Abstract

In this paper, we report on recent digitization efforts of the linguistic atlas of German-speaking Switzerland (Sprachatlas der deutschen Schweiz, henceforth SDS). The SDS project was initiated in the 1930s, its data collection (in 557 locations) took place in the 1940s and 1950s, and the publication of the resulting eight volumes (totalling about 1500 hand-drawn maps) occurred between 1962 and 1997. While the SDS is often considered to have pioneered the concept of small-scale linguistic atlas (Kleinraumatlas), its achievement before the digital era prevented it from being used for quantitative analyses and computational work, and no digitization efforts had been undertaken until recently. Here, we would like to report on the advancement of our SDS digitization project and of some potential uses of the resulting dataset. Because of time and resource constraints, we had to restrict the SDS data in three ways: (1) selection of a subset of the maps, (2) removal of 8 inquiry points located on Italian territory, (3) removal of rarely used variants in each map. We will discuss the linguistic and dialectological reasonings [...]
Digitizing the linguistic atlas of German-speaking Switzerland

Yves Scherrer\textsuperscript{1}, Sandra Kellerhals\textsuperscript{2}

\textsuperscript{1}Department of Linguistics / Computer Science Center, University of Geneva
\textsuperscript{2}Department of Geography, University of Zurich

Methods in Dialectology XV, Groningen
Outline

1. Introduction
2. Digitizing SDS maps
3. Map selection
4. Content selection
5. The digitization process
6. Access and applications
The linguistic landscape of Switzerland

Diglossia:
Standard German / Swiss German
French
Italian
Romansh

http://www.bfs.admin.ch/bfs/portal/de/index/regionen/thematische_karten/maps/bevoelkerung/sprachen_religionen.html
The linguistic landscape of Switzerland

Diglossia:
Standard German / Swiss German

This talk focuses on the dialects spoken in the red area and on the variation between these dialects.

http://www.bfs.admin.ch/bfs/portal/de/index/regionen/thematische_karten/maps/bevoelkerung/sprachen_reli.html
The linguistic atlas of German-speaking Switzerland

Sprachatlas der deutschen Schweiz (SDS)

Data collection:
- From 1938 to 1958
- 565 inquiry points in Switzerland + 8 in Northern Italy
  - About every third village of German-speaking Switzerland
- Trained fieldworkers interview informants on-site and transcribe responses directly

Publication:
- 8 volumes published between 1962 and 1997
- 1548 hand-drawn maps
- 2 volumes on phonetics/phonology, 1 volume on morphology, 5 volumes on lexicon
The linguistic atlas of German-speaking Switzerland

Sprachatlas der deutschen Schweiz (SDS)

Data collection:
- From 1938 to 1958
- 565 inquiry points in Switzerland + 8 in Northern Italy
  - About every third village of German-speaking Switzerland
- Trained fieldworkers interview informants on-site and transcribe responses directly

Publication:
- 8 volumes published between 1962 and 1997
- 1548 hand-drawn maps
- 2 volumes on phonetics/phonology, 1 volume on morphology, 5 volumes on lexicon
The linguistic atlas of
German-speaking Switzerland

Sprachatlas der deutschen Schweiz (SDS)

Data collection:
- From 1938 to 1958
- 565 inquiry points in Switzerland + 8 in Northern Italy
  - About every third village of German-speaking Switzerland
- Trained fieldworkers interview informants on-site and transcribe responses directly

Publication:
- 8 volumes published between 1962 and 1997
- 1548 hand-drawn maps
- 2 volumes on phonetics/phonology, 1 volume on morphology, 5 volumes on lexicon
An original SDS map
An original SDS map

Map drawing involves a classification step:
Variants are represented by symbols.
An original SDS map

Map drawing involves a classification step:
Variants are represented by symbols.

Several variants may exist at a given point.
Map drawing involves a classification step: Variants are represented by symbols.

Several variants may exist at a given point.

Legend contains additional information.
Outline

1. Introduction
2. Digitizing SDS maps
3. Map selection
4. Content selection
5. The digitization process
6. Access and applications
Why digitize SDS maps?

PhD Y. Scherrer (2012):

- Transform Standard German words/sentences into various Swiss German dialects
- Transformation rules should be sensitive to target dialect
- SDS maps provide the necessary information
  - 2009–2010: digitization of 193 SDS maps by Y. Scherrer (V1)
  - Digital data are the prerequisite for many interesting dialectological/dialectometrical studies

SADS: syntax atlas project at University of Zurich

- Desire to compare SADS data with SDS data (cf. Wednesday’s talk)
- 2013–2014: digitization of additional maps by S. Kellerhals (V2)
Why digitize SDS maps?

PhD Y. Scherrer (2012):

- Transform Standard German words/sentences into various Swiss German dialects
- Transformation rules should be sensitive to target dialect
- SDS maps provide the necessary information
- 2009–2010: digitization of 193 SDS maps by Y. Scherrer
- Digital data are the prerequisite for many interesting dialectological/dialectometrical studies

SADS: syntax atlas project at University of Zurich

- Desire to compare SADS data with SDS data (cf. Wednesday’s talk)
- 2013–2014: digitization of additional maps by S. Kellerhals

Y. Scherrer, S. Kellerhals: Digitizing the linguistic atlas of German-speaking Switzerland
Why digitize SDS maps?

PhD Y. Scherrer (2012):

- Transform Standard German words/sentences into various Swiss German dialects
- Transformation rules should be sensitive to target dialect
- SDS maps provide the necessary information
- 2009–2010: digitization of 193 SDS maps by Y. Scherrer (V1)
- Digital data are the prerequisite for many interesting dialectological/dialectometrical studies

SADS: syntax atlas project at University of Zurich

- Desire to compare SADS data with SDS data (cf. Wednesday’s talk)
- 2013–2014: digitization of additional maps by S. Kellerhals (V2)
Why digitize SDS maps?

PhD Y. Scherrer (2012):

- Transform Standard German words/sentences into various Swiss German dialects
- Transformation rules should be sensitive to target dialect
- SDS maps provide the necessary information
- 2009–2010: digitization of 193 SDS maps by Y. Scherrer (V1)
- Digital data are the prerequisite for many interesting dialectological/dialectometrical studies

SADS: syntax atlas project at University of Zurich

- Desire to compare SADS data with SDS data (cf. Wednesday’s talk)
- 2013–2014: digitization of additional maps by S. Kellerhals (V2)
Why digitize SDS maps?

PhD Y. Scherrer (2012):

- Transform Standard German words/sentences into various Swiss German dialects
- Transformation rules should be sensitive to target dialect
- SDS maps provide the necessary information
- 2009–2010: digitization of 193 SDS maps by Y. Scherrer
- Digital data are the prerequisite for many interesting dialectological/dialectometrical studies

SADS: syntax atlas project at University of Zurich

- Desire to compare SADS data with SDS data (cf. Wednesday’s talk)
- 2013–2014: digitization of additional maps by S. Kellerhals
This talk

A lot of dialectologists are interested in finding a painless way of digitizing paper maps.

- We don’t have a miracle solution, but we would like to share our experience and findings.

Three parts of this talk:

1. Which maps were selected for digitization? Why?
2. What aspects of the maps were taken into account during digitization? Why?
3. What were the different practical steps of the digitization process?
This talk

A lot of dialectologists are interested in finding a painless way of digitizing paper maps.

- We don’t have a miracle solution, but we would like to share our experience and findings.

Three parts of this talk:

1. Which maps were selected for digitization? Why?
2. What aspects of the maps were taken into account during digitization? Why?
3. What were the different practical steps of the digitization process?
Outline

1. Introduction
2. Digitizing SDS maps
3. Map selection
4. Content selection
5. The digitization process
6. Access and applications

Y. Scherrer, S. Kellerhals: Digitizing the linguistic atlas of German-speaking Switzerland
Map selection

What are the criteria that guided the selection process?

1. **Type frequency**
   - Phonetic and morphological phenomena have, by definition, a higher type frequency than lexical phenomena
   - Bias towards SDS volumes 1 to 3

2. **Token frequency**
   - For lexical phenomena, we prefer function words with high token frequency

3. **Represent linguistic phenomena, not words**
   - There are 40 SDS maps that illustrate vowel lengthening in an open syllable on 40 different words.
   - We do not need all of them; we chose 7 representative maps.
What are the criteria that guided the selection process?

1. **Type frequency**
   - Phonetic and morphological phenomena have, by definition, a higher type frequency than lexical phenomena
   - Bias towards SDS volumes 1 to 3

2. **Token frequency**
   - For lexical phenomena, we prefer function words with high token frequency

3. **Represent linguistic phenomena, not words**
   - There are 40 SDS maps that illustrate vowel lengthening in an open syllable on 40 different words.
   - We do not need all of them; we chose 7 representative maps.
Map selection

What are the criteria that guided the selection process?

1. Type frequency
   - Phonetic and morphological phenomena have, by definition, a higher type frequency than lexical phenomena
   - Bias towards SDS volumes 1 to 3

2. Token frequency
   - For lexical phenomena, we prefer function words with high token frequency

3. Represent linguistic phenomena, not words
   - There are 40 SDS maps that illustrate vowel lengthening in an open syllable on 40 different words.
   - We do not need all of them; we chose 7 representative maps.
Map selection

Original SDS maps selected for digitization:

- **Phonetics (Vol. I–II):** 63
- **Morphology (Vol. III):** 91
- **Lexicon (Vol. IV–VIII):** 39

Legend:
- V1
- V2

Y. Scherrer, S. Kellerhals: Digitizing the linguistic atlas of German-speaking Switzerland
Multiple phenomena on one map
Multiple phenomena on one map

Symbol form determines realization of MHG *iu* as öi, oi, öü, üü.
Multiple phenomena on one map

**Symbol form** determines realization of MHG *iu* as *öi*, *oi*, *öü*, *üü*.

**Hachure** determines presence of unrounding (*ü* → *i*, *ö* → *e*).
Multiple phenomena on one map

Symbol form determines realization of MHG *iu* as öi, oi, öü, üü.

Hachure determines presence of unrounding (ü → i, ö → e).

Small circle below symbol determines presence of w in hiatus.
Multiple phenomena on one map

Symbol form determines realization of MHG iu as öi, oi, öü, üü.

Hachure determines presence of unrounding (ü→i, ö→e).

Small circle below symbol determines presence of w in hiatus.

This single SDS map allows us to create three working maps.
Original maps vs working maps

Original maps on the left, working maps on the right.

Phonetics: V1 = 63, V2 = 71
Morphology: V1 = 91, V2 = 102
Lexicon: V1 = 39, V2 = 36

CHERRER, S. KELLERHALS: DIGITIZING THE LINGUISTIC ATLAS OF GERMAN-SPEAKING SWITZERLAND
Outline

1. Introduction
2. Digitizing SDS maps
3. Map selection
4. Content selection
5. The digitization process
6. Access and applications
1. Precision of phonetic distinctions

Das Wort an sich ist in FR 13,14, BE 87, 88, 92, 95, GR 25, 27, 33, WS 9, 11, 22, 34, T 1, T 4 als jung bezeichnet und fehlt in T 1-3, 6-8; älter ist hier überall "förhuus" (in T 1, 6 'Huus'). WS 6, 7: Endung 'in'.

Zeichenerklärung

im nichtschraff. Gebiet

<table>
<thead>
<tr>
<th>u, auch y</th>
<th>y, auch u</th>
<th>geschlossenes u</th>
</tr>
</thead>
<tbody>
<tr>
<td>y, u, u, u</td>
<td>y, u (BA 5)</td>
<td>offen, auch neutral</td>
</tr>
<tr>
<td>θ, θ</td>
<td>θ (SO 6)</td>
<td>Zwischenwerte gegen o</td>
</tr>
<tr>
<td>θ</td>
<td>o</td>
<td>geschlossenes o</td>
</tr>
<tr>
<td>θ</td>
<td>o</td>
<td>offenes o</td>
</tr>
</tbody>
</table>

Zeichen eingeklammert: Beleg aus Frage 196.5 (Zuss.) s. unten II

Nebenzeichen

Fragen 193.4a, 196.5 (nicht durchgehend belegt) und (häufig) Sp., sofern anders als Hauptstelle 193.4:

Punkt: geschlossener als Wert 193.4,
Komma: offener als Wert 193.4,
Punkt und Komma: sowohl geschlossener als auch offener.
1. Precision of phonetic distinctions

Five symbols to represent the phonetic continuum between u and o.

- Danger of fieldworker isoglosses!

The spelling conventions for Swiss German (Dieth 1986) only contains the symbols u, o.
1. Precision of phonetic distinctions

Five symbols to represent the phonetic continuum between \( \text{u} \) and \( \text{o} \).

- Danger of fieldworker isoglosses!

The spelling conventions for Swiss German (Dieth 1986) only contains the symbols \( \text{u}, \text{o} \).

We only distinguish as many variants as can be represented with the Dieth spelling rules:

- Upper 3 variants \( \rightarrow \text{u} \)
- Lower 2 variants \( \rightarrow \text{o} \)
2. Omission of variants with less than 10 occurrences
2. Omission of variants with less than 10 occurrences

The following symbols appear less than 10 times each:

• : : * ~ ~
2. Omission of variants with less than 10 occurrences

The following symbols appear less than 10 times each:

- •
- ↓
- *
- ~
- ~

We omit these data points.

We assume that many of these small-scale variants have disappeared since the data collection.
3. Omission of Italian inquiry points
3. Omission of Italian inquiry points

8 inquiry points are located in Northern Italy, where German-speaking Walser have emigrated in the Middle Ages.

However, these dialects seem to be dying out.
3. Omission of Italian inquiry points

8 inquiry points are located in Northern Italy, where German-speaking Walser have emigrated in the Middle Ages.

However, these dialects seem to be dying out.

We have omitted these points in the digitized maps.
Outline

1. Introduction
2. Digitizing SDS maps
3. Map selection
4. Content selection
5. The digitization process
6. Access and applications
1. Scanning

- The SDS maps are larger than A3, so need special scanning equipment.
- Photographing the maps in good lighting conditions works as well. This has been used for V2 maps.
2. Georeferencing

Associate points on the map with Swiss geographic coordinates.

- Load scanned map into ArcGIS.
- Use the georeferencing tool to annotate four easily identifiable cities with their SwissGrid (CH1903+) coordinates.
2. Georeferencing

Associate points on the map with Swiss geographic coordinates.

- Load scanned map into ArcGIS.
- Use the georeferencing tool to annotate four easily identifiable cities with their SwissGrid (CH1903+) coordinates.

- Basel (611 300 / 267 600)
- St. Gallen (746 200 / 254 600)
- Bern (600 100 / 199 700)
- Chur (759 300 / 191 100)
3. Creating shapefiles
3. Creating shapefiles

Load scanned map in background
3. Creating shapefiles

1. Load scanned map in background
2. Load shapefile of all inquiry points (blue)
3. Creating shapefiles

1. Load scanned map in background

2. Load shapefile of all inquiry points (blue)

3. Create empty shapefile for each variant, add red points by clicking
   Use *snapping* to match positions of blue points
3. Creating shapefiles

1. Load scanned map in background

2. Load shapefile of all inquiry points (blue)

3. Create empty shapefile for each variant, add red points by clicking
   Use *snapping* to match positions of blue points

4. Merge all variants into one shapefile
3. Creating shapefiles
3. Creating shapefiles
The ArcGIS shapefiles can be downloaded freely at the address indicated above.

We are planning to make Excel files available as well.

At the same address, we provide an online viewing tool:
We also provide demos of several applications that rely on digitized SDS data:

Dialectometric studies (cf. our talk on Wednesday)
www.dialektkarten.ch
(http://latIntic.unige.ch/~scherrey/)

We also provide **demos** of several **applications** that rely on digitized SDS data:

**Machine translation into Swiss German (slow demo)**
We also provide **demos** of several **applications** that rely on digitized SDS data:

**Dialect identification**
Conclusion

- We have made available 231 digitized working maps, extracted from 212 original SDS maps.
  - This amounts to about 14% of all SDS maps.
  - Our digitization methods work well, but remain time-consuming.
  - On average, 30 minutes per working map.

- This represents only a small part of the wealth of information available in the SDS, but we are not done yet:
  - Additional maps are being digitized at University of Zurich

- Data are available online – we are happy if you use them and/or improve them!

http://www.dialektkarten.ch
http://latln tic.unige.ch/~scherrey/
Conclusion

- We have made available 231 digitized working maps, extracted from 212 original SDS maps.
  - This amounts to about 14% of all SDS maps.
  - Our digitization methods work well, but remain time-consuming.
  - On average, 30 minutes per working map.

- This represents only a small part of the wealth of information available in the SDS, but we are not done yet:
  - Additional maps are being digitized at University of Zurich

- Data are available online – we are happy if you use them and/or improve them!

http://www.dialektkarten.ch
http://latlntic.unige.ch/~scherrey/
Conclusion

- We have made available 231 digitized working maps, extracted from 212 original SDS maps.
  - This amounts to about 14% of all SDS maps.
  - Our digitization methods work well, but remain time-consuming.
  - On average, 30 minutes per working map.

- This represents only a small part of the wealth of information available in the SDS, but we are not done yet:
  - Additional maps are being digitized at University of Zurich

- Data are available online – we are happy if you use them and/or improve them!

http://www.dialektkarten.ch
http://latlntic.unige.ch/~scherrey/