Rapid Deployment of an Afrikaans-English Speech-to-Speech Translator

Herman Engelbrecht, Tanja Schultz
Outline

• Background and Motivation
• Language Characteristics: Afrikaans
• Development Strategy
• Data Resources
• Component Development
• Integrated Speech Translation System
• Results
• Conclusions
Background – Africa

2000+ living languages
Background – South Africa

Population: 46.6 million
Official Languages: 11
Motivation

• Small HLT community in South Africa and no significant, recent MT research activity.
• S.A. government is interested in building S.A. HLT capacity (especially speech-to-speech translation).
• 1 PhD student was sent to CMU for 3 month fellowship to study speech-to-speech translation in order that S2S translation can be developed for local languages.
• Rapid deployment of S2S system for Afrikaans-English language pair was used as the vehicle for studying S2S translation.
Language Characteristics: Afrikaans

• Afrikaans is a Germanic language and linguistically closely related to Dutch.
• Afrikaans has a more regular grammar than Dutch and the grammar is very analytic.
• Afrikaans text is written using Latin alphabet plus a few diacritics. Afrikaans spelling is more phonetic than Dutch.
• Words are separated by spaces in text – no need for a word boundary segmenter and SMT algorithms can be readily applied.
• 62 phones typically used in spoken Afrikaans.
Development Strategy

• The choice of recognition, translation and synthesis strategies were influenced by the amount of time and labor-intensive work required to implement strategy.

• Data-driven techniques preferred over knowledge-based techniques and the following strategies were adopted:
  • Recognition – SLM based recognition strategy.
  • Translation – Statistical machine translation.
  • Synthesis – Concatenative synthesis.

• The focus was on the development of the ASR, MT and TTS components for the new language.
Development Strategy

- Needed to develop/obtain the following subcomponents for Afrikaans:
  - SMT: Translation Models and Language Models.
  - TTS: Pronunciation Lexicon and Letter-to-Sound Rules.
- For English existing ASR and TTS components developed by CMU were used.
- The domain of the system was constrained by the available data resources to be on parliament debates (Hansards).
Data Resources

- **Text Data:**
  - Parallel Afrikaans-English text corpus (Hansards).

- **Speech Data:**
  - Afrikaans speech data from AST speech corpus (based on SpeechDat corpus).
  - Hansard speech data recorded during fellowship.

- **Pronunciation Lexicon:**
  - 5k lexicon obtained from AST speech corpus (includes pronunciation variants).
  - 37k lexicon University of Stellenbosch (no pronunciation variants, but includes syllable markers).
Data Resources

• **Parallel Text Corpus (Hansards):**
  • 43k parallel sentences.
  • ± 700k words per language.
  • ± 20k vocabulary per language.

• **AST speech data (out-of-domain):**
  • 72% Telephone, 28% Mobile phone.
  • 57% Female, 43% Male.
  • Roughly 6 hours of transcribed speech.
  • 265 speakers, ±40 utterances per speaker.

• **Hansard speech data (in-domain):**
  • 1000 prompted utterances recorded on laptop by two native Afrikaans speakers (male and female).
  • Utterances chosen from parallel text corpus.
Component Development - ASR

- Bootstrapped acoustic models from Global-Phone 7-lingual models using Janus JrTK.
- 39 phone models, 1 silence model:
  - 13 vowels, 26 consonants, no diphthongs. No distinction between long and short vowels.
- Fully continuous 3-state HMM recogniser:
  - 500 triphone models (tied using decision trees).
  - 128 Gaussian per state.
  - 13 MFCCs, power, and first and second derivates are reduced to 32 dimensions using LDA.
  - Trained with VTLN and SAT.
- Training data (out-of-domain):
  - 187 speakers, 7696 utterances.
Component Development - ASR

- Hansard Adaptation data (in-domain):
  - 200 utterances, 2 speakers.
- Hansard Evaluation data (in-domain):
  - 800 utterances, 2 speakers.

<table>
<thead>
<tr>
<th></th>
<th>Unadapted AMs</th>
<th>Adapted AMs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of words</td>
<td>15,259</td>
<td>15,259</td>
</tr>
<tr>
<td>Vocabulary size</td>
<td>2,450</td>
<td>2,450</td>
</tr>
<tr>
<td>Pronunciation variants</td>
<td>1.08</td>
<td>1.08</td>
</tr>
<tr>
<td>Trigam LM perplexity</td>
<td>103.71</td>
<td>103.71</td>
</tr>
<tr>
<td>WER (male)</td>
<td>39.1%</td>
<td>17.6%</td>
</tr>
<tr>
<td>WER (female)</td>
<td>54.0%</td>
<td>22.3%</td>
</tr>
<tr>
<td>WER (total)</td>
<td>46.5%</td>
<td>20.0%</td>
</tr>
</tbody>
</table>
Component Development - SMT

- PESA used for training.
  - IBM1 model.
  - Trigram LMs trained using SriLM software.
  - Trained both Afrikaans-English and English-Afrikaans translation models.
  - Experimented with punctuation included and with punctuation removed from text.

- Hansard Parallel Text Corpus:
  - Train set: 41,239 utterances.
  - Test set: 800 utterances (same as used for ASR).
  - Sentences aligned using Europarl sentence aligner.
### Component Development - SMT

<table>
<thead>
<tr>
<th>Text Data Language</th>
<th>English</th>
<th>Afrikaans</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of Sentences</td>
<td></td>
<td>41,239</td>
</tr>
<tr>
<td>Number of Words</td>
<td>687,154</td>
<td>694,455</td>
</tr>
<tr>
<td>Vocabulary size</td>
<td>17,898</td>
<td>25,623</td>
</tr>
<tr>
<td>LM Perplexity w/o punct.</td>
<td>87.21</td>
<td>103.71</td>
</tr>
<tr>
<td>LM Perplexity with punct.</td>
<td>62.28</td>
<td>72.28</td>
</tr>
</tbody>
</table>

- Europarl Dutch-English with IBM4 translation model was used for comparison as the language pairs and domain are very similar.
Component Development - SMT

<table>
<thead>
<tr>
<th>Results</th>
<th>Afrikaans-English</th>
<th></th>
<th>English-Afrikaans</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>BLEU</td>
<td>NIST</td>
<td>BLEU</td>
<td>NIST</td>
</tr>
<tr>
<td>IBM1 w/o punctuation</td>
<td>34.13</td>
<td>7.65</td>
<td>34.68</td>
<td>7.93</td>
</tr>
<tr>
<td>IBM1 with punctuation</td>
<td>36.11</td>
<td>7.66</td>
<td>34.81</td>
<td>7.73</td>
</tr>
</tbody>
</table>

- Dutch-English with 740k Europarl corpus

<table>
<thead>
<tr>
<th>Results</th>
<th>Dutch-English</th>
<th></th>
<th>English-Dutch</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>BLEU</td>
<td>NIST</td>
<td>BLEU</td>
<td>NIST</td>
</tr>
<tr>
<td>IBM4</td>
<td>26.35</td>
<td>-</td>
<td>22.85</td>
<td>-</td>
</tr>
</tbody>
</table>
Component Development - TTS

- Festival was used to build a male Afrikaans voice:
  - Unit-selection voice.
  - Trained Letter-to-sound rules.
  - Binding of units for unit-selection voice.
  - Phone set is identical to ASR phone set.
- 500 Hansard-domain utterances were used to train voice.
- Afrikaans pronunciation lexicon:
  - 37k vocabulary size.
  - No pronunciation variants.
  - Syllables are marked.
Component Development - TTS

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Train set pronunciations</td>
<td>33,121</td>
</tr>
<tr>
<td>Train set pronunciations</td>
<td>3,680</td>
</tr>
<tr>
<td>Phones correct</td>
<td>97.92%</td>
</tr>
<tr>
<td>Words correct</td>
<td>85.24%</td>
</tr>
</tbody>
</table>

- LTS results comparable to German (89.38% word correct).
- It is difficult to formally evaluate Afrikaans TTS with only 2 native speakers (especially if one is the developer).
- Informal evaluation was performed by simply listening to pronunciations to determine their correctness.
Integrated Translation System

• Description:
  • Based on One4All demo scripts developed by ISL.
  • Best ASR output used as SMT input.
  • Re-used existing English ASR and TTS.

[Image of software interface showing translation between Afrikaans and English]
<table>
<thead>
<tr>
<th>System Input</th>
<th>WER</th>
<th>Afrikaans-English</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>NIST</td>
<td>BLEU</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>SCORE</td>
<td>Rel. Imp.</td>
<td>SCORE</td>
</tr>
<tr>
<td>TEXT w/o punct.</td>
<td>0.0%</td>
<td>7.65</td>
<td>-</td>
<td>34.13</td>
</tr>
<tr>
<td>ASR w/o punct. (Adapted AMs)</td>
<td>20.0%</td>
<td>6.12</td>
<td>-20.0%</td>
<td>25.45</td>
</tr>
<tr>
<td>ASR w/o punct. (Unadapted AMs)</td>
<td>46.5%</td>
<td>4.56</td>
<td>-40.4%</td>
<td>17.39</td>
</tr>
<tr>
<td>TEXT with punct.</td>
<td>0.0%</td>
<td>7.66</td>
<td>-</td>
<td>36.11</td>
</tr>
<tr>
<td>ASR with punct. (Adapted AMs)</td>
<td>20.0%</td>
<td>6.04</td>
<td>-21.1%</td>
<td>24.42</td>
</tr>
<tr>
<td>ASR with punct. (Unadapted AMs)</td>
<td>46.5%</td>
<td>4.40</td>
<td>-42.6%</td>
<td>16.72</td>
</tr>
</tbody>
</table>
Results

Afrikaans-English S2S translation results

Adapted AMs

Unadapted AMs

IBM 1 with punct.

IBM 1 w/o punct.

IBM 1 with punct.

IBM 1 w/o punct.

Carnegie Mellon
Results - Example translation

- Reference sentence:
  “Firstly the lack of nursing staff remains a problem.”

- Source sentence:
  “Ten erste bly die gebrek aan verpleeg personeel ’n probleem.”

- Recognised sentence:
  “Ten eerste by gebrek aan verpleeg personeel probleem.”

- Machine Translation of recognised sentence:
  “Firstly at the lack of nurses problem.”

- Machine Translation of source sentence:
  “Firstly I am glad the lack of nurses a problem.”
## Conclusions – Development Time

<table>
<thead>
<tr>
<th>Component</th>
<th>Development Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>Speech recogniser</td>
<td>8</td>
</tr>
<tr>
<td>Machine translator</td>
<td>1</td>
</tr>
<tr>
<td>Speech synthesis</td>
<td>1</td>
</tr>
<tr>
<td>Integrated System</td>
<td>1</td>
</tr>
<tr>
<td>Evaluation</td>
<td>1</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>12</strong></td>
</tr>
</tbody>
</table>

- SMT in a week – Yes
- Speech-to-speech translation in a week - No
Conclusions

• Demonstrated rapid deployment of S2S translation system under somewhat idealised conditions as most of the data and development tools were readily available.

• Recognition component is still the most challenging component to develop for a new language as evidenced by 20% WER.

• Afrikaans-English SMT results very encouraging when compared to Dutch-English as only a simple translation model was used.

• As expected, errors in recognition degrades the translation.
Future work

• Use more sophisticated translation modelling and schemes.
• Develop local SMT software.
• Start looking at other local language pairs:
  • isiXhosa - English
  • Sepedi – English
• Challenges:
  • Ntu languages are very different from the Germanic languages.
  • Ntu languages only been written languages for ±150 years.
## Component Evaluation - ASR

<table>
<thead>
<tr>
<th></th>
<th>Unadapted AMs</th>
<th>Adapted AMs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of words</td>
<td>15,259</td>
<td>15,259</td>
</tr>
<tr>
<td>Vocabulary size</td>
<td>2,450</td>
<td>2,450</td>
</tr>
<tr>
<td>Pronunciation variants</td>
<td>1.08</td>
<td>1.08</td>
</tr>
<tr>
<td>Trigam LM perplexity</td>
<td>103.71</td>
<td>103.71</td>
</tr>
<tr>
<td>WER (male)</td>
<td>39.1%</td>
<td>17.6%</td>
</tr>
<tr>
<td>WER (female)</td>
<td>54.0%</td>
<td>22.3%</td>
</tr>
<tr>
<td>WER (total)</td>
<td>46.5%</td>
<td>20.0%</td>
</tr>
</tbody>
</table>
Background – South Africa

Population: 46.6 million
Official Languages: 11