Using physics in linguistic research: Language diffusion in Austria and Hungary

K. Prochazka* and G. Vogl

University of Vienna, Faculty of Physics, Boltzmanngasse 5, 1090 Vienna, Austria
*katharina.prochazka@univie.ac.at

map: *The Times Atlas*, Richard Andree (public domain)

Motivation

Languages are an important part of our culturally diverse world, yet many of today's languages are being used less and less. To find out why, one needs to first understand the dynamics behind this language shift.

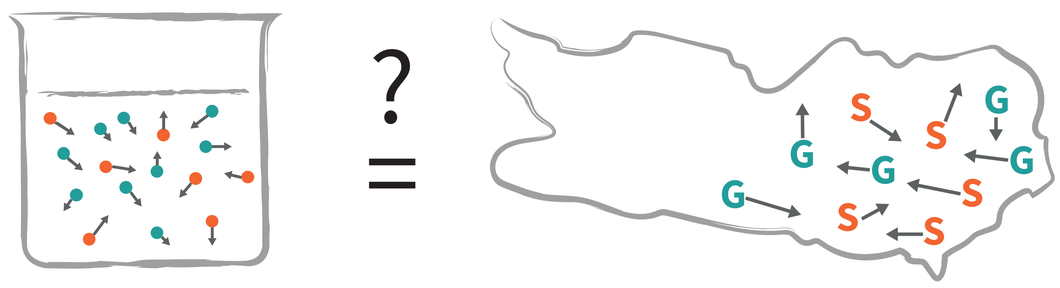
In both Austria and Hungary (the former Austro-Hungarian Empire), use of minority languages has been declining over the past century. Despite supportive measures, language shift (speakers stopping the use of one language for another) is taking place.

One way of monitoring this language shift on a large scale is using methods from the natural sciences where dealing with big sets of data is common.

We present **a microscopic model to follow language shift over time and space** based on data from the census.

The idea: Use physical methods to study language shift

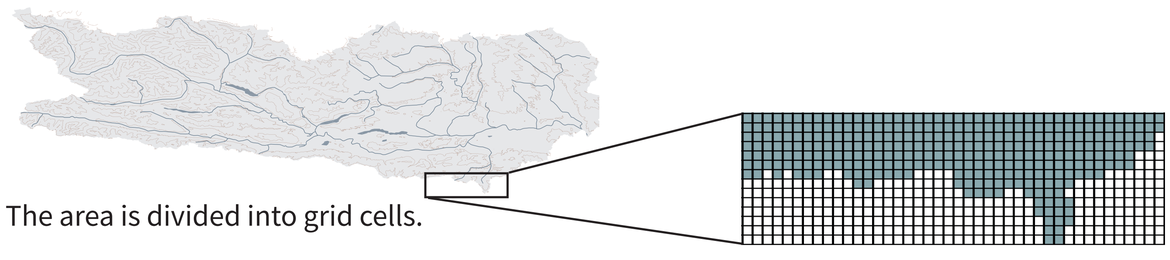
Diffusion of matter ? = **Diffusion of languages**



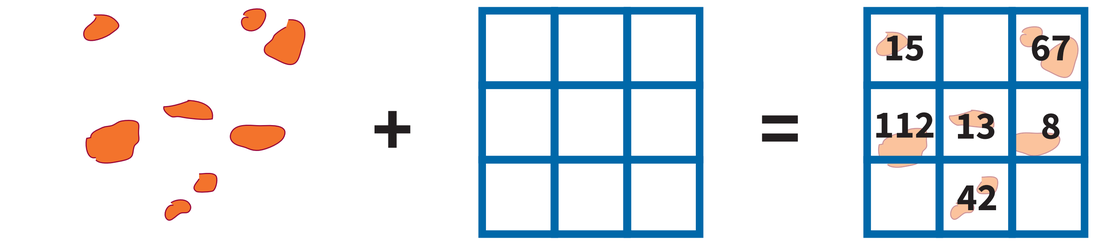
The data

The census was **the primary method for collecting data on languages spoken** in Austria and Hungary. Data was collected **roughly every ten years between 1880 and 2001 (Austria) and between 1881 and 2011 (Hungary)**.

Digitising the census data



The area is divided into grid cells.



Each village (●) is assigned to a grid cell based on its geographic coordinates and speaker numbers from the census are attributed to the grid cells.

Our model: a microscopic approach

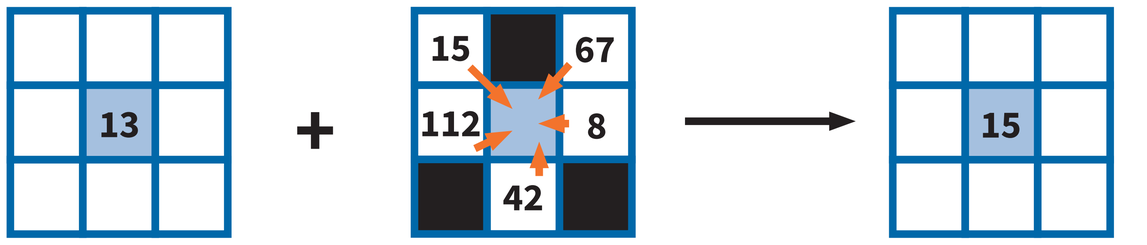
To get the number of speakers of a language α in the next year, each grid cell (□) is updated according to the mathematical rule:

$$n_\alpha(\mathbf{r}, t{+}1) = n_{\text{total}}(\mathbf{r}, t{+}1) \cdot \frac{n_\alpha(\mathbf{r}, t) + F_\alpha(\mathbf{r}, t)}{n_1(\mathbf{r}, t) + F_1(\mathbf{r}, t) + n_2(\mathbf{r}, t) + F_2(\mathbf{r}, t)}$$

this means: the number of speakers in the next year ❸ is proportional to:

❶ the number of current speakers and
❷ the interaction with speakers in other places

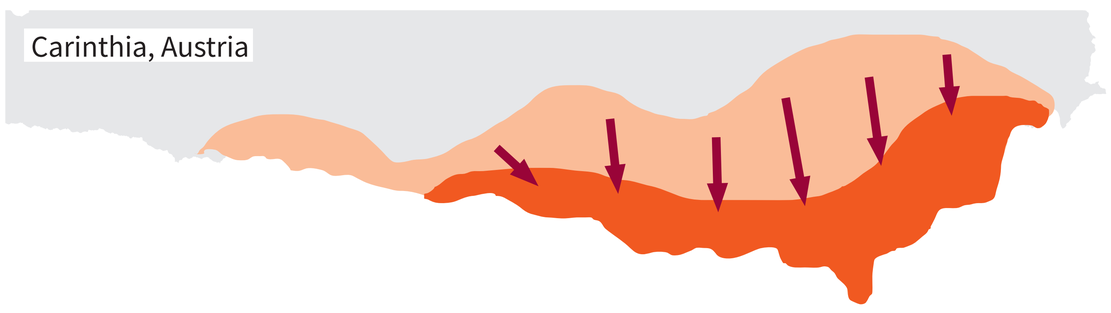
❶ numbers of speakers this year — ❷ interaction with speakers in other places — ❸ numbers of speakers in the next year



Contributions of other cells to the interaction are modelled as Gaussian functions analogous to diffusion models in physics.

Application of the model to different situations

Slovenian in Austria



- the contiguous Slovenian language area is shrinking from one side
  ➞ The **language shift from Slovenian to German can be successfully modelled based on principles of physical diffusion.**
- **Interaction with other speakers of the same language is the most important driving factor for language shift.** External factors such as school or parish language seem to have only a minor influence.

German in Hungary

- dispersed patches of German "language islands" which are separated by larger geographical distances and mostly do not directly interact ➞ The **model needs to be adapted to describe the language shift from German to Hungarian.**
- **Forced assimilation to the Hungarian language through the government *(magyarisation)* is the most important driving factor for language shift.** The language shift process can be seen as shrinking of clusters with only little diffusion between the islands.

Summary

- We have successfully modelled language shift based on principles of physical diffusion.
- **All model parameters can be calculated directly from the census data** ➞ our model is applicable even in situations where data on other factors influencing language use (e. g. status of a language) is either not available or impossible to obtain.
- **Every language shift situation is different and the model must be adapted** to take into account e. g. political decisions which cannot be described as diffusion processes.

http://dcs.univie.ac.at/languagediffusion

K. Prochazka & G. Vogl (2017). *PNAS* 114(17): 4365, DOI: 10.1073/pnas.1617252114.

Acknowledgments
Geographical data for figure backgrounds: Land Kärnten – data.gv.at (license CC BY 3.0 AT) and the GADM database (gadm.org).
Katharina Prochazka is supported by a uni:docs fellowship from the University of Vienna.

dynamics of condensed systems

universität wien