Correlating morphosyntactic dialect variation with geographic distance: Local beats global
Point of Departure

“First Law of Geography”: “Everything is related to everything else, but near things are more related than distant things.” — Waldo Tobler, 1970

Fundamental Dialectological Postulate: „Geographically proximate varieties tend to be more similar than distant ones.“ — Nerbonne & Kleiweg, 2007

- Tested on different types of data (surveys, corpora)
- Several kinds of geographic distance
- Leading to significant and non-significant correlations
- Local analysis was neglected
The linguistic side – The database

- Syntactic Atlas of German-speaking Switzerland (SADS)
- survey conducted between 2000 and 2002 by the German Dept.
- 383 survey sites, ¼ of the German-speaking municipalities in Switzerland
- over 3000 informants
- 118 survey questions about 50 syntax phenomena
- 3 to 26 informants each survey site from all social and age groups (median 5-6)
- greater inter-subject variation due to selection of multiple informants
The geography side - The distances
The geography side - The distances

- Diverse topography
- Constraints on transport routes – and thus on possible contact paths between speakers
- Quantification of language contact also by travel times
- Euclidean distance
  - Travel times in 1850 (for 120 sites)
  - Travel times by car in 1950
  - Travel times by car in 2000
- Global and regional patterns
The geography side - The distances

- Aim is to prove that investigating on a local level brings more
- Which one is the better predictor – Euclidean or distance or travel times. If the latter, does it count which one we use?
Motivation

Take two survey sites and calculate the proportion of answers variants given to a survey question, calculate their difference. (Normally there are much more then 2 answer variants)

Database

Add more questions into the calculation

Methodology

QI.01. Ich habe zu wenig Kleingeld, um ein Billett zu lösen
I don’t have enough change in order to buy a ticket.

Survey site I. - Klosters 0.66 0.33%
Survey site II. - Flühli 0.2 0.8%
Difference: 0.46 0.53

QI.03. Ich habe den Fritz kommen hören.
I have heard Fritz coming.

Ich ha de Fritz ghöört choo. Ich ha de Fritz choo ghöört.

Survey site I. - Klosters 0.2 0.8%
Survey site II. - Flühli 0.4 0.6%
Difference: 0.2 0.2

\[ \Sigma \text{Dif} : 0.46 + 0.53 + 0.2 + 0.2 = 1.39 \text{ (= Synt. Dist.)} \]

Syntactic Distance Euclidean Distance Travel Time

<table>
<thead>
<tr>
<th>Originals</th>
<th>Normalized</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.39</td>
<td>0.35</td>
</tr>
<tr>
<td>159 km</td>
<td>0.47</td>
</tr>
<tr>
<td>177 min.</td>
<td>0.55</td>
</tr>
</tbody>
</table>
Global result - Correlation Scatterplots

Subset of 60 questions

Syntactic distance

Euclidean Distance

Linear trendline
Second order trendline
Third order trendline
**Correlation Scatterplots**

- 120 survey sites for 1850
- 383 survey sites for 1950 and 2000
- Same language data and syntactic distance as of ~2000
- 60 questions used out of 118
Global result - Average syntactic distance

Average syntactic distances
from the given survey site towards all others

Average synt dist.csv.x

- 20.59 - 33
- 33 - 36
- 36 - 39
- 39 - 42
- 42 - 45
- 45 - 48
- 48 - 51
- 51 - 54
- 54 - 57
- 57 - 61.03

25.09.2014
Resulting correlations

Syntactic distance is based on 60 Questions (out of 118)

Correlations on the global level with the **Syntactic distance**:

<table>
<thead>
<tr>
<th>Type of correlation</th>
<th>Euclidean distance</th>
<th>Travel times by car - 1950</th>
<th>Travel times by car - 2000</th>
<th>Travel times in 1850 *</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pearson’s product-moment correlation coefficients</td>
<td>0.68</td>
<td>0.768</td>
<td>0.777</td>
<td>0.785</td>
</tr>
<tr>
<td>Mantel-test results</td>
<td>0.651</td>
<td>0.766</td>
<td>0.762</td>
<td>0.770</td>
</tr>
<tr>
<td>Explains variance in Synt. distance</td>
<td>46.28%</td>
<td>59.01%</td>
<td>60.4%</td>
<td>61.74%</td>
</tr>
</tbody>
</table>

- Analysis on a local subset
- Obtain residuals that show whether geographic distance predicts linguistic distance

* Only for 120 of the survey sites

25.09.2014
Local Patterns
Correlations on this local subset with the **Syntactic distance**:
Correlations on this local subset with the **Syntactic distance**:  

<table>
<thead>
<tr>
<th>Type of correlation</th>
<th>Euclidean distance</th>
<th>Travel times by car - 1950</th>
<th>Travel times by car - 2000</th>
<th>Travel times in 1850</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pearson correlation coefficients</td>
<td>0.516</td>
<td>0.738</td>
<td>0.702</td>
<td>0.8</td>
</tr>
<tr>
<td>Mantel-test results</td>
<td>0.293</td>
<td>0.492</td>
<td>0.450</td>
<td>0.556</td>
</tr>
<tr>
<td>Explains variance in Synt. distance</td>
<td>26.69%</td>
<td>54.55%</td>
<td>49.4%</td>
<td>63.71%</td>
</tr>
</tbody>
</table>

Source: Google Maps
Comparing geographic distance and syntactic distance

60 questions
Survey site: Obersaxen

Motivation
Database
Methodology
Results
Outlook

25.09.2014
Comparing geographic distance and syntactic distance
Summary

- Calculated a Syntactic Distance based on a survey with multiple informants per survey site
- Compared geographic distances to linguistic difference in Swiss German, travel times for different years, using different stat. methods
- At the **global** level, partly confirmed the Fundamental Dialectological Postulate
- **Local** patterns compared to global patterns reveal a deeper structure and offer potential help for linguists
- ...of course there is a lot more to do
Outlook on the topic

- Inverse weighting for correlating linguistic variables
- Finding other interesting local and linguistic subsets
- Results with pre-1850 historical travel time data
- Other proxies of language contact – linguistic gravity (Szmrecsányi, 2012), commuter matrices
- Relating linguistic variation to geographic features (topography, political or cultural borders etc.)
Thank you for your attention!

Péter Jeszenszky  peter.jeszenszky@geo.uzh.ch, Department of Geography
Robert Weibel   robert.weibel@geo.uzh.ch, Department of Geography


III.22 'Sie ist grösser als ich.'
Komparativ
(Ankreuzfrage)
(27.09.12)
Euclidean Distance & Travel Times

As the crow flies
OBERSAXEN

SCHAFFHAUSEN

FREIBURG

VISP