Machine Translation for Swiss German Dialects

SCHERRER, Yves

Abstract

Most machine translation systems apply to written, standardized language varieties. In contrast, we present a model that takes into account some specificities of dialects: - Dialect areas show continuous variation along all levels of linguistic analysis; it is linguistically not satisfactory to model a finite number of distinct dialects as if they were unrelated languages. - Written dialectal data is often not available in sufficient quality and quantity; however, we can rely on atlases, dictionaries and grammars compiled by dialectologists. - Most dialects are closely related to a resource-rich standardized language (in our case, Standard German). We argue that machine translation into Swiss German dialects is best modelled by a traditional rule-based transfer system that can be parametrized by probabilistic geographical maps. We show how such maps can be extracted from dialectological atlases. We will also discuss some issues related to the evaluation of multidialectal machine translation.
Machine Translation for Swiss German Dialects

Yves Scherrer

Université de Genève

yves.scherrer@unige.ch

10/14/2009
1. Introduction: Languages in Switzerland

2. Machine Translation for Swiss German Dialects I
   Motivations and Architecture

3. There’s a map for that!

4. Machine Translation for Swiss German Dialects II
   Incorporating Dialectological Maps

5. Evaluation

6. Conclusion
1 Introduction: Languages in Switzerland

2 Machine Translation for Swiss German Dialects I
   Motivations and Architecture

3 There’s a map for that!

4 Machine Translation for Swiss German Dialects II
   Incorporating Dialectological Maps

5 Evaluation

6 Conclusion
Languages in Switzerland

Languages in Switzerland

Landessprachen in den Gemeinden, 2000
Langues nationales parlées dans les communes, en 2000

Dominierende Landessprache
Langue nationale la plus parlée

- Deutsche Sprache
  - Deutsch/Deutschsprachiger Dominanz
  - Predominance de l’allemand

- Französische Sprache
  - Französischsprachiger Dominanz
  - Predominance du français

- Italienische Sprache
  - Italienischsprachiger Dominanz
  - Predominance de l’italien

- Rätoromanische Sprache
  - Rätoromanischsprachiger Dominanz
  - Predominance du romanche

- Keine Dominanz
  - keine Dominanz
  - aucune prédominance

Values:

- schwach
- faible
- stark
- forte

70 – 84.9 %

≥ 85 %

Dominierende Landessprache
Langue nationale la plus parlée

German / Swiss German
French
Italian
Romansh

http://www.bfs.admin.ch/bfs/portal/de/index/regionen/thematische_karten/maps/bevoelkerung/sprachen_religionen.html

Yves Scherrer (Université de Genève)
Diglossia: A stable language situation in which two language varieties are in concurrent use:

- a standardized high/prestigious variety for formal use,
- a low/vernacular variety for ordinary conversation.

In German-speaking Switzerland:

„High“ variety Standard German, used in writing
„Low“ variety Swiss German dialects, used in speech (and electronic writing)

„High“ and „low“ do not refer to social status, but to modality only.

French-speaking and Italian-speaking Swiss only learn Standard German as a second language.
Languages in Switzerland
Swiss German Dialects

'nothing'

Standard German: 
nichts, (nix)

Swiss German variants:
nüüt
nüt
niit
nünt
nütz
nix

SDS 4/171
Languages in Switzerland

An example

Standard German:

Warst du auf dem Markt einkaufen?
Were you at the market shopping?
'Have you been shopping at the market?'

Swiss German Dialects:

ZH  Bisch uf em Mèrt gsy go poschte?
SZ  Bisch ufem Märcht gsi go poschte?
SO Besch ofem Märed gsy go boschte?
OW Bischt ufm Mär gsi go ichoifä?
SG Bisch uf dä Märt go poschtä?
BE Bisch uf e Märit ga kömerle?
FR Büschu z’Määret gsi ga iichufe?
WS Bisch dü uf um Markt gsi ga ichöifu?

Differences between Standard German and Swiss German:

- Preterite tense replaced by present perfect
- *go/ga* + infinitive construction

Differences within Swiss German dialects:

- No standardized spelling
- Phonetics
- Lexicon
- Morphology
- Syntax

http://als.wikipedia.org/wiki/Alemannischer_Beispielsatz
Languages in Switzerland

An example

Standard German:
Warst du auf dem Markt einkaufen?
Were you at the market shopping?
‘Have you been shopping at the market?’

Swiss German Dialects:
ZH Bisch uf em Mèrt gsy go poschte?
SZ Bisch ufem Märcht gsi go poschte?
SO Besch ofem Märed gsy go boschte?
OW Bisch tuf Märt gsi go ichoifä?
SG Bisch uf dä Märtg go poschtä?
BE Bisch uf e Märit ga kömerle?
FR Büschu z’Määret gsi ga iichufe?
WS Bisch dü uf um Markt gsi ga ichöifu?

Differences between Standard German and Swiss German:
- Preterite tense replaced by present perfect
- go/ga + infinitive construction

Differences within Swiss German dialects:
- No standardized spelling
- Phonetics
- Lexicon
- Morphology
- Syntax

http://als.wikipedia.org/wiki/Alemannischer_Beispielsatz
Languages in Switzerland

An example

Standard German:

Warst du auf dem Markt einkaufen?
Were you at the market shopping?
‘Have you been shopping at the market?’

Swiss German Dialects:

ZH Bisch uf em Mèrt gsy go poschte?
SZ Bisch ufem Märcht gsi go poschte?
SO Besch ofem Märed gsy go boschte?
OW Bischt ufm Märt gsii go ichoifä?
SG Bisch uf dä Märt go poschtä?
BE Bisch uf e Märit ga kömerle?
FR Büschu z’Määret gsi ga iichufe?
WS Bisch dü uf um Markt gsii ga ichöifu?

Differences between Standard German and Swiss German:

- Preterite tense replaced by present perfect
- go/ga + infinitive construction

Differences within Swiss German dialects:

- No standardized spelling
- Phonetics
- Lexicon
- Morphology
- Syntax

http://als.wikipedia.org/wiki/Alemannischer_Beispielsatz
Languages in Switzerland
An example

Standard German:

Warst du auf dem Markt einkaufen?
Were you at the market shopping?
‘Have you been shopping at the market?’

Swiss German Dialects:

ZH  Bisch uf em Mèrt gsy go poschte?
SZ  Bisch ufem Märcht gsi go poschte?
SO  Besch ofem Märed gsy go boschté?
OW  Bischt ufM Márt gsii go ichoifä?
SG  Bisch uf dä Márt go poschtá?
BE  Bisch uf e Märit ga kömerle?
FR  Büschu z’Määret gsi ga iichufe?
WS  Bisch dü uf um Markt gsii ga ichöifu?

Differences between Standard German and Swiss German:

- Preterite tense replaced by present perfect
- go/ga + infinitive construction

Differences within Swiss German dialects:

- No standardized spelling
- Phonetics
- Lexicon
- Morphology
- Syntax

http://als.wikipedia.org/wiki/Alemannischer_Beispielsatz
Languages in Switzerland

An example

Standard German:

Warst du auf dem Markt einkaufen?  
Were you at the market shopping?  
‘Have you been shopping at the market?’

Swiss German Dialects:

ZH  Bisch uf em Mèrt gsy go poschte?  
SZ  Bisch ufem Märcht gsi go poschte?  
SO  Besch ofem Märed gsy go boschte?  
OW  Bischt ufm Märt gsii go ichoifä?  
SG  Bisch uf dä Märt go poschtä?  
BE  Bisch uf e Märit ga kömerle?  
FR  Büschu z’Määret gsi ga ichufu?  
WS  Bisch dü uf um Markt gsii ga ichöifu?

Differences between Standard German and Swiss German:

- Preterite tense replaced by present perfect
- go/ga + infinitive construction

Differences within Swiss German dialects:

- No standardized spelling
- Phonetics
- Lexicon
- Morphology
- Syntax

http://als.wikipedia.org/wiki/Alemannischer_Beispielsatz
Languages in Switzerland
An example

Standard German:
Warst du auf dem Markt einkaufen?
Were you at the market shopping?
‘Have you been shopping at the market?’

Swiss German Dialects:
ZH Bisch uf em Mèrt gsy go poschté?
SZ Bisch ufem Märcht gsi go poschté?
SO Besch ofem Märed gsy go boschté?
OW Bischt ufm Märt gsii go ichoifä?
SG Bisch uf dä Märt go poschtä?
BE Bisch uf e Märit ga kömerle?
FR Büschu z’Määret gsi ga iichufe?
WS Bisch dü uf um Markt gsii ga ichöifu?

Differences between Standard German and Swiss German:
- Preterite tense replaced by present perfect
- *go/ga* + infinitive construction

Differences within Swiss German dialects:
- No standardized spelling
- Phonetics
- Lexicon
- Morphology
- Syntax

http://als.wikipedia.org/wiki/Alemannischer_Beispielsatz
Languages in Switzerland

An example

Standard German:

Warst du auf dem Markt einkaufen?
Were you at the market shopping?
‘Have you been shopping at the market?’

Swiss German Dialects:

ZH  Bisch uf em Mèrt gsy go poschte?
SZ  Bisch ufem Märcht gsi go poschte?
SO  Besch ofem Märed gsy go boschte?
OW  Bischt ufm Mär gsiı go ichoifä?
SG  Bisch uf dä Märzt go poschtä?
BE  Bisch uf e Märit ga kömerle?
FR  Büschu z’Määret gsi ga iichufe?
WS  Bisch dü uf um Markt gsiı ga ichöifu?

Differences between Standard German and Swiss German:

- Preterite tense replaced by present perfect
- go/ga + infinitive construction

Differences within Swiss German dialects:

- No standardized spelling
- Phonetics
- Lexicon
- Morphology
- Syntax

http://als.wikipedia.org/wiki/Alemannischer_Beispielsatz
Languages in Switzerland

An example

Standard German:

Warst du auf dem Markt einkaufen?
Were you at the market shopping?
‘Have you been shopping at the market?’

Swiss German Dialects:

ZH  Bisch uf em Mèrt gsy go poschte?
SZ  Bisch ufem Märcht gsi go poschte?
SO  Besch ofem Máred gsy go boschte?
OW  Bischt ufm Márt gsii go ichoifä?
SG  Bisch uf dä Märt go poschtä?
BE  Bisch uf e Märit ga kömerle?
FR  Büschu z’Määret gsi ga iichufe?
WS  Bisch dü uf um Markt gsii ga ichöifu?

Differences between Standard German and Swiss German:

- Preterite tense replaced by present perfect
- go/ga + infinitive construction

Differences within Swiss German dialects:

- No standardized spelling
- Phonetics
  - Lexicon
  - Morphology
  - Syntax

http://als.wikipedia.org/wiki/Alemannischer_Beispielsatz
Languages in Switzerland
An example

Standard German:
Warst du auf dem Markt einkaufen?
Were you at the market shopping?
‘Have you been shopping at the market?’

Swiss German Dialects:

ZH  Bisch uf em Mèrt gsy go poschte?
SZ  Bisch ufem Märcht gsi go poschte?
SO  Besch ofem Märed gsy go boschte?
OW  Bischt ufm Märt gsii go ichoifä?
SG  Bisch uf dä Märt go poschtä?
BE  Bisch uf e Märit ga kömerle?
FR  Büschu z’Määret gsi ga iichufé?
WS  Bisch dü uf um Markt gsii ga ichöifu?

Differences between Standard German and Swiss German:
• Preterite tense replaced by present perfect
• go/ga + infinitive construction

Differences within Swiss German dialects:
• No standardized spelling
• Phonetics
• Lexicon
• Morphology
• Syntax

http://als.wikipedia.org/wiki/Alemannischer_Beispielsatz
Languages in Switzerland

An example

Standard German:

Warst du auf dem Markt einkaufen?
Were you at the market shopping?
‘Have you been shopping at the market?’

Swiss German Dialects:

ZH   Bisch uf em Mèrt gsy go poschte?
SZ   Bisch ufem Märcht gsi go poschte?
SO   Besch ofem Märed gsy go boschte?
OW   Bischt ufm Märt gsii go ichoifä?
SG   Bisch uf dä Märt go poschtä?
BE   Bisch uf e Märit ga kömerle?
FR   Büschu z’Määret gsi ga iichufe?
WS   Bisch dü uf um Markt gsii ga ichöifu?

Differences between Standard German and Swiss German:

- Preterite tense replaced by present perfect
- *go/ga* + infinitive construction

Differences within Swiss German dialects:

- No standardized spelling
- Phonetics
- Lexicon
- Morphology
- Syntax

http://als.wikipedia.org/wiki/Alemannischer_Beispielsatz
Languages in Switzerland

An example

Standard German:
Warst du auf dem Markt einkaufen?
Were you at the market shopping?
‘Have you been shopping at the market?’

Swiss German Dialects:
ZH Bisch uf em Mèrt gsy go poschte?
SZ Bisch ufem Märcht gsi go poschte?
SO Besch ofem Märed gsy go boschte?
OW Bischt ufm Märt gsii go ichöifä?
SG Bisch uf dä Märt go poschtä?
BE Bisch uf e Märit ga kömerle?
FR Büschu z’Määret gsi ga iichufe?
WS Bisch dü uf um Markt gsii ga ichöifu?

Differences between Standard German and Swiss German:
- Preterite tense replaced by present perfect
- go/ga + infinitive construction

Differences within Swiss German dialects:
- No standardized spelling
- Phonetics
- Lexicon
- Morphology
- Syntax

http://als.wikipedia.org/wiki/Alemannischer_Beispielsatz
Languages in Switzerland

An example

Standard German:

Warst du auf dem Markt einkaufen?
Were you at the market shopping?
‘Have you been shopping at the market?’

Swiss German Dialects:

ZH  Bisch uf em Mèrt gsy go poschte?
SZ  Bisch ufem Märcht gsi go poschte?
SO  Besch ofem Märed gsy go boschte?
OW  Bischt ufm Märt gsii go ichoifä?
SG  Bisch uf dä Märt go poschtä?
BE  Bisch uf e Märit ga kömerle?
FR  Büschu z’Määret gsi ga iichufe?
WS  Bisch dü uf um Markt gsii ga ichöifu?

Differences between Standard German and Swiss German:

- Preterite tense replaced by present perfect
- go/ga + infinitive construction

Differences within Swiss German dialects:

- No standardized spelling
- Phonetics
- Lexicon
- Morphology
- Syntax

http://als.wikipedia.org/wiki/Alemannischer_Beispielsatz
Languages in Switzerland

An example

Standard German:
Warst du auf dem Markt einkaufen?  
Were you at the market shopping?  
‘Have you been shopping at the market?’

Swiss German Dialects:

ZH  Bisch uf em Mèrt gsy go poschte?
SZ  Bisch ufem Märcht gsi go poschte?
SO  Besch ofem Märed gsy go boschte?
OW  Bischt ufM Märt gsii go ichoifä?
SG  Bisch uf dä Märt go poschtä?
BE  Bisch uf e Märit ga kömerle?
FR  Büschu z’Määret gsi ga iichufe?
WS  Bisch dü uf um Markt gsii ga ichöifu?

Differences between Standard German and Swiss German:
- Preterite tense replaced by present perfect
- go/ga + infinitive construction

Differences within Swiss German dialects:
- No standardized spelling
- Phonetics
- Lexicon
- Morphology
- Syntax

http://als.wikipedia.org/wiki/Alemannischer_Beispielsatz
1. Introduction: Languages in Switzerland

2. Machine Translation for Swiss German Dialects I
   Motivations and Architecture

3. There’s a map for that!

4. Machine Translation for Swiss German Dialects II
   Incorporating Dialectological Maps

5. Evaluation

6. Conclusion
Research Questions:

- A lot of dialectological research on Swiss German dialects. How can we integrate this work into a Computational Linguistics application?
- Most variation between dialects is thought to be continuous. How can we model continuity? Can we create one system for the entire Swiss German dialect area?
- Swiss German dialects are closely related to Standard German, for which a lot of resources are available. How can we draw on these resources to build models for Swiss German dialects?
Potential Applications:

- Educational tools for learning dialects
  - Not for Swiss Germans: they understand all dialects + Standard German
  - However, French-speaking and Italian-speaking Swiss, and Germans, might want to learn dialects on the basis of their Standard German knowledge

- Mass-media
  - Subtitling for the deaf (and for Germans)
  - Preparation of dialect-spoken programs

- Industry
  - Spoken Dialogue systems are not currently used in Switzerland because of the complex linguistic situation
Our Approach:

- Machine Translation from Standard German to (written representations of) all Swiss German dialects
- Rule-based transfer model, parametrized by dialectological maps

Why translate into dialects?

- Sidesteps the issue of dialect normalization (lack of spelling conventions)
- Easier to apply dialectological data to standardized input

Why RBMT?

- No training corpora available for multiple dialects
- Explicit integration of dialectological data
- Bottleneck of Transfer-RBMT: Source language parser
- Bottleneck of Phrase-based SMT: Target language model
Machine Translation

Our Approach:
- Machine Translation from Standard German to (written representations of) all Swiss German dialects
- Rule-based transfer model, parametrized by dialectological maps

Why translate into dialects?
- Sidesteps the issue of dialect normalization (lack of spelling conventions)
- Easier to apply dialectological data to standardized input

Why RBMT?
- No training corpora available for multiple dialects
- Explicit integration of dialectological data
- Bottleneck of Transfer-RBMT: Source language parser
  Bottleneck of Phrase-based SMT: Target language model
Our Approach:
- Machine Translation from Standard German to (written representations of) all Swiss German dialects
- Rule-based transfer model, parametrized by dialectological maps

Why translate into dialects?
- Sidesteps the issue of dialect normalization (lack of spelling conventions)
- Easier to apply dialectological data to standardized input

Why RBMT?
- No training corpora available for multiple dialects
- Explicit integration of dialectological data
- Bottleneck of Transfer-RBMT: Source language parser
- Bottleneck of Phrase-based SMT: Target language model
System Architecture

Choice of target dialect

Std German text

Std German parser

Std German parse tree

Hand-crafted transfer rules

Dialect parse tree

Linearization (Generation)

Dialect text
Choice of target dialect

Std German text → Std German parser → Std German parse tree → Hand-crafted transfer rules → Dialect parse tree → Linearization (Generation) → Dialect text

Goal:
- One model for the entire Swiss German dialect landscape.
- Transfer rules are connected to probability maps.
System Architecture

What needs to be done

Syntax:
- Some local changes, a few long distance changes
- A full parse of the source text helps...

Lexicon:
- No explicit translation lexicon for low-frequency content words
  - Mere phonetic adaptations of the Standard German words
  - Store regular expressions („phonetic rules“) instead of words
- Explicit translation lexicon for high-frequency content words and function words
  - Either completely different lemmas, or irregular phonetic changes
  - „Lexical rules“
- Need lemmatization of source words

Morphology:
- Rebuild Swiss German inflection from stems + morphological features
- Need morphological disambiguation of source words
System Architecture
What needs to be done

Syntax:
- Some local changes, a few long distance changes
- A full parse of the source text helps...

Lexicon:
- No explicit translation lexicon for low-frequency content words
  - Mere phonetic adaptations of the Standard German words
  - Store regular expressions ("phonetic rules") instead of words
- Explicit translation lexicon for high-frequency content words and function words
  - Either completely different lemmas, or irregular phonetic changes
  - "Lexical rules"
- Need lemmatization of source words

Morphology:
- Rebuild Swiss German inflection from stems + morphological features
- Need morphological disambiguation of source words
Syntax:
- Some local changes, a few long distance changes
- A full parse of the source text helps...

Lexicon:
- No explicit translation lexicon for low-frequency content words
  - Mere phonetic adaptations of the Standard German words
  - Store regular expressions ("phonetic rules") instead of words
- Explicit translation lexicon for high-frequency content words and function words
  - Either completely different lemmas, or irregular phonetic changes
  - "Lexical rules"
- Need lemmatization of source words

Morphology:
- Rebuild Swiss German inflection from stems + morphological features
- Need morphological disambiguation of source words
System Architecture

Choice of target dialect

Std German text

Std German parser

Std German parse tree

Hand-crafted transfer rules

Dialect parse tree

Linearization (Generation)

Dialect text
System Architecture

Choice of target dialect

Std German text

Std German parser

Std German parse tree

Hand-crafted transfer rules

Dialect parse tree

Linearization (Generation)

Dialect text

- Lemmatization
- Morphological disambiguation
- Part-of-speech tags
- Phrase structure or dependencies
System Architecture

- Choice of target dialect
  - Std German text
    - Std German parser
      - Std German parse tree
        - Hand-crafted transfer rules
          - Dialect parse tree
            - Linearization (Generation)
              - Dialect text

1. Lexical rules: substitution of non-related lexemes
   - May modify other nodes (gender, verb arguments)
2. Syntactic transfer rules
3. Phonetic adaptation of related word roots
4. Generation of inflected word forms (morphology)
5. Phonetic adaptation of inflected word forms
   - Accounts for interactions between adapted root and inflection morphemes
System Architecture

Choice of target dialect

Std German text

Std German parser

Std German parse tree

Hand-crafted transfer rules

Dialect parse tree

Linearization (Generation)

Dialect text

Elision and Sandhi phenomena
1) Some details

2) Details about map integration
Currently using the Fips parser, but considering other systems. XML-based format to represent sentence structure.

**Fips:**
- A rule-based parser developed at University of Geneva for several languages (Wehrli 2007).
- Based on Government & Binding theory of Generative Grammar.
- As other unification-based approaches, morphological and lexical disambiguation is done during parsing.
- Coverage of the German parser is limited.

**Statistical Approaches:**
- Generally use different models for morphological disambiguation, lemmatization, POS tagging, and full parsing.
- Likely to improve coverage and accuracy.
Outline

1. Introduction: Languages in Switzerland
2. Machine Translation for Swiss German Dialects I
   Motivations and Architecture
3. There’s a map for that!
4. Machine Translation for Swiss German Dialects II
   Incorporating Dialectological Maps
5. Evaluation
6. Conclusion
Dialectological Maps

Sources

**SDS** (*Sprachatlas der deutschen Schweiz*):
- Data collection 1939-1958, publication 1962-1997 (8 volumes)
- Hand-drawn maps

<table>
<thead>
<tr>
<th></th>
<th>Total maps</th>
<th>Target</th>
<th>Current</th>
</tr>
</thead>
<tbody>
<tr>
<td>Phonetics/phonology</td>
<td>300</td>
<td>80</td>
<td>1</td>
</tr>
<tr>
<td>Morphology + irregular verbs</td>
<td>210</td>
<td>100</td>
<td>20</td>
</tr>
<tr>
<td>Lexicon (function + content words)</td>
<td>800</td>
<td>100</td>
<td>10</td>
</tr>
</tbody>
</table>

**SADS** (*Syntaktischer Atlas der deutschen Schweiz*):
- Data collection recently completed, publication scheduled 2010
- Access to numerical raw data

<table>
<thead>
<tr>
<th></th>
<th>Total questions</th>
<th>Target maps</th>
<th>Current</th>
</tr>
</thead>
<tbody>
<tr>
<td>Syntax</td>
<td>118</td>
<td>30</td>
<td>2</td>
</tr>
</tbody>
</table>
Criteria for map selection:

- High-frequency phenomena (type or token frequency)
- Phenomena still used in present-day dialectal speech
- Phenomena whose dialectological description is complete

Rules that apply uniformly to the entire Swiss German dialect area are not counted here.
Scan images of hand-drawn dialect atlas maps
Digitize images to create point maps
Interpolate point maps to create surface maps
  Constraints? Algorithm?
Dialectological Maps

Original hand-drawn map from the SDS atlas

Yves Scherrer (Université de Genève)  MT for Swiss German Dialects

10/14/2009  21 / 40
Dialectological Maps

Digitized point map
Dialectological Maps
Interpolated surface maps for each variant

Black: p=1
White: p=0
The current interpolation algorithm (Inverse Distance Weighted) is based on distance only. Do topographical, political, religious borders influence dialects? Should they be included in the model?

The current algorithm presupposes continuous dialectal variation. But some dialect boundaries are quite clear-cut...

- 19th century debate in dialectology
- Rumpf et al. (in press) use a Voronoi mosaic (i.e. polygons)

Atlas data is from the 1950s. Can we model diachronic dialect change?

Christen (1998):
- Spatially limited variants tend to disappear
- Cities influence dialects of their suburbs
- Lexicon: convergence within Swiss German dialects and with Standard German
The current interpolation algorithm (Inverse Distance Weighted) is based on distance only. Do topographical, political, religious borders influence dialects? Should they be included in the model?

The current algorithm presupposes continuous dialectal variation. But some dialect boundaries are quite clear-cut...  
- 19th century debate in dialectology
- Rumpf et al. (in press) use a Voronoi mosaic (i.e. polygons)

Atlas data is from the 1950s. Can we model diachronic dialect change? Christen (1998):
- Spatially limited variants tend to disappear
- Cities influence dialects of their suburbs
- Lexicon: convergence within Swiss German dialects and with Standard German
The current interpolation algorithm (Inverse Distance Weighted) is based on distance only. Do topographical, political, religious borders influence dialects? Should they be included in the model?

The current algorithm presupposes continuous dialectal variation. But some dialect boundaries are quite clear-cut…

- 19th century debate in dialectology
- Rumpf et al. (in press) use a Voronoi mosaic (i.e. polygons)

Atlas data is from the 1950s. Can we model diachronic dialect change?

Christen (1998):

- Spatially limited variants tend to disappear
- Cities influence dialects of their suburbs
- Lexicon: convergence within Swiss German dialects and with Standard German
1. Introduction: Languages in Switzerland
2. Machine Translation for Swiss German Dialects I
   - Motivations and Architecture
3. There’s a map for that!
4. Machine Translation for Swiss German Dialects II
   - Incorporating Dialectological Maps
5. Evaluation
6. Conclusion
Map-based Machine Translation

Choice of target dialect

- Std German text
  - Std German parser
    - Std German parse tree
      - Hand-crafted transfer rules
        - Dialect parse tree
          - Linearization (Generation)
            - Dialect text

- One model for the entire Swiss German dialect landscape.
- Transfer rules are connected to probability maps.
- Most times, there is more than one correct translation:
  - Dialects are not normed.
  - Different dialects can influence each other in transition zones.
  - Every translation is ranked according to the values read in the probability maps.
Map-based Machine Translation

Each transfer rule covers one linguistic variable and is implemented as one procedure. It is linked to a set of surface maps covering the probabilities of each variant.

Basic Algorithm

For each rule:
- if the rule applies to current configuration (word or sentence):
  - open all linked variant maps
  - read off probabilities of variants at target dialect coordinates
  - for each variant whose probability > threshold:
    - copy the original parse tree of the sentence
    - apply the transformation of the variant to this copy
    - adjust its probability
    - store the copy for further rule applications

This algorithm creates a trellis of modified parse trees, growing exponentially with rule alternatives. This is a potential problem that we have not yet addressed.
Map-based Machine Translation

Each transfer rule covers one linguistic variable and is implemented as one procedure. It is linked to a set of surface maps covering the probabilities of each variant.

Basic Algorithm

For each rule:
- if the rule applies to current configuration (word or sentence):
  - open all linked variant maps
  - read off probabilities of variants at target dialect coordinates
  - for each variant whose probability > threshold:
    - copy the original parse tree of the sentence
    - apply the transformation of the variant to this copy
    - adjust its probability
    - store the copy for further rule applications

This algorithm creates a trellis of modified parse trees, growing exponentially with rule alternatives. This is a potential problem that we have not yet addressed.
Map-based Machine Translation

Each transfer rule covers one linguistic variable and is implemented as one procedure. It is linked to a set of surface maps covering the probabilities of each variant.

Basic Algorithm

For each rule:
- if the rule applies to current configuration (word or sentence):
  - open all linked variant maps
  - read off probabilities of variants at target dialect coordinates
  - for each variant whose probability > threshold:
    - copy the original parse tree of the sentence
    - apply the transformation of the variant to this copy
    - adjust its probability
    - store the copy for further rule applications

This algorithm creates a trellis of modified parse trees, growing exponentially with rule alternatives. This is a potential problem that we have not yet addressed.
Map-based Machine Translation

An example: Prepositional Dative Marking

In Standard German:
- Inflectional marking of dative case
- Dative case used for indirect objects, and after some prepositions
  
  Ich gebe der Mutter ein Buch.
  mit der Mutter

In most Swiss German dialects:
- Like in Standard German
  
  Ich gib de Mueter es Buech.
  mit de Mueter

In some Central Swiss German dialects:
- Inflectional marking of dative case
- Dative case used after some prepositions
- Indirect objects are marked with a prepositional phrase (whose preposition selects dative case)
  
  Ich gib i de Mueter es Buech.
  mit de Mueter
In Standard German:

- Inflectional marking of dative case
- Dative case used for indirect objects, and after some prepositions

Ich gebe der Mutter ein Buch.
mit der Mutter

In most Swiss German dialects:

- Like in Standard German

Ich gib de Mueter es Buech.
mit de Mueter

In some Central Swiss German dialects:

- Inflectional marking of dative case
- Dative case used after some prepositions
- Indirect objects are marked with a prepositional phrase (whose preposition selects dative case)

Ich gib i de Mueter es Buech.
mit de Mueter
In Standard German:
- Inflectional marking of dative case
- Dative case used for indirect objects, and after some prepositions

Ich gebe der Mutter ein Buch.
mit der Mutter

In most Swiss German dialects:
- Like in Standard German

Ich gib de Mueter es Buech.
mit de Mueter

In some Central Swiss German dialects:
- Inflectional marking of dative case
- Dative case used after some prepositions
- Indirect objects are marked with a prepositional phrase (whose preposition selects dative case)

Ich gib i de Mueter es Buech.
mit de Mueter
Three variants (SADS I/4):

As in Std German

PDM with $i$

PDM with $a$

In the village of Einsiedeln, 82% of informants used $a$-marking, and 18% used $i$-marking.

Ich gebe der Mutter das Buch (1.0)

Ich gebe $\langle a \rangle$ der Mutter das Buch (0.82)

Ich gebe $\langle i \rangle$ der Mutter das Buch (0.18)
Map-based Machine Translation
An example: Prepositional Dative Marking

Three variants (SADS I/4):

As in Std German

PDM with \(i\)

PDM with \(a\)

In the village of Einsiedeln, 82% of informants used \(a\)-marking, and 18% used \(i\)-marking.

Ich gebe der Mutter das Buch (1.0)

\[
\text{Ich gebe } \langle a \rangle \text{ der Mutter das Buch (0.82)}
\]

\[
\text{Ich gebe } \langle i \rangle \text{ der Mutter das Buch (0.18)}
\]
1. Introduction: Languages in Switzerland
2. Machine Translation for Swiss German Dialects I
   Motivations and Architecture
3. There’s a map for that!
4. Machine Translation for Swiss German Dialects II
   Incorporating Dialectological Maps
5. Evaluation
6. Conclusion
Evaluation
Requirements for a test corpus

- Written material
- Consistent spelling if possible
- Parallel texts: Swiss German dialect — Standard German
- Different Swiss German dialects
- Annotated with author’s dialect (as precisely as possible)
Available Swiss German dialect data

- Blogs, Chat data
  - No author localization (but can sometimes be inferred from content)
  - Huge differences in spelling

- Dialect literature
  - Restricted to a few regions, no exact author localization
  - Sensible spelling rules
  - Copyright restrictions

- Swiss German Wikipedia
  - Coarse-grained author localization (~30 dialect categories)
  - Finer-grained localization might be obtained by contacting the authors directly
  - Most articles translated/adapted from Standard German articles (good/bad?)
  - Different spelling conventions, but fairly consistent

- Scientific data (transcriptions)
  - Little recent (i.e. computerized) transcriptions
  - No OCR because of diacritics
Evaluation
Building a test corpus

1. Obtain Swiss German texts with localization annotation
2. Manually translate these texts to Standard German
   - Easier to control than manually translating into Swiss German dialects
Evaluation

Test procedure I

D1 original

D2 original

D3 original

Std

Std

Std

Std

D1 MT

D1 MT

D2 MT

D2 MT

D3 MT
1. Translate the Standard German text into $n$ dialects.

2. Compare the $n$ target texts with the (one) Swiss German source text. Hypothesis: $D2_{original}$ is more similar to $D2_{MT}$ than to $D1_{MT}$ or $D3_{MT}$.
Translate the Standard German text into $n$ dialects.

Compare the $n$ target texts with the (one) Swiss German source text. Hypothesis: $D_{2_{\text{original}}}$ is more similar to $D_{2_{MT}}$ than to $D_{1_{MT}}$ or $D_{3_{MT}}$.
Evaluation

Metrics

Example:
- **Ref** poschte
- **MT1** poschtä
- **MT2** poste
- **MT3** gfrüüre

Common MT evaluation metrics are not adapted for dialects.

- **BLEU** Binary measure of word identity/difference
- **METEOR** Binary measure of word stem identity/difference
- **Levenshtein distance** Letter-based measure of word difference; bias towards constant word order when applied to whole sentences

**Solution:** Some combination of (weighted) Levenshtein distance with a traditional MT metric.
Evaluation

Metrics

Example:

Ref poschte
MT1 poschtä
MT2 poste
MT3 gfrüüre

Common MT evaluation metrics are not adapted for dialects.

BLEU  Binary measure of word identity/difference
METEOR Binary measure of word stem identity/difference
Levenshtein distance Letter-based measure of word difference; bias towards constant word order when applied to whole sentences

Solution: Some combination of (weighted) Levenshtein distance with a traditional MT metric.
Outline

1. Introduction: Languages in Switzerland
2. Machine Translation for Swiss German Dialects I
   Motivations and Architecture
3. There’s a map for that!
4. Machine Translation for Swiss German Dialects II
   Incorporating Dialectological Maps
5. Evaluation
6. Conclusion
Basic model architecture defined, database implemented

A lot of manual work to be done:
- Map selection and digitization
- Rule implementation
- Creation of evaluation corpus

Dialect MT is feasible only because the languages are closely related.
- There is always a default option: leave it as it is in Standard German
Conclusion

Further research

- Test more sophisticated interpolation methods to model dialect change.
- Currently, source language parsing and transfer are strictly separated.
  - Could a „synchronous parsing“ approach help?
  - Can synchronous parsing be adapted to > 2 languages? To a continuum?
- Implemented rules and digitized maps can be used for other purposes
  - Dialect identification
  - Dialect-to-Standard translation
  - Dialect-to-Dialect translation
- Dialect speech synthesis to respect modality-related diglossia.