Universität Zürich[UZH]

Geographisches Institut
Geographische Informationssysteme

Dr. Peter Ranacher
peter.ranacher@geo.uzh.ch

Master Topic: Geographical Representation of Language Diversity

Short description
There are two big databases of the languages of the world: Ethnologue (Eberhard et al. 2021), a paywalled publication documenting the language distribution of the world including polygon data, but no academic references; and Glottolog (Hammarström et al. 2020), which started out as academic bibliography about the languages of the world and their classification, which contains point locations for most languages, but no speaker counts or ranges.

For working with world-wide language samples that take into account geography, this situation is deeply dissatisfying. The GIS unit has initiated the "Glottography" project to implement a Web-GIS application that collects published geographical distributions of languages and presents them in a unified manner. Related projects exist for specific parts of the world (Haynie & Gavin 2019; Roose 2021). While the challenges in displaying the raw data (which source claims which language is spoken where) are more on the technical side, the aggregation and weighing of different sources to display a consensus map is an open research topic, which requires various GIS methodology. The following research questions stand out and would make good starting points for Master's theses in GIScience.

What geographic and cultural properties govern the shape of language ranges?
The region where a language is spoken is often not known precisely (Luebbering 2011; Luebbering et al. 2013). Given a rough outline, or point data, on the distribution of a language, can we extract the likely actual range of a language? What other information do we need to do this? One particular method here are voter models, where individuals adapt their language locally to their neighbors (Burridge 2017; Coelho et al. 2021). These kinds of models highlight the importance of coastal patterns and of the spatial distribution of populations. However, rivers and mountain ranges can also be important geographical features for where a language is spoken. Another obvious driver are historical political boundaries (Edwards 2016), which are again likely based on geographical features and (historical) population
densities. Travel effort (Rantanen et al. 2021) seems to be another big driver in more scarcely populated regions of the world. There is little systematic work studying how much effect each of these potential drivers has on the actual outline of a language range.

How can we derive a helpful consensus from different sources?
Because ranges are fuzzy and uncertain, sources differ in the ranges they give for an individual language. Yet on language maps, ranges are usually given as polygons with explicit boundaries or individual point locations. How can we model the uncertainty of language ranges and extract it from a set of partly overlapping ranges from different expert maps? And how can we turn our internal model of fuzzy language ranges back into a single set of language polygons, to show the "true" language range on a map? This research question is related to map crowdsourcing (Budig et al. 2016; Montgomery & Stoeckle 2013), but also to other aggregation methods for large-scale maps eg. of ecoregions (Dinerstein et al. 2017).

What can we infer language ranges of the past?
Touching upon the former two research questions, but going further: Can we infer the spatio-temporal distribution of languages where we miss data? What expansion processes do languages undergo, and how can we use this knowledge to infer the range of a language at some point in the past if we have its present range and a past locations? This question may be approached using more detailed data for specific regions or time frames; alternatively, it could focus on quite abstract general processes such as random walks, which are already used in linguistic phylogeography (Neureiter et al.; Bouckaert et al. 2012).

Further research questions are possible, and we will be happy to accommodate your own ideas. We have a few less developed directions, related to map generalization (in particular removal of language enclaves) for different zoom levels, or multilinguality (both traditional multilinguality, and presence of heritage speakers in metropole regions due to modern global migration). We also have topics in collaboration with the other GIScience groups, if you want to work on visualizing world-wide language

diversity in the web app or on extracting language ranges from text descriptions in sources using named entity recognition with a gazetteer.

Contact

If you are interested, contact

- Dr. Peter Ranacher peter.ranacher@geo.uzh.ch

References

- 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. (doi:10.1126/science.1219669)
- Budig, Benedikt & van Dijk, Thomas C. & Feitsch, Fabian & Arteaga, Mauricio Giraldo. 2016. Polygon consensus: smart crowdsourcing for extracting building footprints from historical maps. Proceedings of the 24th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems, 1–4. Burlingame California: ACM. (doi:10.1145/2996913.2996951) (https://dl.acm.org/doi/10.1145/2996913.2996951) (Accessed November 24, 2021.)
- Burridge, James. 2017. Spatial Evolution of Human Dialects. Physical Review X 7(3). 031008. (doi:10.1103/PhysRevX.7.031008)
- Coelho, Marco Tulio Pacheco & Haynie, Hannah J. & Bowern, Claire & Colwell, Robert K. & Greenhill, Simon J. & Kirby, Kathryn R. & Rangel, Thiago F. & Gavin, Michael C. 2021. Demographic shifts, inter-group contact, and environmental conditions drive language extinction and diversification. SocArXiv. (doi:10.31235/osf.io/xqr2u) (https://osf.io/preprints/socarxiv/xqr2u/) (Accessed August 29, 2021.)
- Dinerstein, Eric & Olson, David & Joshi, Anup & Vynne, Carly & Burgess, Neil D. & Wikramanayake, Eric & Hahn, Nathan et al. 2017. An Ecoregion-Based Approach to Protecting Half the Terrestrial Realm. BioScience. Oxford Academic 67(6). 534-545. (doi:10.1093/biosci/bix014)
- Eberhard, David M. & Simons, Gary F. & Fennig, Charles D. (eds.). 2021. Ethnologue: Languages of the World. Twenty-fourth edition. Dallas, TX: SIL International. (http://www.ethnologue.com)
- Edwards, Owen. 2016. Metathesis and unmetathesis: parallelism and complementarity in Amarasi, Timor. Canberra: Australian National University. (PhD dissertation.)
- Hammarström, Harald & Forkel, Robert & Haspelmath, Martin & Bank, Sebastian. 2020. Glottolog 4.2.1. Jena.(doi:10.5281/zenodo.3754591) (https://glottolog.org/ accessed 2020-08-10)
- Haynie, Hannah J. & Gavin, Michael C. 2019. Modern Language Range Mapping for the Study of Language Diversity. SocArXiv. (doi:10.31235/osf.io/9fu7g) (https://osf.io/preprints/socarxiv/9fu7g/) (Accessed August 29, 2021.)
- Luebbering, Candice R. & Kolivras, Korine N. & Prisley, Stephen P. 2013. The lay of the language: surveying the cartographic characteristics of language maps. Cartography and Geographic Information Science. Taylor & Francis 40(5). 383–400. (doi:10.1080/15230406.2013.809892)
- Luebbering, Candice Rae. 2011. The Cartographic Representation of Language: Understanding language map construction and visualizing language diversity. Virginia Tech. (https://vtechworks.lib.vt.edu/handle/10919/37543) (Accessed December 1, 2021.)
- Montgomery, Chris & Stoeckle, Philipp. 2013. Geographic information systems and perceptual dialectology: a method for processing draw-a-map data. Journal of Linguistic Geography. Cambridge University Press 1(1). 52–85. (doi:10.1017/jlg.2013.4)
- Neureiter, Nico & Ranacher, Peter & van Gijn, Rik & Bickel, Balthasar & Weibel, Robert. Can Bayesian phylogeography reconstruct migrations and expansions in linguistic evolution? Royal Society Open Science. Royal Society 8(1). 201079. (doi:10.1098/rsos.201079)
- Rantanen, Timo & Tolvanen, Harri & Honkola, Terhi & Vesakoski, Outi. 2021. A comprehensive spatial model for historical travel effort - a case study in Finland. Fennia - International Journal of Geography 199(1). (doi:10.11143/fennia.98357) (https://fennia.journal.fi/article/view/98357) (Accessed November 24, 2021.)
- Roose, Meeli. 2021. URHIA. geonode.utu.fi. (https://geonode.utu.fi/maps/74) (Accessed November 24, 2021.)