Expanding a Lexicon to be Multi-Lingual Using Cognates

Bryan Wilkinson

**Problem**

How to make a static knowledge resource multi-lingual using automatic or semi-automatic techniques

**Linguistic Background**

- Historical Linguistics
- Word Relations
  - Cognates
  - Loan-Words

<table>
<thead>
<tr>
<th>Language</th>
<th>Word</th>
</tr>
</thead>
<tbody>
<tr>
<td>Latin</td>
<td>campus</td>
</tr>
<tr>
<td>Spanish</td>
<td>campo</td>
</tr>
<tr>
<td>Italian</td>
<td>campo</td>
</tr>
<tr>
<td>French</td>
<td>champ</td>
</tr>
</tbody>
</table>

**OntoSem**

- Text Processing System
- Ontology (Language Independent)
- Lexicon (Language Dependent)

**Bootstrapping The Lexicon with Cognates**

- Associate cognates with same concept as word in L1
- Use syntactic information such as part of speech to get arguments for word
- Use lexical restrictions found in ontology to associate concepts for new words

**Evaluation**

- For cognate identification use precision and recall
- For Bootstrapping, look only at precision

**Future Work**

- Expand to use loan words
- Continue looking for better cognate recognition

**Related Work**

- Various Merging Algorithms
  - BiFrameNet
  - Padó’s FrameNet Work

**Cognate Recognition**

- Phonetic Based
- Orthographically Based

**Cysouw’s cognate alignment**

<table>
<thead>
<tr>
<th>English</th>
<th>Spanish</th>
<th>Italian</th>
<th>French</th>
</tr>
</thead>
<tbody>
<tr>
<td>photo</td>
<td>foto</td>
<td>vibration</td>
<td>vibracion</td>
</tr>
<tr>
<td>tourism</td>
<td>turismo</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>