Text-to-Speech (TTS) for Seven Swiss German Dialects

SwissText, June 18, 2019
Christof Traber, Schamai Safra, Bleicke Holm,
Dominic Schnyder, Philipp Lichtenberg
SlowSoft GmbH, Zurich
Text-to-Speech (TTS) for Seven Swiss German Dialects

SwissText, June 18, 2019
Christof Traber, Schamai Safra, Bleicke Holm, Dominic Schnyder, Philipp Lichtenberg
SlowSoft GmbH, Zurich
Project Background

• Presenter worked entire life on TTS systems for various languages
• It was high time to attack Swiss German dialect as native language
• Started in a commercial setup in 2015; Swiss German since 2017
• TTS is usually highly underestimated in its complexity
• Swiss German is much bigger challenge than Standard German TTS
  – dialects are not standardized, vary from place to place, even from speaker to speaker → chose one reference variant
  – no standardized way of writing; difficult to do NLP
  – few corpora available
"Grüezi, und herzlich willkommen zur Konferenz SwissText 2019 dd#zu Winterthur."

Currently:
- 6 dialects, initial quality
- 3 male, 3 female voices
Input Formats For Swiss German I: Pseudo-Phonetic Writing Style

Must be used in applications such as reading aloud of short messages

Example:
{Ich chume hüüt echli spöter, wil ich de Zug verpasst han.} (ZH)
{l chom hüt achli spööter, wäli dä Zug vopasst ha.} (SG)

Writing not standardized, depending on dialect and author

→ most difficult application, not yet started
Input Formats For Swiss German II: Full Standard German

Most desirable input form for most applications (e.g., dialog systems): can be used identically for all dialects

"Peter arbeitete gestern bis in die Nacht."

→ "Der Peter *hat* gestern bis in die Nacht *geschafft.*"

- Translation needed from Standard to Swiss German as first step (lexical substitutions, past tense -> perfect, genitive -> dative etc.)
- Translation possible by rules, but needs extremely good input analysis
Input Formats For Swiss German III: Normalized Swiss German

Use Standard German words to denote Swiss German words

"Los[dd], der Peter ist heute go[dd] posten[dd] gegangen."

→Easiest for TTS, somewhat awkward to write, but still very useful for many applications (esp. dialog systems)

However: Different variants needed for different dialects
1) "Da können Sie Ihre Bestellung aufgeben." (Non-BE, Non-VS)
2) "Hier könnt Ihr Eure Bestellung aufgeben." (BE, VS)
Envisaged Input Format For Swiss German: "Dialect-Independent Normalized Swiss German"

Minimal aim: Same input should work for all dialects (some lexical substitutions, uniform notation of different polite forms)

Currently: "Hier können#po Sie#po Ihre#po Bestellung aufgeben."

1) GR: {Do könd Sie Ihri Pstellig ufgää.} ("hier"→"da")
2) BE: {Hie schöid Dihr Eui Pstellig ufgää.} ("können Sie"→"könnt Ihr")
TTS System Structure I: End-to-End Learning

"Guten Tag, meine Damen und Herren"

Advantage:
• Avoids linguistics/phonetics

Major drawbacks:
• Needs huge corpora, for Swiss German more than for Standard German
• Correctability: How to do corrections / additions by user (e.g., how to add name pronunciations)?
• High computational load at run-time (real-time?)
TTS System Structure II: Classical Processing

"Guten Tag, meine Damen u. Herren"

Text Preprocessing

Word Analysis + G2P

Sentence Analysis / Tagging / Disambigu.

Accentuation and Prosodic Phrasing

Sentence-level Phonolog. Transformations

Text-to-Speech (TTS) for Seven Swiss German Dialects

Applied Methods

lexicon-/rule-based

lexicon-based finite-state morphology

parsing (alt.: machine learning, NN)

rule-based (alt.: machine learning, NN)

rule-based (alt.: NN)

statistical models (alt.: NN)

unit selection (alt.: HMM, NN)

Front End (language-dependent)

Phonolog. Repr.

Back End (voice-dependent)

18/6/2019 (C) SlowSoft GmbH
Classical Core NLP Part in TTS (Std. German)

Grapheme-to-phoneme (G2P) Conversion

Lexical/ morphological analysis (FST)

Word analysis variants

Sentence parsing (Disambiguation & Structural Analysis)

Morphosyntactic tree

Phonological representation
Classical Core NLP Part in TTS (Swiss German)

Grapheme-to-Phoneme (G2P) Conversion

Lexical/Morphological Analysis (FST)

Word

Word Analysis Variants

Sentence Parsing (Disambiguation & Structural Analysis)

Morphosyntactic Tree

Substitutions/Insertions in Tree

New Morphosyntactic Tree
### TTS Problems Even With Normalized Text Input

Many homographs only pose problems in Swiss German (function words!)

<table>
<thead>
<tr>
<th>Std. German</th>
<th>Swiss German</th>
</tr>
</thead>
<tbody>
<tr>
<td>&quot;meinen&quot;</td>
<td>[maɪnən]</td>
</tr>
<tr>
<td></td>
<td>[maɪnən]</td>
</tr>
<tr>
<td></td>
<td>[maɪnən]</td>
</tr>
<tr>
<td></td>
<td>[maɪnən]</td>
</tr>
</tbody>
</table>

- "meinen"  
  - (VVINF)
  - (VVFIN/pl1 or VVFIN/pl3)
  - (PPOSS/acc/m/sg)
  - (PPOSS/dat/X/pl)

<table>
<thead>
<tr>
<th>Std. German</th>
<th>Swiss German</th>
</tr>
</thead>
<tbody>
<tr>
<td>&quot;zu&quot;</td>
<td>[tsu]</td>
</tr>
<tr>
<td></td>
<td>[tsu]</td>
</tr>
<tr>
<td></td>
<td>[tsuː]</td>
</tr>
</tbody>
</table>

- "zu"  
  - (APPR)
  - (PTKZU)
  - (PTKVZ)

→ Swiss German requires much more accurate tagging than Standard German
Application Requirement: Correctability

• TTS mercilessly makes NLP errors audible, esp. with Swiss German (TTS acts like magnifying glass for NLP errors)
• Even best NLP will make errors, but TTS output must be correct

→ Correctability needed/offered:
  − markup of word category (STTS tag set)/features
    e.g. "weiss#adjd", "weiss#vvfin", "können#p1"
  − markup of Standard German use (to come)
    e.g. "Maurer#hh", "Gott sei#hh Dank"
  − in extreme case markup of phonetics via SSML (to come)
Envisaged/Demo Applications

• **Accessibility:** Speaking aids for Swiss ALS patients and reading aids for blind people

• **Language learning tools**

• **Dialog systems** in Swiss German for robotics, digital assistants, voicebots

• **Sägemol App** (free) on iOS and Android: Speech-to-speech translator from several languages to Swiss German
  - Apple/Google ASR, Google translation
  - SlowSoft TTS from Standard to SwissGerman

https://www.facebook.com/LuzernerZeitung/videos/551817122000456/
Outlook 2019

• Commercial approach not successful (so far)
  − commercial continuation not clear, but project continues
  − free use of TTS for research and medical purposes

• Voice/dialect-specific improvements
  − improved speech signal generation (using more recorded material)
  − improvements in dialect lexica

• Generic improvements
  − improved prosody control (NN approach, almost ready)
  − improved speech signal generation (transition to deep learning)
  − improved text analysis (esp. tagging)
  − improved rule-based deu->gsw translation
Thank you for your attention!

ZH: "Ich finde, das langt jetzt aber."
BA: "Genau. Also tun wir uns doch zusammen verabschieden."
All: "Vielen Dank für Ihre#po Aufmerksamkeit."