A Quantitative Approach to Swiss German Dialect Syntax

SCHERRER, Yves, STOECKLE, Philipp

Abstract

In the last decades, dialectometry has emerged as a new field of dialectology. As this kind of research requires large amounts of data, many dialectometric studies used data from “traditional” dialect atlases (e.g. ALF, AIS, RND) which were collected by investigating representatives of the oldest dialects available in the survey locations (i.e. the so-called NORMs, cf. Chambers & Trudgill 2004: 29). Moreover, these data contained mostly lexical and phonological (and sometimes morphological) variables, while syntactic phenomena are largely absent in traditional atlases. In this paper we would like to present results of a dialectometric study that focuses on three aspects which have not been given much attention in previous research. The first aspect concerns the research area, German-speaking Switzerland. Although it is one of the liveliest and at the same time best researched dialect areas in Central Europe, until recently (cf. Goebel et al. 2013, Scherrer & Stoeckle accepted) there have been very few dialectometric studies in this area (cf. Kelle 2001). The second aspect regards the investigated linguistic level: our [...]
A Quantitative Approach to Swiss German Dialect Syntax

International Conference on Language Variation in Europe (ICLaVE 9)

6th June 2017

Yves Scherrer (Helsinki) & Philipp Stoeckle (Vienna)
Introduction

Language change – dialect change – syntactic change

• Focus in variationist linguistics: phonology, lexis
• Syntax largely absent in most dialect atlases
  – Not considered to show much variation
  – Changing more slowly
  – More difficult to detect
• But: in recent years there has been a “boom” in research of dialect syntax
  – “European Dialect Syntax (Edisyn)” (www.dialectsyntax.org)
  – Esp. “Syntactic Atlas of German speaking Switzerland (SADS)” (Bucheli & Glaser 2002)
German-speaking Switzerland

- Four official languages in Switzerland (German, French, Italian, Rhaeto-Romance)
- In the German-speaking part: diglossic situation
- (High-Alemannic) dialects spoken in almost every situation (poly-dialectal communication)
- Use of standard language is limited to writing and formal/official situations
Background & Data

“Syntactic Atlas of German-speaking Switzerland”
(Syntaktischer Atlas der deutschen Schweiz, SADS)
(cf. Bucheli & Glaser 2002)

- Project funded by the SNF 2000-2008 (extended through 2014)
- 4 questionnaires including different syntactic phenomena
- 3187 informants at 383 locations (between 3 and 26 informants at each location)
- Different elicitation techniques
  - Translation
  - Completion
  - Multiple Choice (acceptance and preference)

www.ds.uzh.ch/dialektsyntax/
Results from recent studies (with SADS data)

1. Syntax data show a high degree of variation
2. Variation is not random, but rather seems to be structured with respect to ...
   a) Geography: areal patterns can not only be detected for single phenomena (cf. Glaser 2013 for an overview), but also on an aggregate level (cf. Scherrer & Stoeckle 2016)
   b) Socio-demographics: recent studies show systematic variation (esp. with respect to age groups) which points to apparent time change (cf. Friedli 2012, Stoeckle accepted)
Areal patterns of different aggregate datasets (SADS + SDS; Scherrer & Stoeckle 2016)
Apparent time comparison for single SADS variables
(Stoeckle accepted)
Results from recent studies (with phonological data)

“Regional and Sociolinguistic Variation in Southwest Germany” (cf. Hansen-Morath 2016)

- Dialectometric analyses with phonological data show different geographic patterns for older (60-70) and younger (25-35) speakers

Combined Cluster Analysis (Gabmap; noise factor 0.2; group average + weighted average; cf. Hansen & Stoeckle 2012)
Research questions

• Can correlation between syntactic and socio-demographic variation still be detected on an aggregate level?
• What role do socio-demographic factors play in explaining syntactic variation?
Apparent time comparison

- Subdivide SADS informants into 3 age groups
  - „young“:
    - < 40
    - (500)
  - „middle“:
    - 40-64
    - (1266)
  - „old“:
    - ≥ 65
    - (1004)
Experiments

• Hierarchical cluster analysis
• Correlations between age groups
• Mean differences between age groups
Cluster Analysis

Data matrix → Similarity matrix → Value matrix → Visualisation

Items/Maps → Inquiry points

Euclidean similarity

Hierarchical clustering

A Quantitative Approach to Swiss German Dialect Syntax • Yves Scherrer & Philipp Stoeckle
Cluster Analysis (Ward Algorithm)

Hierarchical Cluster Analysis
Classification into 3 groups
"old" / 2 Clusters
EucledRIW
Ward Algorithm

Hierarchical Cluster Analysis
Classification into 3 groups
"young" / 2 Clusters
EucledRIW
Ward Algorithm
Cluster Analysis (Ward Algorithm)

Hierarchical Cluster Analysis
Classification into 3 groups
"old" / 3 Clusters
EuclidRIW
Ward Algorithm

Hierarchical Cluster Analysis
Classification into 3 groups
"young" / 3 Clusters
EuclidRIW
Ward Algorithm
Cluster Analysis (Ward Algorithm)
Cluster Analysis (Ward Algorithm)

Hierarchical Cluster Analysis
Classification into 3 groups
"old" / 5 Clusters
EuclidRIW
Ward Algorithm

Hierarchical Cluster Analysis
Classification into 3 groups
"young" / 5 Clusters
EuclidRIW
Ward Algorithm

A Quantitative Approach to Swiss German Dialect Syntax • Yves Scherrer & Philipp Stoeckle
Cluster Analysis (Ward Algorithm)

Old: 11 clusters @ 1.0

Young: 6 clusters @ 1.0
Correlations

**Similarity matrices**

![Similarity matrices diagram](image)

**Value matrix**

![Value matrix](image)

**Pearson’s correlation**

![Pearson’s correlation](image)
Correlations

• Measure of homogeneity of answers
• High correlations = low mean differences
However...

- We obtain a very similar result by correlating two random partitions of the SADS dataset...
Correlations

• Northeastern and Southwestern dialects are generally characterized by little syntactic variation and greater homogeneity.

• Interestingly, this variation does not seem to be conditioned on age (or gender, for that matter).

• Can we uncover parts of the variation that are age-related and not generic?
  – Correlations with chance correction.
Chance correction

- Age correlation value – random correlation value
Mean differences between age groups

Data matrices → Value matrix → Visualisation

Items/Maps

Inquiry points

Items/Maps

Value

Mean difference

Color code

Inquiry points

Inquiry points

A Quantitative Approach to Swiss German Dialect Syntax • Yves Scherrer & Philipp Stoeckle
Mean differences between age groups

• Where do survey answers differ most between age groups?
Mean differences between age groups

Mean differences
Young - Old
- 0.15 - 0.23
- 0.24 - 0.27
- 0.28 - 0.33
- 0.34 - 0.40
- 0.41 - 0.57

A Quantitative Approach to Swiss German Dialect Syntax • Yves Scherrer & Philipp Stoeckle
Mean differences between age groups

• Zones with higher-than-average differences lie at the edge between lowlands and alpine dialects

• Compatible with findings from cluster analysis

• Not all variants are expected to show age variation

• Can we find the variants that influence age variation most?
  – Mean differences with weighted variants
Variant weighting

• Mixed effects model to predict age group on the basis of variants, inquiry point as random factor
• Weight data matrices with resulting coefficients
  – Similar in spirit to e.g. Goebl’s GIW
Conclusions

• We investigate age as a socio-demographic factor of syntactic change
  – Apparent time analysis with 3 age groups
• Previous work:
  – Clear age effects in single variants
• Cluster analysis:
  – Alpine dialects show more heterogeneity with older than with younger speakers
  – Convergence of alpine towards lowlands dialects?
• Correlations and mean differences:
  – Some areas of instability are identified, but it is difficult to tease apart age variation from other types of variation
• Variant weighting allows for some strengthening of the (weak) signal
• What about the massive differences in random correlations?
  – Why are some areas much more homogeneous than others?
  – Is this all free variation?
  – Is this type of variation due to a small number of variants? Which ones?