The BBN RT03 BN Arabic System

Long Nguyen, John Makhoul, Mohamed Noamany

RT-03 Workshop
Boston, MA, May 19-21, 2003

Overview

• Development test set
• Improvements
• Evaluation result
Language-Specific Treatment

- Transcripts: Used a modified Buckwalter format
  For example: ‘JxbAr ywm kAml nJtykm bhA fy brnAmjnA AlMxbAry Al$Aml Al$AEp AlrAbEp wAlE$rwn…’

- ‘Phonetic’ dictionary: Used grapheme-phoneme, one-to-one mapping rule

- High OOV rate mainly due to affixes
  – ~5% OOV rate for a 65k-word lexicon

Development Test Set

- Selected four episodes from the two audio sources in the TDT4 Arabic corpus
  - NTV: Nile TV shows from Cairo, Egypt
  - VOA: Uncle Sam’s radio show
  - Broadcast in the second half of Jan ’01
  - First 30 minutes from each episode (~2 hours)

<table>
<thead>
<tr>
<th></th>
<th>NTV</th>
<th>VOA</th>
<th>All</th>
</tr>
</thead>
<tbody>
<tr>
<td>Baseline (GI, 1xRT)</td>
<td>28.6</td>
<td>20.9</td>
<td>24.6</td>
</tr>
</tbody>
</table>
**Improvements**

- Used GD acoustic models (36h of LDC data), MLLR adaptation, and ran at 10xRT
- LM: 65k-word lexicon, 15M bigrams, 33M trigrams, trained on mostly in-house data
- Added 38h of TDT4 data

<table>
<thead>
<tr>
<th></th>
<th>NTV</th>
<th>VOA</th>
<th>All</th>
</tr>
</thead>
<tbody>
<tr>
<td>0. Baseline</td>
<td>28.6</td>
<td>20.9</td>
<td>24.6</td>
</tr>
<tr>
<td>1. GD, adaptation, 10x</td>
<td>23.6</td>
<td>16.8</td>
<td>20.0</td>
</tr>
<tr>
<td>2. + TDT4 data (38h)</td>
<td>22.3</td>
<td>14.5</td>
<td>18.2</td>
</tr>
</tbody>
</table>

**Evaluation Results**

- System ran at 7xRT
- Evaluation result (26.3%) is significantly higher than the result on the dev test
  - Self-score using NIST’s reference is 26.7%
  - There are problems in the reference
- [Adjudication ongoing]
Summary

• Achieved good WER reduction on the development set
• Need to agree on a transcription standard for Arabic