Principle-Based Preparation of Authentic Bilingual Text Resources

Michelle Vanni, Ph.D.
Army Research Laboratory
1. The context: Speech recognition for military

2. The research questions: Where does the material fit?

3. The problem: Material and task description

4. The principles: Constraints on organization

5. The examples: What you would do & why
Vision and Objective

• Automated Speech Recognition (ASR) technology trained on authentically accented data for operations
• High quality ASR for military-relevant languages spoken in operational scenarios
• Algorithms adapting general purpose ASR technology to military operational needs

Problems Being Addressed

• Algorithms to adapt ASR to new types of variation
• Expeditionary Force: local populace & coalition partners
• Army Challenges
  1 Situation Awareness: Adversary intent & capabilities
  2 Security Force Assistance: HQ ASR for effectiveness

Impact

• Understand foreign media and captured document content
• HLT-equipped soldiers: Train & serve with coalition partners
• Focus: Variations of high military, low commercial value
Given the modest amounts of bilingual in-domain speech data available, which approaches to Automatic Speech Recognition (ASR) adaptation have the most impact on language modeling for Army-centric technologies?

Can ASR software components and algorithms be trained to achieve better performance with African-Accented speakers?

Is it possible to generalize—and to what extent—ASR adaptation algorithms designed to address individual speaker differences, over sets of non-native pronunciations present in communities?
**Strategy:**

1. Improve techniques for ASR adaptation on bi-modal—speech-text aligned—accented data.
3. Assess for general purpose tech to process the speech of African accented high-resource languages

**Technical Barriers:**

1. Valuable on individual non-native variations, maximum likelihood linear regression (MLLR) & maximum a posteriori (MAP) adaptation likely improve high-resource language ASR on similar variations in accented speakers, e.g., French and accented French: However experimentation for specific Army operations is required.

**Approach:**

1. Experiment with language modeling software offering Deep Neural Network technology on compiled parallel aligned data sets for low resource languages of military interest.
2. Test a new unsupervised morphological analyzer on Pashto data compiled in domains of military interest.
3. Use speech data collected in Cameroon and Gabon to test accuracy of a French speech recognizer with one type of African-accented speech. Adapt the French speech recognizer with a modest amount of Cameroon-accented French and compare accuracy using Word Error Rate.
PROBLEM

- Lots of raw data in an operational format

- In this case, Power Point slides
  - Could be bilingual web data or digitized books

- Bi-text needed for multiple purposes
  - Speech recognition pronouncing dictionary
  - Machine translation domain adaptation
  - Glossaries and Translation Memory

First Step: Change the format!
  - Find a suitable editing environment
  - In this case, MS Excel worked fine
A SAMPLE OF THE TEXT EDITING PRINCIPLES:

1. Each row represents a single semantic unit, word, phrase or clause and can be simple or complex.

2. Punctuation:
   a. Periods only after full clauses, with or without grammatical subjects and where appropriate by convention.
   b. Commas as appropriate
   c. Usually delete colons, except when required on the basis of content.

3. Capitalization: As appears in raw text for application-specific pre-processing

.... AND SO ON
A SAMPLE OF THE TEXT EDITING PRINCIPLES:

4. **Insertion**: Without substitution of original material, syntactic support structures for creation of a corpus usable for training machine translation of genres other than the genre of provenance.

5. **Insertion**: Conventionally accepted orthographic forms, without substitution of forms presented in original data.

6. **Insertion**: Without substitution of original material, of semantically accurate and similarly structured translation, when given dynamic equivalent rendering is structurally divergent.

... AND SO ON
### Examples

<table>
<thead>
<tr>
<th>EEIFA</th>
<th>EEFI</th>
</tr>
</thead>
<tbody>
<tr>
<td>les éléments essentiels d'information des forces amies</td>
<td>essential elements of friendly information</td>
</tr>
<tr>
<td>ce que nous voulons cacher de la menace</td>
<td>what we want to hide from the threat</td>
</tr>
<tr>
<td>la SECOP</td>
<td>OPSEC</td>
</tr>
<tr>
<td>COA</td>
<td>COA</td>
</tr>
<tr>
<td>cause de l’action</td>
<td>cause of action</td>
</tr>
<tr>
<td>zone géographique</td>
<td>geographic area</td>
</tr>
<tr>
<td>où l'information pour répondre à un EIP peut être recueillie</td>
<td>where information to answer a PIR can be collected</td>
</tr>
<tr>
<td>zone géographique où l'information pour répondre à un EIP ou confirmer / refuser un COA de la menace peut être recueillie</td>
<td>geographical area where information to answer a PIR or confirm/deny a threat COA can be collected</td>
</tr>
</tbody>
</table>
Questions?

Contact us at:

michelle.t.vanni.civ@mail.mil