model sketch for the Hungarian data in section 9. Another possibility would be to examine in more detail the analogies between languages and atoms and transferring different diffusion models from physics to linguistics (or see which ones are applicable at all). After all, I have barely scratched the surface concerning the mathematical treatment of atom movement in physics. Diffusion models are as diverse as language shift situations, which is perhaps the key point when it comes to modeling language shift: Ultimately, we need to keep in mind that language shift situations are all both complex and, more importantly, different, hence there is no “model of everything”, just like we have no “theory of everything” in (theoretical) physics. Nevertheless, much like theoretical physicists, we should keep pushing on, to expand the boundaries of our current framework—especially using unconventional tools, as has been done in this work.

Interactive maps

One benefit of mathematical modeling and computer simulations is that they can interpolate, i. e. estimate missing data points such as the years between censuses. This makes it possible to get a continuous visualization of the language shift process. To this end, the simulated results generated for Carinthia have been implemented as interactive maps on the project webpage at https://dcs.univie.ac.at/languagediffusion/ (English version) and https://dcs.univie.ac.at/sprachdiffusion/ (German version).

The maps display simulated data for Austria-Hungary (1880–1910) and modern Austria (1971–2001) and allow the user to follow the development of Slovenian in Carinthia on a yearly basis by scrolling—a digital flip book. Different viewing options are available: we can look at the percentage of Slovenian speakers, the absolute number of Slovenian speakers or the change in Slovenian speakers in absolute numbers (as pictured in figure 26 and 27, but per year). A screenshot is shown below in figure 34.

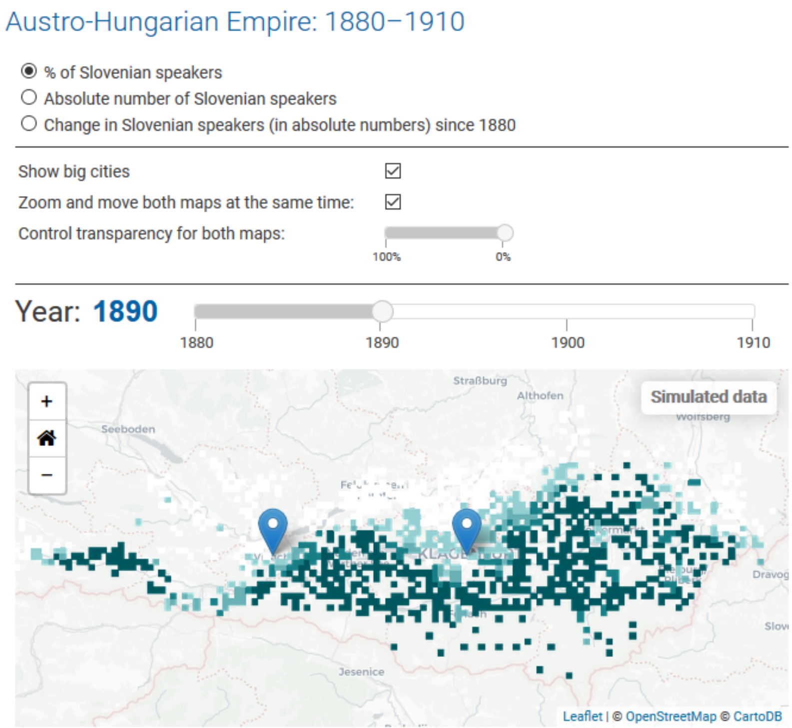


Figure 34: Screenshot of the interactive maps for Carinthia on the project webpage at https://dcs.univie.ac.at/languagediffusion/ (English version) and https://dcs.univie.ac.at/sprachdiffusion/ (German version)

Sources and references

This section is meant to give an overview of the tools used in the making of this thesis, grouped thematically. Specific sources pertaining to text sections or images are referenced directly in the text and figure captions (see the section "References" starting on page 125).

This thesis was written in LaTeX and is set in the following typefaces: Adobe Text Pro (body text), Source Sans Pro (headings), Droid Sans (text in figures). Unless otherwise noted, graphics were created by the author using Python, Adobe InDesign, Paint.NET and Inkscape, with help from QGIS. Some color schemes are based on colors from ColorBrewer (http://colorbrewer2.org). All simulations were performed using GNU Octave (version 4.0.0 or 4.2.0).

Census data

The digitized data used for modeling in this thesis can be downloaded from https://doi.org/10.6084/m9.figshare.4535399 (Carinthia, Austria) and https://doi.org/10.6084/m9.figshare.6115634 (southern Hungary). The following list gives the original source publications of this data. Data for individual settlements for the second modeling period (1971–2001) in Austria were commissioned by a special evaluation from Statistics Austria and cannot be made available due to data protection regulations.

Other census results referenced in this thesis, but not used for modeling: references [194–198].

Census forms (questionnaires and instructions) for recent censuses were obtained from the World Bank Central Microdata Catalog, http://microdata.worldbank.org/index.php/catalog (last accessed January 31, 2019).

Austria/Cisleithania

K.K. Statistische Central-Commission (1883): *Special-Orts-Repertorien der im Oesterreichischen Reichsrathe vertretenen Königreiche und Länder. V. Kärnten*. Wien: Alfred Hölder.

K.K. Statistische Central-Commission (1894): *Special-Orts-Repertorien der im Österreichischen Reichsrathe vertretenen Königreiche und Länder. Neubearbeitung auf Grund der Ergebnisse der Volkszählung vom 31. Dezember 1890. V. Kärnten*. Wien: Alfred Hölder.

K.K. Statistische Zentralkommission (1905): *Gemeindelexikon der im Reichsrate vertretenen Königreiche und Länder, Bearbeitet auf Grund der Ergebnisse der Volkszählung vom 31. Dezember 1900. V. Kärnten*. Wien: k.k. Hof- und Staatsdruckerei.

Statistische Zentralkommission (1918): *Spezialortsrepertorium der Österreichischen Länder. Bearbeitet auf Grund der Ergebnisse der Volkszählung vom 31. Dezember 1910. V. Kärnten.* Wien: Verlag der Staatsdruckerei.

Hungary/Transleithania

Az országos magyar kir. statisztikai hivatal (1882): *Az 1881. év elején végrehajtott népszámlálás főbb eredményei megyék és községek szerint részletezve, II. kötet.* Budapest: Pesti Könyvnyomda-Részvény-Társaság. https://library.hungaricana.hu/en/view/NEDA_1881_02/ (last accessed January 31, 2019).

Az országos m. kir. statisztikai hivatal (1892): *A Magyar Korona országainak helységnévtára a nagyméltóságú magy. kir. kereskedelemügyi minister úr rendeletéből, szerkeszti Dr Jekelefalussy József.* Budapest: Pesti Könyvnyomda-Részvény-Társaság. https://library.hungaricana.hu/en/view/NEDA_1892_helysegnevtar/ (last accessed January 31, 2019).

A magyar kir. központi statisztikai hivatal (1902): *A magyar korona országainak 1900. évi népszámlálása. A népesség általános leírása községenkint.* Budapest: Pesti Könyvnyomda-Részvény-Társaság. https://library.hungaricana.hu/en/view/NEDA_1900_01/ (last accessed January 31, 2019).

A magyar kir. központi statisztikai hivatal (1912): *A magyar szent korona országainak 1910. évi népszámlálása. A népesség főbb adatai községek és népesebb puszták, telepek szerint.* Budapest: Az athenaeum irodalmi és nyomdai r.-társulat nyomása. https://library.hungaricana.hu/en/view/NEDA_1910_01/ (last accessed January 31, 2019).

A magyar kir. központi statisztikai hivatal (1923): *Az 1920. évi népszámlálás. A népesség főbb demográfiai adatai. Községek és népesebb puszták, telepek szerint.* Budapest: Pesti Könyvnyomda-Részvény-Társaság. https://library.hungaricana.hu/en/view/NEDA_1920_01/ (last accessed January 31, 2019).

A magyar kir. központi statisztikai hivatal (1932): *Az 1930. évi népszámlálás. Demográfiai adatok községek és külterületi lakotthelyek szerint.* Budapest: Stephaneum nyomda részvénytársaság. https://library.hungaricana.hu/en/view/NEDA_1930_01/ (last accessed January 31, 2019).

Geographical data

Geographical data was obtained from the following sources (last accessed January 31, 2019):

- GADM (https://gadm.org, free for academic and non-commercial use)
- GeoNames (https://www.geonames.org, CC BY 4.0 license)
- Land Kärnten – data.gv.at (https://www.data.gv.at/auftritte/?organisation=land-kaernten, CC BY 3.0 license)
- Mosaic Project (http://censusmosaic.org/data/historical-gis-files, references [138; 139], free for non-commercial scientific use)
- Natural Earth (https://www.naturalearthdata.com, public domain)

- OpenDEM (https://opendem.info, Open Data Commons Open Database License)
- OpenStreetMap (https://www.openstreetmap.org, Open Data Commons Open Database License & CC BY-SA 2.0 license)

Legal texts

All legal texts were used in the current version (January 31, 2019). Full names in the primary language of the country and the year of approval are given in parentheses.

Austria/Cisleithania

Austro-Hungarian Compromise (*Österreichisch-Ungarischer Ausgleich* 1867, in Hungarian: *Osztrák-magyar kiegyezés*)

December Constitution (*Dezemberverfassung* 1867)

Treaty of Saint-Germain-en-Laye (*Staatsvertrag von Saint-Germain-en-Laye* 1919)

Austrian State Treaty (*Staatsvertrag betreffend die Wiederherstellung eines unabhängigen und demokratischen Österreich* 1955)

Minority School Act for Carinthia (*Minderheiten-Schulgesetz für Kärnten* 1959)

Ethnic Groups Act (*Volksgruppengesetz/Bundesgesetz über die Rechtsstellung der Volksgruppen in Österreich* 1976)

Hungary/Transleithania

Austro-Hungarian Compromise (*Osztrák-magyar kiegyezés* 1867, in German: *Österreichisch-Ungarischer Ausgleich*)

Hungarian Nationalities Law, Act XLIV/1868 (*1868. évi XLIV. törvényczikk a nemzetiségi egyenjogúság tárgyában*)

Croatian–Hungarian Compromise (*Horvát–magyar kiegyezés* 1868, in Croatian: *Hrvatsko-ugarska nagodba*)

Act on the Rights of Nationalities, Act CLXXIX/2011 (*2011. évi CLXXIX. törvény a nemzetiségek jogairól*)

References

[1] Prochazka, Katharina/Vogl, Gero (2017): Quantifying the driving factors for language shift in a bilingual region. *Proceedings of the National Academy of Sciences* 114 (17): 4365–4369. (cit. on pp. 7, 37, 73, 79, 83, 86, 88–91, 94)

[2] Prochazka, Katharina/Vogl, Gero (2018): Are languages like atoms? On modelling language spread as a physicist. *Glottotheory* 9 (1): 77–88. (cit. on pp. 7, 25, 30, 35, 36)

[3] Prochazka, Katharina (2018): Minderheitensprachen zählen! Über Sprachzählungen und Minderheiten(-sprachen). *Wiener Linguistische Gazette* 83: 1–26. (cit. on pp. 7, 40, 42, 44, 46, 48)

[4] Prochazka, Katharina (in print): Sprachwechsel in Südkärnten: Quantitative Beschreibung und Modellierung als Diffusionsprozess. In: Bülow, Lars/Fischer, Ann-Kathrin/Herbert, Kristina (eds.), *Dimensionen des sprachlichen Raums: Variation – Mehrsprachigkeit – Konzeptualisierung*. Wien: Peter Lang. (cit. on pp. 7, 69, 71, 95)

[5] Snow, Charles Percy (1959): *The two cultures and the scientific revolution: The Rede Lecture, 1959*. New York: Cambridge University Press. (cit. on p. 13)

[6] Rietveld, Toni/van Hout, Roeland (2005): Quantitative Methods. In: Ammon, Ulrich/Dittmar, Norbert/Mattheier, Klaus J./Trudgill, Peter (eds.), *Sociolinguistics. An International Handbook of the Science of Language and Society*. Handbücher zur Sprach- und Kommunikationswissenschaft / Handbooks of Linguistics and Communication Science (HSK) 3.2. Berlin: Walter de Gruyter, 965–978. (cit. on p. 13)

[7] Kallmeyer, Werner (2005): Qualitative Methoden. In: Ammon, Ulrich/Dittmar, Norbert/Mattheier, Klaus J./Trudgill, Peter (eds.), *Sociolinguistics. An International Handbook of the Science of Language and Society*. Handbücher zur Sprach- und Kommunikationswissenschaft / Handbooks of Linguistics and Communication Science (HSK) 3.2. Berlin: Walter de Gruyter, 978–992. (cit. on p. 13)

[8] Haspelmath, Martin/Sims, Andrea (2013): *Understanding morphology*. New York: Routledge. (cit. on p. 14)

[9] Labov, William (2006): *The Social Stratification of English in New York City*. Cambridge: Cambridge University Press. (cit. on pp. 14, 37)

[10] Dittmar, Norbert (2004): Register. In: Ammon, Ulrich/Dittmar, Norbert/Mattheier, Klaus J./Trudgill, Peter (eds.), *Sociolinguistics. An International Handbook of the Science of Language and Society*. Handbücher zur Sprach- und Kommunikationswissenschaft / Handbooks of Linguistics and Communication Science (HSK) 3.1. Berlin: Walter de Gruyter, 216–226. (cit. on p. 18)

[11] Hornberger, Nancy H. (2010): Language Shift and Language Revitalization. In: Kaplan, Robert B. (ed.), *The Oxford Handbook of Applied Linguistics*. Oxford: Oxford University Press, 412–420. (cit. on pp. 19, 20)

[12] Weinreich, Uriel (1953): *Languages in contact. Findings and problems*. New York: Publications of the Linguistic Circle of New York. (cit. on p. 19)

[13] Ammon, Ulrich (1986): *Sprache – Varietät/Standardvarietät – Dialekt*. Series B Nr. 147. Essen: Linguistic Agency University of Duisburg-Essen. (cit. on p. 19)

[14] Pohl, Heinz-Dieter (2009): Sprachkontakt in Kärnten. In: Elmentaler, Michael (ed.), *Deutsch und seine Nachbarn*. Frankfurt/Main: Peter Lang, 117–132. (cit. on pp. 19, 68)

[15] Heller, Monica (2005): Language and Identity. In: Ammon, Ulrich/Dittmar, Norbert/Mattheier, Klaus J./Trudgill, Peter (eds.), *Sociolinguistics. An International Handbook of the Science of Language and Society*. Handbücher zur Sprach- und Kommunikationswissenschaft / Handbooks of Linguistics and Communication Science (HSK) 3.2. Berlin: Walter de Gruyter, 1582–1586. (cit. on p. 19)

[16] García, Ofelia (2010): Language Spread and Its Study in the Twenty-First Century. In: Kaplan, Robert B. (ed.), *The Oxford Handbook of Applied Linguistics*. Oxford: Oxford University Press, 398–411. (cit. on p. 20)

[17] Auer, Peter/Schmidt, Jürgen Erich (eds.), (2010): *Language and Space. An International Handbook of Linguistic Variation. Volume 1: Theories and Methods*. Handbücher zur Sprach- und Kom-

munikationswissenschaft / Handbooks of Linguistics and Communication Science (HSK) 30.1. Berlin: Walter de Gruyter. (cit. on p. 20)

[18] Lameli, Alfred/Kehrein, Roland/Rabanus, Stefan (eds.), (2010): *Language and Space. An International Handbook of Linguistic Variation. Volume 2: Language Mapping*. Handbücher zur Sprach- und Kommunikationswissenschaft / Handbooks of Linguistics and Communication Science (HSK) 30.2. Berlin: Walter de Gruyter. (cit. on p. 20)

[19] Rindler Schjerve, Rosita (1996): Domänenuntersuchungen. In: Goebl, Hans/Nelde, Peter H./Starý, Zdeněk/Wölck, Wolfgang (eds.), *Contact Linguistics. An International Handbook of Contemporary Research*. Handbücher zur Sprach- und Kommunikationswissenschaft / Handbooks of Linguistics and Communication Science (HSK) 12.1. Berlin: Walter de Gruyter, 796–804. (cit. on p. 20)

[20] Britain, David (2013): Space, diffusion and mobility. In: Chambers, Jack/Schilling, Natalie (eds.), *The Handbook of Language Variation and Change, Second Edition*. Oxford: Wiley Online Library, 471–500. (cit. on p. 20)

[21] Cook, Vivian J. (1992): Evidence for multi-competence. *Language Learning* 42 (4): 557–591. (cit. on p. 20)

[22] Cook, Vivian/Wei, Li (2016): *The Cambridge Handbook of Linguistic Multi-competence*. Cambridge: Cambridge University Press. (cit. on p. 20)

[23] Haugen, Einar Ingvald (1953): *The Norwegian language in America: A study in bilingual behavior. Volume II: The American Dialects of Norwegian*. Philadelphia: University of Pennsylvania Press. (cit. on p. 20)

[24] Weber, Jean-Jacques/Horner, Kristine (2013): *Introducing multilingualism: A social approach*. New York/London: Routledge. (cit. on p. 20)

[25] Fishman, Joshua (1991): *Reversing Language Shift*. Clevedon: Multilingual Matters. (cit. on pp. 20–22, 112)

[26] UNESCO Ad Hoc Expert Group on Endangered Languages (2003): *Language Vitality and Endangerment*. http://unesdoc.unesco.org/images/0018/001836/183699E.pdf (last accessed January 31, 2019). (cit. on pp. 20–22, 112)

[27] Tsunoda, Tasaku (2005): *Language Endangerment and Language Revitalization: An Introduction*. Berlin: Mouton de Gruyter. (cit. on pp. 20–22, 112)

[28] Duchêne, Alexandre/Heller, Monica (2008): Discourses of endangerment: Sociolinguistics, globalization and social order. In: Duchêne, Alexandre/Heller, Monica (eds.), *Discourses of Endangerment: Ideology and Interest in the Defence of Languages*. London: Continuum, 1–13. (cit. on pp. 20, 26)

[29] Gal, Susan (1996): Language shift. In: Goebl, Hans/Nelde, Peter H./Starý, Zdeněk/Wölck, Wolfgang (eds.), *Contact Linguistics. An International Handbook of Contemporary Research*. Handbücher zur Sprach- und Kommunikationswissenschaft / Handbooks of Linguistics and Communication Science (HSK) 12.1. Berlin: Walter de Gruyter, 586–593. (cit. on pp. 20, 26)

[30] Gamerith, Werner (1994): *Ethnizität und ihr zeitlich-räumlicher Wandel anhand von Volkszählungsergebnissen: Das Beispiel der Kärntner Slowenen*. Klagenfurter Geographische Schriften Heft 12. Klagenfurt: Institut für Geographie an der Universität. (cit. on pp. 21, 45, 48, 50, 52, 69)

[31] Rindler Schjerve, Rosita (2004): Minderheit/Minority. In: Ammon, Ulrich/Dittmar, Norbert/Mattheier, Klaus J./Trudgill, Peter (eds.), *Sociolinguistics. An International Handbook of the Science*

of Language and Society. Handbücher zur Sprach- und Kommunikationswissenschaft / Handbooks of Linguistics and Communication Science (HSK) 3.1. Berlin: Walter de Gruyter, 480–486. (cit. on p. 21)

[32] Baker, Colin (2001): *Foundations of bilingual education and bilingualism*. Clevedon: Multilingual Matters. (cit. on pp. 21, 22)

[33] Lieberson, Stanley (1982): Forces affecting language spread: Some basic propositions. In: Cooper, Robert L. (ed.), *Language spread: Studies in diffusion and social change*. Bloomington: Indiana University Press, 37–62. (cit. on pp. 21, 22, 99)

[34] Trudgill, Peter (1974): Linguistic change and diffusion: Description and explanation in sociolinguistic dialect geography. *Language in Society* 3 (2): 215–246. (cit. on pp. 21–23, 81, 99, 117)

[35] Mühlhäusler, Peter (2010): Ecology of Languages. In: Kaplan, Robert B. (ed.), *The Oxford Handbook of Applied Linguistics*. Oxford: Oxford University Press, 421–434. (cit. on p. 21)

[36] Nettle, Daniel (1998): Explaining global patterns of language diversity. *Journal of Anthropological Archaeology* 17 (4): 354–374. (cit. on p. 21)

[37] Mufwene, Salikoko S. (2001): *The ecology of language evolution*. Cambridge: Cambridge University Press. (cit. on p. 21)

[38] Amano, Tatsuya/Sandel, Brody/Eager, Heidi/Bulteau, Edouard/Svenning, Jens-Christian/Dalsgaard, Bo/Rahbek, Carsten/Davies, Richard G./Sutherland, William J. (2014): Global distribution and drivers of language extinction risk. *Proceedings of the Royal Society of London B: Biological Sciences* 281 (1793): 20141574. (cit. on p. 21)

[39] Sutherland, William J. (2003): Parallel extinction risk and global distribution of languages and species. *Nature* 423 (6937): 276–279. (cit. on p. 21)

[40] Trudgill, Peter (1986): *Dialects in contact*. Oxford: Blackwell. (cit. on p. 21)

[41] Kerswill, Paul (2002): Models of linguistic change and diffusion: new evidence from dialect levelling in British English. *Reading working papers in linguistics* 6 (2): 187–216. (cit. on p. 21)

[42] Britain, David (2002): Space and spatial diffusion. In: Chambers, J. K./Trudgill, Peter/Schilling-Estes, Natalie (eds.), *The Handbook of Language Variation and Change*. Oxford: Blackwell, 603–637. (cit. on p. 21)

[43] Labov, William (2007): Transmission and diffusion. *Language* 83 (2): 344–387. (cit. on p. 21)

[44] Nerbonne, John (2010): Measuring the diffusion of linguistic change. *Phil. Trans. R. Soc. B* 365 (1559): 3821–3828. (cit. on p. 21)

[45] Schmidt, Johannes (1872): *Die Verwandtschaftsverhältnisse der indogermanischen Sprachen*. Weimar: Böhlau. (cit. on p. 21)

[46] Bloomfield, Leonard (1933): *Language*. New York. (cit. on p. 21)

[47] Labov, William (2001): *Principles of Linguistic Change. Vol. 2: Social Factors*. Oxford. (cit. on p. 21)

[48] Murray, James D. (2002): *Mathematical Biology. Volume I.* New York: Springer. (cit. on pp. 23, 24, 33, 74)

[49] Altmann, Gabriel/Meyer, Peter (2005): Physicists look at language. In: Altmann, Gabriel/Levickij, Viktor/Perebyinis, Valentina (eds.), *Problemy kvantitativnoj lingvistiki*. Černivci: RUTA, 42–59. (cit. on pp. 24, 28)

[50] Schmidt, Jürgen Erich/Herrgen, Joachim (2011): *Sprachdynamik. Eine Einführung in die moderne Regionalsprachenforschung*. Berlin: Erich Schmidt. (cit. on p. 24)

[51] Mehrer, Helmut (2007): *Diffusion in solids*. Berlin: Springer. (cit. on pp. 24, 27)

[52] Skellam, John Gordon (1951): Random dispersal in theoretical populations. *Biometrika* 38 (1/2): 196–218. (cit. on p. 24)

[53] Noble, J. V. (1974): Geographic and temporal development of plagues. *Nature* 250 (5469): 726–728. (cit. on p. 24)

[54] Vogl, Gero/Smolik, Manfred/Stadler, Lorenz-Mathias/Leitner, Michael/Essl, Franz/Dullinger, Stefan/Kleinbauer, Ingrid/Peterseil, Johannes (2008): Modelling the spread of ragweed: effects of habitat, climate change and diffusion. *The European Physical Journal Special Topics* 161 (1): 167–173. (cit. on p. 24)

[55] Smolik, Manfred G./Dullinger, Stefan/Essl, Franz/Kleinbauer, Ingrid/Leitner, Michael/Peterseil, Johannes/Stadler, Lorenz-Mathias/Vogl, Gero (2010): Integrating species distribution models and interacting particle systems to predict the spread of an invasive alien plant. *Journal of Biogeography* 37 (3): 411–422. (cit. on p. 24)

[56] Richter, Robert/Berger, Uwe E./Dullinger, Stefan/Essl, Franz/Leitner, Michael/Smith, Matthew/Vogl, Gero (2013): Spread of invasive ragweed: climate change, management and how to reduce allergy costs. *Journal of Applied Ecology* 50 (6): 1422–1430. (cit. on p. 24)

[57] Belik, Vitaly/Geisel, Theo/Brockmann, Dirk (2011): Natural human mobility patterns and spatial spread of infectious diseases. *Physical Review X* 1: 011001. (cit. on p. 24)

[58] Leitner, Michael/Kühn, Ingolf (2018): Dispersal in Plants and Animals. In: Bunde, Armin/Caro, Jürgen/Kärger, Jörg/Vogl, Gero (eds.), *Diffusive Spreading in Nature, Technology and Society*. Berlin: Springer, 29–47. (cit. on p. 24)

[59] Rogers, Everett M. (2003): *Diffusion of innovations*. New York: Free Press. (cit. on p. 24)

[60] Fort, Joaquim (2012): Synthesis between demic and cultural diffusion in the Neolithic transition in Europe. *Proceedings of the National Academy of Sciences* 109 (46): 18669–18673. (cit. on p. 24)

[61] Lemmen, Carsten/Gronenborn, Detlef (2018): The Diffusion of Humans and Cultures in the Course of the Spread of Farming. In: Bunde, Armin/Caro, Jürgen/Kärger, Jörg/Vogl, Gero (eds.), *Diffusive Spreading in Nature, Technology and Society*. Berlin: Springer, 333–349. (cit. on p. 24)

[62] Ember, Carol R./Ember, Melvin (2010): *Cultural Anthropology. Thirteenth Edition. International Edition*. Boston: Pearson. (cit. on p. 25)

[63] Gray, Russell D./Atkinson, Quentin D. (2003): Language-tree divergence times support the Anatolian theory of Indo-European origin. *Nature* 426 (6965): 435–439. (cit. on p. 25)

[64] Bouckaert, Remco/Lemey, Philippe/Dunn, Michael/Greenhill, Simon J./Alekseyenko, Alexander V./Drummond, Alexei J./Gray, Russell D./Suchard, Marc A./Atkinson, Quentin D. (2012): Mapping the origins and expansion of the Indo-European language family. *Science* 337 (6097): 957–960. (cit. on p. 25)

[65] Burridge, James (2017): Spatial evolution of human dialects. *Physical Review X* 7: 031008. (cit. on p. 25)

[66] Burridge, James (2018): Unifying models of dialect spread and extinction using surface tension dynamics. *Royal Society Open Science* 5 (1): 171446. (cit. on p. 25)

[67] Eisenstein, Jacob/O'Connor, Brendan/Smith, Noah A./Xing, Eric P. (2014): Diffusion of lexical change in social media. *PLoS one* 9 (11): e113114. (cit. on p. 25)

[68] List, Johann-Mattis/Pathmanathan, Jananan Sylvestre/Lopez, Philippe/Bapteste, Eric (2016): Unity and disunity in evolutionary sciences: process-based analogies open common research avenues for biology and linguistics. *Biology Direct* 11: 39. (cit. on p. 27)

[69] Heitjans, Paul/Kärger, Jörg (eds.), (2005): *Diffusion in condensed matter: methods, materials, models*. Berlin: Springer Science & Business Media. (cit. on p. 27)

[70] Hartmann, Stephan (2005): *The World as a Process: Simulations in the Natural and Social Sciences*. http://philsci-archive.pitt.edu/2412/ (last accessed January 31, 2019). (cit. on p. 29)

[71] Castellano, Claudio/Fortunato, Santo/Loreto, Vittorio (2009): Statistical physics of social dynamics. *Reviews of Modern Physics* 81 (2): 591. (cit. on p. 32)

[72] Carro, Adrián/Toral, Raúl/San Miguel, Maxi (2016): Coupled dynamics of node and link states in complex networks: A model for language competition. *New Journal of Physics* 18 (11): 113056. (cit. on p. 32)

[73] Chapel, Laetitia/Castelló, Xavier/Bernard, Claire/Deffuant, Guillaume/Eguíluz, Víctor M/Martin, Sophie/San Miguel, Maxi (2010): Viability and resilience of languages in competition. *PLoS ONE* 5 (1): e8681. (cit. on p. 32)

[74] Patriarca, Marco/Castelló, Xavier/Uriarte, Jose-Ramon/Eguíluz, Víctor M/San Miguel, Maxi (2012): Modeling two-language competition dynamics. *Advances in Complex Systems* 15 (3/4): 1250048. (cit. on p. 32)

[75] Solé, Ricard V./Corominas-Murtra, Bernat/Fortuny, Jordi (2010): Diversity, competition, extinction: the ecophysics of language change. *Journal of The Royal Society Interface* 7 (53): 1647–1664. (cit. on p. 32)

[76] Kandler, Anne (2009): Demography and language competition. *Human Biology* 81 (2-3): 181–210. (cit. on p. 32)

[77] Kandler, Anne/Unger, Roman/Steele, James (2010): Language shift, bilingualism and the future of Britain's Celtic languages. *Philosophical Transactions of the Royal Society of London B: Biological Sciences* 365 (1559): 3855–3864. (cit. on pp. 32, 37)

[78] Isern, Neus/Fort, Joaquim (2014): Language extinction and linguistic fronts. *Journal of The Royal Society Interface* 11 (94): 20140028. (cit. on p. 32)

[79] Walters, Caroline E. (2014): A reaction–diffusion model for competing languages. *Meccanica* 49 (9): 2189–2206. (cit. on pp. 32, 33)

[80] Patriarca, Marco/Leppänen, Teemu (2004): Modeling language competition. *Physica A: Statistical Mechanics and its Applications* 338 (1): 296–299. (cit. on p. 32)

[81] Abrams, Daniel M./Strogatz, Steven H. (2003): Linguistics: Modelling the dynamics of language death. *Nature* 424 (6951): 900–900. (cit. on pp. 32, 37)

[82] Minett, James W./Wang, William S.-Y. (2008): Modelling endangered languages: The effects of bilingualism and social structure. *Lingua* 118 (1): 19–45. (cit. on p. 32)

[83] Mira, Jorge/Paredes, Ángel (2005): Interlinguistic similarity and language death dynamics. *Europhysics Letters* 69 (6): 1031. (cit. on p. 32)

[84] Wyburn, John/Hayward, John (2008): The future of bilingualism: an application of the Baggs and Freedman model. *Journal of Mathematical Sociology* 32 (4): 267–284. (cit. on p. 32)

[85] Zhang, Menghan/Gong, Tao (2013): Principles of parametric estimation in modeling language competition. *Proceedings of the National Academy of Sciences* 110 (24): 9698–9703. (cit. on p. 32)

[86] Kandler, Anne/Steele, James (2008): Ecological models of language competition. *Biological Theory* 3 (2): 164–173. (cit. on p. 33)

[87] Bacaër, Nicolas (2011): *A Short History of Mathematical Population Dynamics*. London: Springer. (cit. on p. 33)

[88] Schulze, Christian/Stauffer, Dietrich/Wichmann, Søren (2007): Birth, survival and death of languages by Monte Carlo simulation. *Communications in Computational Physics* 3 (2): 271–294. (cit. on p. 34)

[89] Kretzschmar, William A./Juuso, Ilkka (2014): Cellular Automata for Modeling Language Change. In: Wąs, Jarosław/Sirakoulis, Georgios Ch./Bandini, Stefania (eds.), *Proceedings of the 11th International Conference on Cellular Automata for Research and Industry, ACRI 2014*. Cham: Springer, 339–348. (cit. on p. 34)

[90] Beltran, Francesc S./Herrando, Salvador/Ferreres, Doris/Adell, Marc-Antoni/Estreder, Violant/Ruiz-Soler, Marcos (2009): Forecasting a language shift based on cellular automata. *Journal of Artificial Societies and Social Simulation* 12 (3): 5. (cit. on p. 34)

[91] de Oliveira, Viviane M/Gomes, Marcelo AF/Tsang, Ing Ren (2006): Theoretical model for the evolution of the linguistic diversity. *Physica A: Statistical Mechanics and its Applications* 361 (1): 361–370. (cit. on p. 34)

[92] Lipowska, Dorota/Lipowski, Adam (2017): Language competition in a population of migrating agents. *Physical Review E* 95 (5): 052308. (cit. on p. 34)

[93] Wyburn, John/Hayward, John (2019): An application of an analogue of the partition function to the evolution of diglossia. *Physica A: Statistical Mechanics and its Applications* 516: 447–463. (cit. on p. 34)

[94] Hegselmann, Rainer (1996): Understanding Social Dynamics: The Cellular Automata Approach. In: Troitzsch, Klaus/Mueller, Ulrich/Gilbert, Nigel/Doran, Jim E. (eds.), *Social Science Microsimulation*. Berlin: Springer, 282–306. (cit. on p. 34)

[95] Gilbert, Nigel (2008): *Agent-based models*. Thousand Oaks: Sage. (cit. on p. 34)

[96] Yang, Xin-She/Young, Y. (2006): Cellular Automata, PDEs, and Pattern Formation. In: Olariu, Stephan/Zomaya, Albert Y. (eds.), *Handbook of Bioinspired Algorithms and Applications*. Boca Raton: Chapman & Hall/CRC, 273–284. (cit. on p. 35)

[97] Hornsby, Michael (2015): *Revitalizing Minority Languages: New Speakers of Breton, Yiddish and Lemko*. Basingstoke: Palgrave Macmillan. (cit. on p. 39)

[98] Labov, William (1994): *Principles of Linguistic Change. Vol. 1: Internal Factors*. Oxford. (cit. on p. 39)

[99] Knooihuizen, Remco (2006): Language shift, ethnolinguistic vitality and historical sociolinguistics: testing the models. In: *Proceedings of the Postgraduate Conference 2006, Lingustics & English Language*. Edinburgh: University of Edinburgh. http://www.lel.ed.ac.uk/~pgc/archive/2006/2006knooihuizen.pdf (last accessed January 31, 2019). (cit. on p. 40)

[100] Gorter, Durk (2013): Linguistic Landscapes in a Multilingual World. *Annual Review of Applied Linguistics* 33: 190–212. (cit. on p. 40)

[101] Kühnel, Ferdinand (2015): The Silent Disappearance of Ethnic Minorities from Gravestones: Ethnic Homogenization in Carinthia's Bilingual Areas since 1918. In: Wakounig, Marija/Kühnel, Ferdinand (eds.), *Central Europe (Re-)visited. A Multi-Perspective Approach to a Region*. Europa Orientalis 17. Wien/Berlin: LIT Verlag, 199–212. (cit. on p. 40)

[102] Göderle, Wolfgang (2016b): *Zensus und Ethnizität: zur Herstellung von Wissen über soziale Wirklichkeiten im Habsburgerreich zwischen 1848 und 1910*. Göttingen: Wallstein Verlag. (cit. on p. 41)

[103] de Vries, John (2005): Language Censuses. In: Ammon, Ulrich/Dittmar, Norbert/Mattheier, Klaus J./Trudgill, Peter (eds.), *Sociolinguistics. An International Handbook of the Science of Lan-

guage and Society. Handbücher zur Sprach- und Kommunikationswissenschaft / Handbooks of Linguistics and Communication Science (HSK) 3.2. Berlin: Walter de Gruyter, 1104–1116. (cit. on pp. 41, 42)

[104] de Vries, John (1990): On coming to our census: A layman's guide to demolinguistics. *Journal of Multilingual & Multicultural Development* 11 (1-2): 57–76. (cit. on pp. 41, 44)

[105] de Vries, John (1996): Language demography. In: Goebl, Hans/Nelde, Peter H./Starý, Zdeněk/Wölck, Wolfgang (eds.), *Contact Linguistics. An International Handbook of Contemporary Research*. Handbücher zur Sprach- und Kommunikationswissenschaft / Handbooks of Linguistics and Communication Science (HSK) 12.1. Berlin: Walter de Gruyter, 771–776. (cit. on p. 41)

[106] Lieberson, Stanley (1966): Language questions in censuses. *Sociological Inquiry* 36 (2): 262–279. (cit. on p. 41)

[107] Mackey, William/Cartwright, Donald (1979): Geocoding language loss from census data. In: Mackey, William/Ornstein, Jacob (eds.), *Sociolinguistic Studies in Language Contact: Methods and Cases*. The Hague: Mouton, 69–96. (cit. on pp. 42–44, 47, 48)

[108] Duchêne, Alexandre/Humbert, Philippe N./Coray, Renata (2018): How to ask questions on language? Ideological struggles in the making of a state survey. *International Journal of the Sociology of Language* 252: 45–72. (cit. on p. 43)

[109] Busch, Brigitta (2015): Über das Kategorisieren von Sprachen und Sprecher_innen. Zur Dekonstruktion von Sprachstatistiken. In: Thoma, Nadja/Knappik, Magdalena (eds.), *Sprache und Bildung in Migrationsgesellschaften*. Bielefeld: transcript, 45–68. (cit. on p. 43)

[110] Busch, Brigitta (2016): *Categorizing languages and speakers: Why linguists should mistrust census data and statistics*. Working Papers in Urban Language & Literacies. http://heteroglossia.net/fileadmin/user_upload/publication/WP189_Busch_2016._Categorizing_languages.pdf (last accessed January 31, 2019). (cit. on p. 43)

[111] Statistik Austria (2007): *Volkszählung 2001. Benutzerhandbuch. Arbeitsbehelf.* Wien: Statistik Austria. (cit. on p. 43)

[112] Brix, Emil (1981): Die Kärntner Volksabstimmmung von 1920 im Kontext der österreichischen Nationalitätenstatistik 1880–1934. In: Rumpler, Helmut (ed.), *Kärntens Volksabstimmung 1920: wissenschaftliche Kontroversen und historisch-politische Diskussionen anläßlich des internationalen Symposions Klagenfurt 1980*. Klagenfurt: Kärntner Druck-und Verlagsgesellschaft, 233–253. (cit. on pp. 43, 89)

[113] Brix, Emil (1982): *Die Umgangssprachen in Altösterreich zwischen Agitation und Assimilation. Die Sprachenstatistik in den zisleithanischen Volkszählungen 1880–1910*. Wien: Böhlau. (cit. on p. 44)

[114] Duchêne, Alexandre/Humbert, Philippe N. (2018): Surveying languages: the art of governing speakers with numbers. *International Journal of the Sociology of Language* 252: 1–20. (cit. on p. 45)

[115] Unkart, Ralf/Glantschnig, Gerold/Ogris, Alfred (1984): *Zur Lage der Slowenen in Kärnten. Die slowenische Volksgruppe und die Wahlkreiseinteilung 1979 – eine Dokumentation*. Das Kärntner Landesarchiv 11. Klagenfurt: Verlag des Kärntner Landesarchivs. (cit. on pp. 45, 49, 72)

[116] Inzko, Valentin/Lukan, Walter/Moritsch, Andreas (1988): *Geschichte der Kärntner Slowenen: von 1918 bis zur Gegenwart unter Berücksichtigung der gesamtslowenischen Geschichte*. Klagenfurt: Hermagoras. (cit. on pp. 45, 46, 49, 53, 69, 95)

[117] Pan, Christoph (2011): Minderheiten als Mehrwert: Volksgruppenzählung in Europa, Italien und Südtirol. In: Karpf, Peter/Kassl, Thomas/Platzer, Werner/Puschnig, Udo (eds.), *Zählen Minder-*

heiten? – Volksgruppen zählen! Kärnten-Dokumentation Band 27. Klagenfurt: Amt der Kärntner Landesregierung, Volksgruppenbüro, 116–128. (cit. on p. 45)

[118] Horowitz, Donald L. (2001): *Ethnic Groups in Conflict*. Berkeley: Univ. of California Press. (cit. on p. 46)

[119] Ibounig, Peter (1986): *Die Kärntner Slowenen im Spiegel der Volkszählung 1981*. Kärnten-Dokumentation Band 2. Klagenfurt: Amt d. Kärntner Landesregierung. (cit. on pp. 47, 48)

[120] Pohl, Heinz-Dieter (2004): Sprache und Politik, gezeigt am Glottonym *Windisch*. In: Krisch, Thomas/Lindner, Thomas/Müller, Ulrich (eds.), *Analecta homini universali dicata, Festschrift für Oswald Panagl zum 65. Geburtstag*. Stuttgart: Verlag Hans Dieter Heinz, 625–636. (cit. on pp. 47, 72)

[121] Priestly, Tom (1997): On the development of the 'Windischentheorie'. *International Journal of the Sociology of Language* 124: 75–98. (cit. on pp. 47, 72)

[122] Winkler, Wilhelm (1923): *Die Bedeutung der Statistik für den Schutz der nationalen Minderheiten*. Leipzig/Wien: Deuticke. (cit. on p. 47)

[123] Exner, Gudrun/Kytir, Josef/Pinwinkler, Alexander (2004): *Bevölkerungswissenschaft in Österreich in der Zwischenkriegszeit (1918–1938): Personen, Institutionen, Diskurse*. Schriften des Instituts für Demographie der Österreichischen Akademie der Wissenschaften 18. Wien: Böhlau. (cit. on p. 47)

[124] Ibounig, Peter (2011): Volksgruppen zählen – mittels Volkszählungen? Die Angaben zur Umgangssprache in den österreichischen Volkszählungen. In: Karpf, Peter/Kassl, Thomas/Platzer, Werner/Puschnig, Udo (eds.), *Zählen Minderheiten? – Volksgruppen zählen!* Kärnten-Dokumentation Band 27. Klagenfurt: Amt der Kärntner Landesregierung, Volksgruppenbüro, 63–73. (cit. on pp. 48, 69)

[125] Boysen, Thede (2011): Die Vermessung von Volksgruppen: Menetekel oder positive Diskriminierung? Nationale Minderheiten in Deutschland zwischen Bekenntnisprinzip und Social Engineering. In: Karpf, Peter/Kassl, Thomas/Platzer, Werner/Puschnig, Udo (eds.), *Zählen Minderheiten? – Volksgruppen zählen!* Kärnten-Dokumentation Band 27. Klagenfurt: Amt der Kärntner Landesregierung, Volksgruppenbüro, 7–12. (cit. on p. 49)

[126] Enotna Lista (o. J.): *Die Ortstafelfrage. Fragen und Antworten zur Ortstafelerkenntnis*. http://www.elnet.at/dossier/fragen_und_antworten_zur_ortstafelerkenntnis/56 (last accessed January 31, 2019). (cit. on p. 49)

[127] Tóth, Ágnes/Vékás, János (2014): National and ethnic minorities in Hungary in the period 2001–2011 – Ethno-demographic trends as reflected in the census data. *Hungarian Statistical Review* 17: 95–112. (cit. on pp. 50, 52, 102)

[128] Göderle, Wolfgang (2016a): Administration, Science, and the State: The 1869 Population Census in Austria-Hungary. *Austrian History Yearbook* 47: 61–88. (cit. on p. 50)

[129] Teibenbacher, Peter/Kramer, Diether/Göderle, Wolfgang (2012): *An Inventory of Austrian Census Materials, 1857–1910. Final Report*. Mosaic Working Paper WP2012-007. http://censusmosaic.org/sites/default/files/downloads/publications/mosaicWP/MOSAIC-WP-2012-007.pdf (last accessed January 31, 2019). (cit. on p. 50)

[130] Ladstätter, Johannes (1973): Wandel der Erhebungs- und Aufarbeitungsziele der Volkszählungen seit 1869. In: Helczmanovszki, Heimold (ed.), *Beiträge zur Bevölkerungs- und Sozialgeschichte*

Österreichs. Nebst einem Überblick über die Entwicklung der Bevölkerungs- und Sozialstatistik. Wien: Verlag für Geschichte und Politik, 267–294. (cit. on p. 50)

[131] Anonymous (1881): *Lehrer-Kalender und Schematismus des Sämmtlichen Lehrpersonales der Volksschulen in Kärnten*. Klagenfurt. (cit. on pp. 51, 93, 96)

[132] Katholische Kirche Diözese Gurk (1880): *Geistlicher Personalstand der Diözese Gurk im Jahre 1880*. Klagenfurt. (cit. on pp. 51, 93)

[133] K.K. Statistische Central-Commission (1883): *Special-Orts-Repertorien der im Oesterreichischen Reichsrathe vertretenen Königreiche und Länder. V. Kärnten*. Wien: Alfred Hölder. (cit. on pp. 51, 123)

[134] K.K. Statistische Central-Commission (1894): *Special-Orts-Repertorien der im Österreichischen Reichsrathe vertretenen Königreiche und Länder. Neubearbeitung auf Grund der Ergebnisse der Volkszählung vom 31. Dezember 1890. V. Kärnten*. Wien: Alfred Hölder. (cit. on pp. 51, 59, 60, 123)

[135] K.K. Statistische Zentralkommission (1905): *Gemeindelexikon der im Reichsrate vertretenen Königreiche und Länder, Bearbeitet auf Grund der Ergebnisse der Volkszählung vom 31. Dezember 1900. V. Kärnten*. Wien: k.k. Hof- und Staatsdruckerei. (cit. on pp. 51, 123)

[136] Statistische Zentralkommission (1918): *Spezialortsrepertorium der Österreichischen Länder. Bearbeitet auf Grund der Ergebnisse der Volkszählung vom 31. Dezember 1910. V. Kärnten*. Wien: Verlag der Staatsdruckerei. (cit. on pp. 51, 124)

[137] Seewann, Gerhard (2012b): *Geschichte der Deutschen in Ungarn. Band 2: 1860 bis 2006*. Marburg: Herder-Institut. (cit. on pp. 53, 62, 68, 100, 112)

[138] Rumpler, Helmut/Seger, Martin (eds.), (2010): *Die Habsburgermonarchie 1848–1918, Band IX/2: Soziale Strukturen. Die Gesellschaft der Habsburgermonarchie im Kartenbild. Verwaltungs-, Sozial- und Infrastrukturen. Nach dem Zensus von 1910*. Wien: Verlag der ÖAW. (cit. on pp. 54, 57, 72, 104, 124)

[139] Max Planck Institute for Demographic Research/Chair for Geodesy and Geoinformatics, University of Rostock (2012): *MPIDR Population History GIS Collection (slightly modified version of a GIS-File by Rumpler and Seger 2010)*. http://censusmosaic.org/data/historical-gis-files (last accessed January 31, 2019). (cit. on pp. 54, 57, 72, 104, 124)

[140] Rumpler, Helmut/Urbanitsch, Peter (eds.), (2000a): *Die Habsburgermonarchie 1848–1918. Band VII: Verfassung und Parlamentarismus. 1. Teilband: Verfassungsrecht, Verfassungswirklichkeit, zentrale Repräsentativkörperschaften*. Wien: Verlag der Österr. Akademie der Wissenschaften. (cit. on p. 56)

[141] Rumpler, Helmut/Urbanitsch, Peter (eds.), (2000b): *Die Habsburgermonarchie 1848–1918. Band VII: Verfassung und Parlamentarismus. 2. Teilband: Die regionalen Repräsentativkörperschaften*. Wien: Verlag der Österr. Akademie der Wissenschaften. (cit. on p. 56)

[142] Haslinger, Peter (2012): How to Run a Multilingual Society: Statehood, Administration and Regional Dynamics in Austria-Hungary, 1867–1914. In: Augusteijn, Joost/Storm, Eric (eds.), *Region and State in Nineteenth-Century Europe. Nation-Building, Regional Identities and Separatism*. Basingstoke: Palgrave Macmillan, 111–128. (cit. on pp. 56, 62)

[143] Wikimedia Commons (2017b): *File:Austria-Hungary1899.JPG*. https://commons.wikimedia.org/w/index.php?title=File:Austria-Hungary1899.JPG&oldid=247966342 (last accessed January 31, 2019). (cit. on p. 57)

[144] Urbanitsch, Peter (1980): Erläuterungen zur Sprachen- bzw. Nationalitätenkarte der österreichisch-ungarischen Monarchie. In: Wandruszka, Adam/Urbanitsch, Peter (eds.), *Die Habsburgermonarchie 1848–1918. Band III. Die Völker des Reiches. 2. Teilband*. Wien: Verlag der ÖAW, explanatory supplement included with the book. (cit. on p. 57)

[145] Jernudd, Björn/Nekvapil, Jiří (2012): History of the field: a sketch. In: Spolsky, Bernard (ed.), *The Cambridge Handbook of Language Policy*. Cambridge: Cambridge University Press, 16–36. (cit. on p. 61)

[146] Johnson, David Cassels (2013): *Language Policy*. Basingstoke: Palgrave Macmillan. (cit. on p. 61)

[147] Spolsky, Bernard (2012): What is language policy? In: Spolsky, Bernard (ed.), *The Cambridge Handbook of Language Policy*. Cambridge: Cambridge University Press, 3–15. (cit. on p. 61)

[148] Wikimedia Commons (2017a): *File:Austria Hungary ethnic.svg*. https://commons.wikimedia.org/w/index.php?title=File:Austria_Hungary_ethnic.svg&oldid=271592294 (last accessed January 31, 2019). (cit. on p. 62)

[149] Glück, Helmut/Rödel, Michael (eds.), (2016): *Metzler Lexikon Sprache*. Stuttgart: J.B. Metzler Verlag. (cit. on p. 62)

[150] Haslinger, Peter (2008): Sprachenpolitik, Sprachendynamik und imperiale Herrschaft in der Habsburgermonarchie 1740–1914. *Zeitschrift für Ostmitteleuropa-Forschung* 57 (1): 81–111. (cit. on p. 62)

[151] Marácz, László (2012): Multilingualism in the Transleithanian part of the Austro-Hungarian Empire (1867–1918): Policy and practice. *Jezikoslovlje* 13 (2): 269–298. (cit. on p. 62)

[152] Stourzh, Gerald (1980): Die Gleichberechtigung der Volksstämme als Verfassungsprinzip 1848–1918. In: Wandruszka, Adam/Urbanitsch, Peter (eds.), *Die Habsburgermonarchie 1848–1918. Band III. Die Völker des Reiches. 2. Teilband*. Wien: Verlag der ÖAW, 975–1206. (cit. on p. 62)

[153] Kardum, Lidia/Gajski, Bruno Korea (2012): Language as a political issue: The case of language of command in the Austro-Hungarian army. *Jezikoslovlje* 13 (2): 351–371. (cit. on p. 62)

[154] Newerkla, Stefan Michael (2003): The seamy side of the Habsburgs' liberal language policy: Intended and factual reality of language use in Plzen's educational system. In: Rindler Schjerve, Rosita (ed.), *Diglossia and Power: Language Policies and Practice in the 19th Century Habsburg Empire*. Berlin-New York: Mouton de Gruyter, 167–195. (cit. on p. 63)

[155] Tilkovszky, Loránt/Till, Johanna (1991): *Zeitgeschichte der Ungarndeutschen seit 1919: mit einer Vorgeschichte*. Budapest: Corvina. (cit. on p. 63)

[156] Bindorffer, Györgyi (2004): Glaubensleben und Religion der Ungarndeutschen: Verbindender und trennender Katholizismus. *Acta Ethnographica Hungarica* 49 (1–2): 141–150. (cit. on p. 68)

[157] Hutterer, Claus Jürgen (1991b): Hochsprache und Mundart bei den Deutschen in Ungarn. In: Manherz, Karl (ed.), *Aufsätze zur deutschen Dialektologie*. Budapest: Tankönyvkiadó, 313–344. (cit. on p. 68)

[158] Hutterer, Claus Jürgen (1991a): Die deutsche Volksgruppe in Ungarn. In: Manherz, Karl (ed.), *Aufsätze zur deutschen Dialektologie*. Budapest: Tankönyvkiadó, 253–280. (cit. on pp. 68, 100)

[159] Hessky, Regina (1997): Ungarisch – Deutsch. In: Goebl, Hans/Nelde, Peter H./Starý, Zdeněk/Wölck, Wolfgang (eds.), *Contact Linguistics. An International Handbook of Contemporary Research*. Handbücher zur Sprach- und Kommunikationswissenschaft / Handbooks of Linguistics and Communication Science (HSK) 12.2. Berlin: Walter de Gruyter, 1723–1731. (cit. on pp. 68, 100)

[160] Suppan, Arnold (2004): Zur Geschichte Südkärntens. Aus der Perspektive einer zweisprachigen Region. In: Pandel, Martin/Polzer, Miroslav/Polzer-Scrienz, Mirjam/Vospernik, Reginald (eds.), *Ortstafelkonflikt in Kärnten – Krise oder Chance?* Wien: Braumüller, 128–199. (cit. on p. 69)

[161] Busch, Brigitta (2001): Slovenian in Carinthia. In: Extra, Guus/Dorter, Gurk (eds.), *The other languages of Europe*. Clevedon: Multilingual Matters, 119–136. (cit. on p. 69)

[162] Priestly, Tom (2000): Slovene in Austria. In: Wirrer, Jan (ed.), *Minderheiten- und Regionalsprachen in Europa*. Wiesbaden: Westdeutscher Verlag, 222–234. (cit. on p. 69)

[163] Priestly, Tom (1990): Surrender to Symbolic Domination, or Resistance: Patterns of Language Use Among Slovene Speakers in Two Carinthian Communities. *Slovene Studies* 12 (2): 183–203. (cit. on p. 71)

[164] Priestly, Tom (1994): Effects of educational and social mobility on language maintenance, language attitudes and language structure: The case of Sele in Carinthia. *Journal of Multilingual & Multicultural Development* 15 (2-3): 199–217. (cit. on p. 71)

[165] Busch, Brigitta (1996): *Lepena. Ein Dorf macht Schule. Eine Mikrountersuchung sozialer und kultureller Gegensätze*. Klagenfurt: Drava-Verlag. (cit. on p. 71)

[166] Maurer-Lausegger, Herta (2001): Zur soziolinguistischen Situation in der gemischtsprachigen Gemeinde Diex/Djekše in Kärnten: ein historische Streifzug. In: Pittner, Karin/Pittner, Robert J. (eds.), München: Lincom Europa, 109–124. (cit. on p. 71)

[167] Merkač, Franc/Jurić-Pahor, Marija (1984): *O vaškem vsakdanu*. Klagenfurt: Drava. (cit. on p. 71)

[168] Fisher, Ronald Aylmer (1937): The wave of advance of advantageous genes. *Annals of Eugenics* 7 (4): 355–369. (cit. on p. 74)

[169] Williams, Colin H. (2004): The Geography of Language. In: Ammon, Ulrich/Dittmar, Norbert/Mattheier, Klaus J./Trudgill, Peter (eds.), *Sociolinguistics. An International Handbook of the Science of Language and Society*. Handbücher zur Sprach- und Kommunikationswissenschaft / Handbooks of Linguistics and Communication Science (HSK) 3.1. Berlin: Walter de Gruyter, 130–145. (cit. on p. 75)

[170] Crank, John (1975): *The Mathematics of Diffusion*. Oxford: Clarendon. (cit. on p. 82)

[171] Jost, Wilhelm (1960): *Diffusion in Solids, Liquids, Gases*. New York: Academic Press. (cit. on p. 82)

[172] Philibert, Jean (1991): *Atom movements. Diffusion and mass transport in solids*. Les Ulis: Les Éditions de Physique. (cit. on p. 82)

[173] Cushman-Roisin, Benoit (2012): *Course Notes (ENGS 43). Chapter 2 – Diffusion*. http://www.dartmouth.edu/~cushman/courses/engs43/Chapter2.pdf (last accessed January 31, 2019). Dartmouth College. (cit. on p. 82)

[174] Nelder, John A./Mead, Roger (1965): A simplex method for function minimization. *The Computer Journal* 7 (4): 308–313. (cit. on pp. 84, 106)

[175] Fishman, Joshua (1972): *The Sociology of Language*. Rowley: Newbury House. (cit. on p. 92)

[176] Kurz, Maria (1990): *Zur Lage der Slowenen in Kärnten. Der Streit um die Volksschule in Kärnten (1867–1914)*. Klagenfurt: Kärntner Landesarchiv. (cit. on pp. 93, 95, 96)

[177] Veiter, Theodor (1936): *Die Slowenische Volksgruppe in Kärnten. Geschichte, Rechtslage, Problemstellung*. Wien: Reinhold-Verlag. (cit. on p. 93)

[178] Fischer, Gero (1980): *Das Slowenische in Kärnten*. Wien: Kattnig. (cit. on pp. 95, 96)

[179] Klemenčič, Matjaž (2011): The Austrian Policy towards Elementary Bilingual Education in Southern Carinthia and its Changes between 1958 and 1988. *Razprave in Gradivo* 65: 88–113. (cit. on p. 96)

[180] Zupančič, Jernej (2008): The Economic Situation of Carinthian Slovenes After WW II. *Slovene Studies* 30 (2): 235–239. (cit. on p. 96)

[181] Weichselbraun, Anna (2014): 'People here speak five languages!': The reindexicalization of minority language practice among Carinthian Slovenes in Vienna, Austria. *Language in Society* 43 (4): 421–444. (cit. on p. 97)

[182] ORF (2016): *Immer mehr wählen zweisprachigen Unterricht*. http://kaernten.orf.at/news/stories/2801452/ (last accessed January 31, 2019). (cit. on p. 97)

[183] Priestly, Tom/Comanaru, Ruxandra (2009): Identity among the Minority Slovenes of Carinthia, Austria. *Razprave in Gradivo* 58: 6–23. (cit. on p. 97)

[184] Seewann, Gerhard (2012a): *Geschichte der Deutschen in Ungarn. Band 1: Vom Frühmittelalter bis 1860*. Marburg: Herder-Institut. (cit. on p. 100)

[185] Swanson, John C. (2017): *Tangible Belonging: Negotiating Germanness in Twentieth-Century Hungary*. Pittsburgh: University of Pittsburgh Press. (cit. on p. 100)

[186] Manherz, Karl (ed.), (1998): *Die Ungarndeutschen*. Budapest: Press Publica. (cit. on p. 101)

[187] Bartha, Csilla/Borbély, Anna (2006): Dimensions of linguistic otherness: prospects of minority language maintenance in Hungary. *Language Policy* 5 (3): 335–363. (cit. on p. 102)

[188] Az országos magyar kir. statisztikai hivatal (1882): *Az 1881. év elején végrehajtott népszámlálás főbb eredményei megyék és községek szerint részletezve, II. kötet*. Budapest: Pesti Könyvnyomda-Részvény-Társaság. https://library.hungaricana.hu/en/view/NEDA_1881_02/ (last accessed January 31, 2019). (cit. on pp. 103, 124)

[189] Az országos m. kir. statisztikai hivatal (1892): *A Magyar Korona országainak helységnévtára a nagyméltóságú magy. kir. kereskedelemügyi minister úr rendeletéből, szerkeszti Dr Jekelefalussy József*. Budapest: Pesti Könyvnyomda-Részvény-Társaság. https://library.hungaricana.hu/en/view/NEDA_1892_helysegnevtar/ (last accessed January 31, 2019). (cit. on pp. 103, 124)

[190] A magyar kir. központi statisztikai hivatal (1902): *A magyar korona országainak 1900. évi népszámlálása. A népesség általános leírása községenkint*. Budapest: Pesti Könyvnyomda-Részvény-Társaság. https://library.hungaricana.hu/en/view/NEDA_1900_01/ (last accessed January 31, 2019). (cit. on pp. 103, 124)

[191] A magyar kir. központi statisztikai hivatal (1912): *A magyar szent korona országainak 1910. évi népszámlálása. A népesség főbb adatai községek és népesebb puszták, telepek szerint*. Budapest: Az athenaeum irodalmi és nyomdai r.-társulat nyomása. https://library.hungaricana.hu/en/view/NEDA_1910_01/ (last accessed January 31, 2019). (cit. on pp. 103, 124)

[192] A magyar kir. központi statisztikai hivatal (1923): *Az 1920. évi népszámlálás. A népesség főbb demográfiai adatai. Községek és népesebb puszták, telepek szerint*. Budapest: Pesti Könyvnyomda-Részvény-Társaság. https://library.hungaricana.hu/en/view/NEDA_1920_01/ (last accessed January 31, 2019). (cit. on pp. 103, 124)

[193] A magyar kir. központi statisztikai hivatal (1932): *Az 1930. évi népszámlálás. Demográfiai adatok községek és külterületi lakotthelyek szerint*. Budapest: Stephaneum nyomda részvénytársaság. https://library.hungaricana.hu/en/view/NEDA_1930_01/ (last accessed January 31, 2019). (cit. on pp. 103, 124)

[194] Központi statisztikai hivatal és dokumentációs szolgálat magyar országos levéltár (1976): *Az. 1941. évi népszámlálás. 2. Demográfiai adatok községek szerint*. Budapest: Történeti statisztikai kötetek. https://library.hungaricana.hu/hu/view/NEDA_1941_02/ (last accessed January 31, 2019). (cit. on pp. 103, 123)

[195] Központi statistzikai hivatal / Hungarian Central Statistical Office (2004a): *Népszámlálás / Population Census 2001. Baranya county. Retrospective data. Table 1.1.9 Population by mother tongue, nationality and sex, 1930, 1960–2001.* http://www.nepszamlalas2001.hu/eng/volumes/06/02/tabeng/1/load01_9_0.html (last accessed January 31, 2019). (cit. on pp. 103, 123)

[196] Központi statistzikai hivatal / Hungarian Central Statistical Office (2004b): *Népszámlálás / Population Census 2001. Tolna county. Retrospective data. Table 1.1.9 Population by mother tongue, nationality and sex, 1930, 1960–2001.* http://www.nepszamlalas2001.hu/eng/volumes/06/17/tabeng/1/load01_9_0.html (last accessed January 31, 2019). (cit. on pp. 103, 123)

[197] Központi Statistzikai Hivatal / Hungarian Central Statistical Office (2013a): *Népszámlálás / Population Census 2011. Baranya county. Retrospective data. Table 1.1.6.2 Population by nationality, mother tongue, language spoken with family members or friends and affinity with nationalities' cultural values.* http://www.ksh.hu/nepszamlalas/tables_regional_02 (last accessed January 31, 2019). (cit. on pp. 103, 123)

[198] Központi Statistzikai Hivatal / Hungarian Central Statistical Office (2013b): *Népszámlálás / Population Census 2011. Tolna county. Retrospective data. Table 1.1.6.2 Population by nationality, mother tongue, language spoken with family members or friends and affinity with nationalities' cultural values.* http://www.ksh.hu/nepszamlalas/tables_regional_17 (last accessed January 31, 2019). (cit. on pp. 103, 123)

[199] Gal, Susan (1993): Diversity and contestation in linguistic ideologies: German speakers in Hungary. *Language in Society* 22 (3): 337–359. (cit. on p. 112)

[200] Gal, Susan (1995): Cultural bases of language use among German-speakers in Hungary. *International Journal of the Sociology of Language* 11 (1): 93–102. (cit. on p. 112)

[201] Nelde, Peter Hans (2000): Bilingualism among ethnic Germans in Hungary. In: Wolff, Stefan (ed.), *German Minorities in Europe: Ethnic Identity and Cultural Belonging*. New York: Berghahn Books, 125–134. (cit. on p. 112)

[202] Erb, Maria (2010): Sprachgebrauch der Ungarndeutschen: Geschichte – Tendenzen – Perspektiven. In: Kostrzewa, Frank/Rada, Roberta V. (eds.), *Deutsch als Fremd-und Minderheitensprache in Ungarn. Historische Entwicklung, aktuelle Tendenzen und Zukunftsperspektiven*. Hohengehren: Schneider, 118–146. (cit. on p. 112)

[203] Bindorffer, Györgyi (1998): No language, no ethnicity? Ethnic identity, language and cultural representation among Hungarian Germans. *Review of Sociology of the Hungarian Sociological Association* 1–2: 143–157. (cit. on p. 112)

[204] Maitz, Péter/Sándor, Klára (2009): Changes in the linguistic marketplace: The case of German in Hungary. In: Carl, Jenny/Stevenson, Patrick (eds.), *Language, Discourse and Identity in Central Europe*. Hampshire: Palgrave Macmillan, 149–164. (cit. on p. 112)

[205] Judson, Pieter M. (2006): *Guardians of the Nation: Activists on the language frontiers of imperial Austria*. Cambridge, Massachusetts: Harvard University Press. (cit. on p. 120)

[206] Zahra, Tara (2010): Imagined Noncommunities: National Indifference as a Category of Analysis. *Slavic Review* 69 (1): 93–119. (cit. on p. 120)

[207] Rumpler, Helmut/Urbanitsch, Peter (eds.), (2006): *Die Habsburgermonarchie 1848–1918. Band VIII: Politische Öffentlichkeit und Zivilgesellschaft. Teil 1: Vereine, Parteien und Interessenverbände als Träger der politischen Partizipation*. Wien: Verlag der Österr. Akademie der Wissenschaften. (cit. on p. 120)

[208] Hechter, Michael (1975): *Internal Colonialism: The Celtic Fringe in Bristish National Development, 1536–1966*. London: Routledge & Kegan. (cit. on p. 120)

List of figures

Figure 1 Percentage of “Slovenian” in census results in southern Carinthia, Austria . . . 11

Figure 2 General principle of modeling and simulation employed in this thesis 30

Figure 3 Overview of censuses in Austria(-Hungary) with data on language use and the way the language-related question is phrased 52

Figure 4 Areas used for data evaluation and modeling in Carinthia and in southern Hungary and their location in Europe 1880 and Europe today 54

Figure 5 Map of Austria-Hungary and its position in Europe 57

Figure 6 A page showing results of the Austrian census 1890 59

Figure 7 A page showing results of the Hungarian census 1880 60

Figure 8 Distribution of ethnic groups of Austria-Hungary in 1910, based on census results on language 62

Figure 9 Percentage of the two language groups (majority and minority language) in the population from 1880 to 1910, calculated from census results 64

Figure 10 Distribution of settlement sizes (number of inhabitants) in Carinthia and southern Hungary 66

Figure 11 Maps of areas of Carinthia and southern Hungary on the same scale with a grid overlay 66

Figure 12 Schematic representation of the settlement structure in Carinthia and southern Hungary 67

Figure 13 Percentage of “Slovenian” in census results in southern Carinthia, Austria . . . 70

Figure 14 Map of Carinthia with its districts (today’s borders). Grey shading roughly indicates the area used for modeling. 72

Figure 15 Creating an abstraction of the geographic area by overlaying a regular grid . . 73

Figure 16 Assigning speaker numbers from individual settlements to grid cells 73

Figure 17 Cells with $\geq$ 50% Slovenian in southern Carinthia for four different years (start and end of each modeling period) 76

Figure 18 Cells with $\geq$ 25% Slovenian in southern Carinthia for four different years (start and end of each modeling period) 77

Figure 19 Cells with $\geq$ 17.5% Slovenian in southern Carinthia for four different years (start and end of each modeling period) 77

Figure 20 Percentage of Slovenian speakers in 1910 and schematic of the language border for 1880 and 1910 79

Figure 21 Schematic representation of how speaker numbers in the following year are calculated 81

Figure 22 Measures of the model's reliability compared with census data: mean absolute error per category in speakers and relative error per category in percentage of speakers 86

Figure 23 Total number of Slovenian speakers in southern Carinthia as estimated by the simulation and according to census data 88

Figure 24 Fraction of Slovenian speakers in southern Carinthia as estimated by the simulation and according to census data 88

Figure 25 Residuals for the number of Slovenian speakers: census data for the last year of each period minus simulated result at the end of each period 90

Figure 26 Increase and decrease in the number of Slovenian speakers in southern Carinthia between 1880 and 1910 91

Figure 27 Increase and decrease in the number of Slovenian speakers in southern Carinthia between 1971 and 2001 94

Figure 28 Percentage of "German" in census results in southern Hungary (counties Baranya and Tolna) 102

Figure 29 Percentage of "German" in census results (mother tongue) for the counties Baranya and Tolna from 1880 until 2011 103

Figure 30 Map of Hungary with its counties (today's borders). Grey shading roughly indicates the area used for modeling. 104

Figure 31 Increase and decrease in the number of German speakers in southern Hungary between 1880 and 1930 109

Figure 32 Residuals for the number of German speakers: census data for 1930 minus simulated result for 1930 110

Figure 33 Development of German speakers in southern Hungary (sum of counties Baranya and Tolna) as estimated by the simulation and according to census data 110

Figure 34 Screenshot of the interactive maps for Carinthia on the project webpage https://dcs.univie.ac.at/languagediffusion/ 122

List of tables

Table 1 A non-exhaustive overview of factors influencing language use and possibly leading to language shift . . . 22

Table 2 Comparison of the Austrian and Hungarian census in Austria-Hungary during the time period 1880–1910 . . . 58

Table 3 Summary of data on the two language contact situations investigated in Austria-Hungary . . . 65

Table 4 Comparison values for the goodness of fit of the baseline model, the interaction model, and the interaction model with habitat, for Slovenian–German language shift in Carinthia . . . 87

Table 5 Fit values for the diffusivities D_α for both time periods . . . 91

Table 6 Comparison values for the goodness of fit of the baseline model and the interaction model, for German–Hungarian language shift in southern Hungary . . 108

Acknowledgements

This thesis presents the results of over three years of work, most of which is obviously my accomplishment. However, I benefitted immensely from the input of others and this work would have been impossible without the support of many amazing people I met along the way.

First and foremost, I would like to extend my sincere gratitude to my team of advisors for taking on this project and me as a student. Most people have one (official) advisor, I had three: meeting Gero Vogl was what initially led to this thesis, and I am glad it did. I learned so much working with Gero owing to his extensive knowledge and experience in and outside physics. His enthusiasm and pragmatic outlook kept this work (and me) going during the inevitable ups and downs of scientific research. I am also grateful for him making numerous phone calls instead of me.

Manzi Glauninger finally got to be my official advisor this time around, and I think he and I would agree that it was a very good choice. He was always available for questions, and made sure to remind me that I am half a linguist after all and should keep the linguistic side of things in mind. I consider myself lucky that he dedicated part of his time to this work and hope to have provided a supplement to his informal studies of quantum physics.

Christoph Dellago completed the team, and I am likewise thankful for his support, his time and his helpful input. Coming from computational physics, he added a third perspective, which ultimately enhanced this work—and my own knowledge.

Thanks also goes to Stephan Puchegger for being my partner in crime. Without exaggeration: I could not have done it without you.

I very much appreciate Herwig Peterlik giving me a place to begin this thesis, and Stefan Michael Newerkla giving me a place to finish it. Likewise, I want to thank my colleagues in both places: the working group *Dynamics of Condensed Systems* at the Faculty of Physics, and the Viennese part of the SFB *German in Austria* (particularly the team of Task Cluster C). Special thanks to my (part-time) office companions Jakob Fellinger and Agnes Kim—after having an office all to myself for most of my PhD time, you both were a welcome change.

I would like to express my appreciation to Sheila Embleton and Jörg Kärger for agreeing to review this thesis. Thank you!

Thank you also to Emmerich Kelih for his encouragement, and to Marija Wakounig and Ferdinand Kühnel for their collaboration and for opening professional doors. I am happy to have met the three of you.

Michael Leitner discussed the Hungarian data with us, which has been a great help and led me to consider new interpretations of the data.

Tamás Tölgyesi provided me with valuable assistance in translating the Hungarian census data.

I am especially grateful to the following people and organizations for their assistance:

- Statistik Austria: Eva-Maria Asamer, Adelheid Bauer, Alois Gehart and Wolfgang Zöllner
- Landesstelle für Statistik, Amt der Kärntner Landesregierung: Peter Ibounig
- Büro des Landeshauptmanns für das Land Kärnten: Leo Murer
- Archiv der Diözese Gurk: Robert Kluger
- Kärntner Landesarchiv: Thomas Zeloth and Wilhelm Wadl
- Universitätsbibliothek Klagenfurt
- Zentrum der Ungarndeutschen an der ELTE Budapest: Maria Erb, Elisabeth Knipf-Komlósi, Karl Manherz, Márta Müller
- Teodor Domej
- Werner Drobesch
- All the other Carinthian people we talked to, them being Carinthian Slovenes or not.

I feel very privileged to have had the opportunity to do this PhD project, so I would like to extend my thanks to everyone who supported making it happen in the first place: the aforementioned people as well as Robin Golser, Ursula Ludacka, Ewa Partyka-Jankowska, Thomas Pichler, Manuel Ross, Markus Stana, Yvonne Simon and Christoph Tietz, and the many other people who believed in me.

This work was financially supported by a uni:docs fellowship from the University of Vienna and the Austrian Science Fund (FWF): Special Research Programme (SFB) F60-G23 "German in Austria (DiÖ): Variation – Contact – Perception" (project F 6005-G23 "German in the context of the other languages in the Habsburg State (19th century) and 2nd Austrian Republic"). Parts of this research were completed as a fellow of the Vienna Doctoral School Physics.

Maja Gušavac is gratefully acknowledged for proof-reading this thesis.

Lastly, thank you to everyone who cheered me on. It was a long and exciting road to get here, and I am glad I did not have to go it alone.

English abstract

Language as part of our culture and identity is constantly changing. Likewise, language use is changing: people sometimes stop using one language and shift to another. What causes such a language shift? This thesis examines this question from a new perspective: through the application of methods from physics (mathematical modeling), it aims to describe the language shift process and give an explanation for why people shift their language. To do so, language shift is interpreted as a spreading process – a new behavior, the use of a particular language, spreads. This linguistic diffusion (spread of languages) and the resulting language shift are explored in analogy to physical diffusion (spread of atoms).

Based on descriptions of physical diffusion, a mathematical model for language shift is developed. The mathematical model uses the microscopic approach of cellular automata to simulate the behavior of individual settlements in the studied area, just like the behavior of individual atoms, and their movement can be simulated. In contrast to macroscopic approaches such as reaction-diffusion equations (differential equations based on the diffusion equation, which consider the concentration of a substance or, in the case of language, the percentage of speakers in the population), the microscopic approach can represent the actual conditions in more detail. As a test for the model, empirical data on language from the census is used—while keeping in mind that it presents self-reported statements and any language survey should be regarded critically.

This work looks at the temporal and spatial changes in language use in two different language contact situations in Austria-Hungary and today's Austria and Hungary: Slovenian-German in Carinthia and German-Hungarian in southern Hungary. In these situations, there is a decline in the use of the minority language in favor of the majority language over time. As a first step, the two situations are compared to explore how local conditions influence possible mathematical models. Then, the influence of different factors on language shift is examined during the actual modeling procedure. The results show that the newly developed methodology is suitable for the description of language shift.

For Carinthia, interaction with other people who speak the same language was identified as the most important factor for language maintenance. Bilingual schools and parish language did not seem to have a noticeable impact. However, a difference in the development in rural vs. urban areas was found. In southern Hungary, the microscopic model did not prove as suitable because in this case, there seem to be additional factors at play (e. g. migration) which are not (yet) included in the model. In addition, differences in the data material (the Austrian census asks for everyday language, the Hungarian one for mother tongue) might play a role.

Deutsche Zusammenfassung

Sprache als Teil unserer Kultur und Identifikationsmerkmal unterliegt einer ständigen Veränderung. Ebenso verändert sich auch der Sprachgebrauch: Menschen hören manchmal auf, eine Sprache zu verwenden, sie wechseln also ihre Sprache. Wieso aber wechseln Menschen ihre Sprache? Wie kommt es zu diesem Sprachwechsel? Diese Fragestellung wird in dieser Arbeit von einer neuen Perspektive beleuchtet: Durch die Anwendung von Methoden aus der Physik (mathematischer Modellierung) sollen der Sprachwechselprozess – wie und wo Menschen ihre Sprache wechseln – beschrieben werden sowie eine Erklärung, warum Menschen ihre Sprache wechseln, versucht werden. Dabei wird Sprachwechsel als Ausbreitungsprozess betrachtet – ein neues Verhalten, der Gebrauch einer bestimmten Sprache, verbreitet sich. In Analogie zur physikalischen Diffusion, der Ausbreitung von Atomen, wird also die linguistische Diffusion, die Ausbreitung von Sprachen, und der daraus resultierende Sprachwechsel erforscht.

Dazu wird ein mathematisches Modell für Sprachwechsel entwickelt. Das Modell verfolgt den mikroskopischen Ansatz der zellulären Automaten, bei dem das Verhalten einzelner Orte im Untersuchungsgebiet ähnlich wie das Verhalten einzelner Atome und ihrer Bewegungen simuliert wird. Im Gegensatz zu makroskopischen Ansätzen wie Reaktions-Diffusions-Gleichungen[94] kann der mikroskopische Ansatz die tatsächlichen Verhältnisse detaillierter abbilden. Als empirische Daten zum Test des Modells werden Volkszählungsdaten herangezogen, wobei immer im Hinterkopf behalten werden sollte, dass Sprachzählungen – wie eben auch die Volkszählung – kritisch betrachtet werden müssen anstatt sie als absolute Tatsache zu akzeptieren.

Die Dissertation betrachtet und modelliert die zeitliche und räumliche Veränderung des Sprachgebrauchs in zwei unterschiedlichen Sprachkontaktsituationen in Österreich-Ungarn bzw. dem heutigen Österreich und Ungarn: Slowenisch–Deutsch in Kärnten und Deutsch–Ungarisch in Südungarn. In diesen Situationen kommt es über die Zeit gesehen zu einem Rückgang der Verwendung der Minderheitensprache zugunsten der Mehrheitssprache. Zunächst wird durch einen Vergleich der zwei Sprachkontaktsituationen analysiert, wie die lokalen Gegebenheiten mögliche mathematische Modelle des Sprachwechsels beeinflussen. Durch die eigentliche Modellierung wird dann der Einfluss verschiedener Randbedingungen auf den Sprachwechsel untersucht. Die Ergebnisse bestätigen, dass die neu entwickelte Methodik geeignet ist, um Sprachwechsel großflächig zu beschreiben.

94 Differentialgleichungen in Anlehnung an die Diffusionsgleichung, welche die Konzentration eines Stoffes bzw. im Fall von Sprache den Prozentsatz von SprecherInnen in der Bevölkerung betrachten

In Kärnten wurde Interaktion mit anderen Menschen, die die gleiche Sprache sprechen, als der wichtigste Faktor für den Spracherhalt identifiziert. Utraquistische (zweisprachige) Schulen sowie die Pfarrsprache zeigten keinen merklichen Einfluss. Es ist jedoch ein Unterscheid zwischen der Entwicklung in urbanen und ländlichen Gegenden ersichtlich. In Südungarn hingegen war das mikroskopische Modell weniger gut für eine Beschreibung geeignet, da dort zusätzliche Faktoren wie z.B. Migration ins Gewicht fallen, die in dem hier entwickelten Modell (noch) nicht enthalten sind. Außerdem spielen potenziell Unterschiede im Datenmaterial (die österreichische Volkszählung fragt nach der Umgangssprache, die ungarische nach der Muttersprache) eine Rolle.