FrameNet translation
using bilingual dictionaries
with evaluation on the English-French pair

Claire.Mouton@gmail.com
Gael.de-Chalendar@cea.fr
Benoit.Richert@student.ecp.fr
Agenda

• Introduction
• Proposed approach
• Evaluation
• Resource enrichment
• Conclusions
FrameNet: a resource for Semantic Role Labeling

- Semantic Role Labeling (SRL)
  - Detect and identify **predicate** of a given situation
  - Detect and identify **roles** of a given situation
  - Aims at helping Textual entailment, Question-Answering systems...

- **FrameNet**
  - Language: English
  - Structure: Frame = set of triggering predicates + set of specific roles
  - Number of predicate-frame pairs: more than 10,000
  - Number of roles: 250 (specific subset for each frame)

- Example

  **Attempt_suasion**
  
  [Advise, beg, discourage, encourage, exhort, press, urge (...)]

  [A number of embassies]_{SPEAKER} are warning [their citizens]_{ADRESSEE} [against traveling to Thailand's capital]_{CONTENT}. 
• **Real need for other languages than English**
  - **Case of French**
    - Volem [Fernandez et al., 02]
      * Semantic resource for French, Spanish and Catalan
      * 1,500 verbs
      * ~20 generic semantic roles
    - **Comparison to FrameNet**
      * Much lower coverage
      * Less specific roles
      * Only verbs, no other part-of-speech
      * Entries are verbs (and not sets of predicates grouped by "senses" as in FrameNet)
    - FrameNet transposition to French [Pado and Pitel, 07]
      * ~7000 predicate-frame pairs
      * Precision 77%
• Introduction
• Proposed approach
• Evaluation
• Resource enrichment
• Conclusions
Overview of the proposed method

• For each frame and each predicate in this frame
  ▪ Extraction of translation pairs from bilingual dictionaries
  ▪ Base score representing the confidence we have in the translation of the given predicate in the given frame
  ▪ 5 variations of this score based on different heuristics

• Linear combination of the scores

• Filtering with a parameter threshold

• Run with different parameters and weights on a development set to find the best settings
Extraction of translation pairs

- **Bilingual dictionaries we use in our experiments**
  - Wiktionary
    - Creative Commons license
    - 27,109 French-English translation pairs in January 2009 version
    - Distinction of senses for some of the translations
  - EuRADic
    - Distributed by ELDA
    - 243,539 entries

- **Extraction of translation pairs**
  - English Lexical Unit (LU) present in predicates of a frame → French Lexical Unit(s) (LU)
  - 2 different resources by dividing EuRADic and Wiktionary results
• **Score S1: redundancy of translations**
  - If many English LU of the same frame translate to the same French LU → confidence for the translation to be correct is high.
  - French LU-Frame score=Nb of translation pairs for the LU in the given frame
  - If a translation pair is found in several sense distinctions in the Wiktionary, they are all summed up.
  - Example:
    - **Ingestion**
      ```
      remettre.v {put back.v:1} 1
      boire.v {quaff.v:1, drink.v:2} 3
      alimenter.v {feed.v:1} 1
      déjeuner.v {lunch.v:1, dine.v:1, feed.v:1, eat.v:1} 4
      ...
      ```

  Wiktionary
  - *consume liquid through the mouth*
    - `drink.v` → `boire.v`
  - *consume alcoholic beverages*
    - `drink.v` → `boire.v`
Structural Scores I

- **Structural score S2: polysemy of source LU**
  - **Hypothesis**
    - Polysemous source LU (present in more than one frame) → higher risk that translation is erroneous
    - $S_2 = \text{confidence score } S_1$ lowered depending on the number of frames containing the source LU
  - **Example**
    - *rise* appears in 9 different frames
      - **Getting_up**
        - get up → se lever
        - rise → augmenter
        - → se lever

Se lever: $S_1 = 2$, $S_2 = 2/10^\alpha$
Augmenter: $S_1 = 1$, $S_2 = 1/9^\alpha$
• **Structural score S3: number of English LUs in the frame**
  ▪ **Hypothesis**
    ▪ Source frame contains lots of LUs → higher risk that redundant translations appear
    ▪ $S3 = \text{confidence score } S1 \text{ lowered depending on the number of source LUs in the given frame}$
  ▪ **Example**
    ▪ **Containers** has 116 English LUs
      \(bac.n\) is the French translation of 15 of the English LUs
      (\textit{WRONG}) \(nigaud.n \leftarrow \text{mug}\) is the French translation of 1 English LU
    ▪ **Operational\_testing** has 8 English LUs
      \(tester.v\) is the French translation of 1 of the English LUs
      \[
      \begin{align*}
      \text{bac\_Containers : } & S1 = 23 & S3 = 15/116^\alpha \\
      \text{nigaud\_Containers : } & S1 = 1 & S3 = 1/116^\alpha \\
      \text{tester\_Operational\_testing : } & S1 = 1 & S3 = 1/8^\alpha
      \end{align*}
      \]
**Target score S4: number of translation pairs**

- **Hypothesis**
  - High number of translation pairs
    - higher risk that redundant translations appear
  - S4 = confidence score S1 lowered depending on the number of translation pairs for the given frame

- **Example**
  - Same idea as previous score
• **Target score S5: number of LUs in the target frame**
  - Hypothesis
    - Target frame contains lots of LUs
      - Some LUs may carry slightly different meanings
    - $S_5 = \text{confidence score } S_1 \text{ lowered depending on the number of target LUs in the given frame}$

• **Target score S6: polysemy of the target LU**
  - Hypothesis
    - Polysemous target LU (present in more than one frame)
      - LU less informative in the given frame
    - $S_6 = \text{confidence score } S_1 \text{ lowered depending on the number of frames containing the target LU}$
  - Example
    - *Prendre* appears in 83 frames and *Porter* appears in 75 frames
Agenda

- Introduction
- Proposed approach
- Evaluation
- Resource enrichment
- Conclusions
Experimental setup

- **Evaluation criteria**
  - Precision, Recall, $F_{0.5}$-measure
  - Computed on each frame and averaged

- **Two FrameNet subsets**
  - Obtained from the union of FrameNet.FR [Pado and Pitel, 07], unfiltered translations with EuRADic and with Wiktionary
  - Subset 1: Development set
    - Sample of 10 frames: Nb of LUs representative of the global distribution (quantiles)
    - Manually corrected
  - Subset 2: Test set
    - Sample of 10 frames: the ones used by [Pado and Pitel, 07]
    - Manually corrected

- **Scores combination and parameter settings**
  - Normalization and linear combination
  - Maximization of recall at $P_{0.95}$ and maximization of $F_{0.5}$-measure
<table>
<thead>
<tr>
<th>Resource</th>
<th>Linear combination</th>
<th>All frames</th>
<th>Test Set</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>#LU-Frame</td>
<td>#Frames</td>
<td>P</td>
</tr>
<tr>
<td>Berkeley FrameNet</td>
<td>11,171</td>
<td>796</td>
<td></td>
</tr>
<tr>
<td>FrameNet.FR [Pado and Pitel, 07]</td>
<td>6,659</td>
<td>480</td>
<td>77%</td>
</tr>
<tr>
<td>Wi_nofilter</td>
<td>19,912</td>
<td>781</td>
<td>70%</td>
</tr>
<tr>
<td>Wi_P_{095}</td>
<td>$\frac{1}{4}S2 + \frac{1}{4}S5 + \frac{1}{2}S6$</td>
<td>2,889</td>
<td>686</td>
</tr>
<tr>
<td>Wi_F_{0.5} max</td>
<td>$\frac{1}{4}S1 + \frac{1}{2}S4 + \frac{1}{4}S6$</td>
<td>15,720</td>
<td>781</td>
</tr>
<tr>
<td>Eu_nofilter</td>
<td>57,787</td>
<td>795</td>
<td>58%</td>
</tr>
<tr>
<td>Eu_P_{095}</td>
<td>$\frac{3}{4}S2 + \frac{1}{4}S6$</td>
<td>616</td>
<td>210</td>
</tr>
<tr>
<td>Eu_F_{0.5} max</td>
<td>$\frac{1}{4}S2 + \frac{3}{4}S6$</td>
<td>24,885</td>
<td>767</td>
</tr>
<tr>
<td>Wi_F_{0.5} max U Eu_F_{0.5} max</td>
<td>34,121</td>
<td>793</td>
<td>70%</td>
</tr>
<tr>
<td>Wiktionary \ EuRADic</td>
<td>12,211</td>
<td>773</td>
<td>82%</td>
</tr>
<tr>
<td>Wi_F_{0.5} max \ Eu_F_{0.5} max</td>
<td>6,484</td>
<td>724</td>
<td>95%</td>
</tr>
</tbody>
</table>
• Introduction
• Proposed approach
• Evaluation
• **Resource enrichment**
• Conclusions
Enrichment by similarity

• **Resources used to perform the enrichment**
  - Semantic spaces computed with MI on syntactical co-occurrences
  - Cosine similarity

• **Classification of nouns**
  - Classes ↔ frames
  - Learning data ↔ set of triggering Lus of each frame
  - K-NN classifier on multi-represented data [Kriegel et al, 05]
  - In every semantic space, weights the confidence on the neighbors by taking into account density of neighbors belonging to the same class

• **Variation of parameters**
  - K: 10, 25, 50
  - Filter thresholds
  - Selection of semantic spaces
  - Use of the size of the classes in confidence vector
  - Use of the translation score S1 into the learning process
### Setting parameters

- Optimizing precision / coverage against union of three resources:
  - FrameNet.FR [Pado and Pitel, 07]
  - Translation using Wiktionary
  - Translation using EuRADic

### Results

<table>
<thead>
<tr>
<th>Resource</th>
<th>All frames</th>
<th>Test set</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>#LU-frame</td>
<td>#New attributions</td>
</tr>
<tr>
<td>Berkeley FrameNet</td>
<td>11,171</td>
<td>7,581 (79%)</td>
</tr>
<tr>
<td>FN.1 precision</td>
<td>9,536</td>
<td></td>
</tr>
<tr>
<td>FN.2 coverage</td>
<td>27,371</td>
<td>24,997 (91%)</td>
</tr>
<tr>
<td>TFN+EFN.1</td>
<td>15,132</td>
<td>8,648 (57%)</td>
</tr>
</tbody>
</table>

### Comments

- TFN + EFN.1 = (Wi_F\_0.5 max \( \cap \) Eu_F\_0.5 max) \( \cup \) FN.1
- Combined resource: 15,132 pairs with an estimated precision of 86%
Agenda

• Introduction
• Proposed approach
• Evaluation
• Resource enrichment
• Conclusions
Conclusions and future work

• New approach to transfer FrameNet into another language
  ▪ Validated for French

• Resources resulting from translation
  ▪ A robust one: 95% estimated precision - 58% of BerkeleyFN size
  ▪ A balanced one: 70% estimated precision – 3 times BerkeleyFN size

• Enrichment
  ▪ Performed on nouns
  ▪ Significant results incite to go further with verbs and adjectives

• Future work
  ▪ Try to apply the translation method to the heads of the phrases filling the different roles in order to build learning data for a SRL system.
Questions
• **Approaches with bilingual corpora**
  - German: [Pado and Lapata, 05]
  - French: [Pado and Pitel, 07]
  - Italian: [Tonelli and Pianta, 08], [Basili et al.09]

• **Approaches with bilingual dictionaries and filtering**
  - Chinese: [Fung and Chen, 04]
<table>
<thead>
<tr>
<th>Resource</th>
<th>$\alpha$</th>
<th>$P$</th>
<th>$R$</th>
<th>$F_{0.5}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wiktio</td>
<td></td>
<td>63%</td>
<td>40%</td>
<td>53%</td>
</tr>
<tr>
<td>Wiktio+S1 $F_{0.5}max$</td>
<td>1</td>
<td>63%</td>
<td>40%</td>
<td>53%</td>
</tr>
<tr>
<td>Wiktio+S2 $F_{0.5}max$</td>
<td>1</td>
<td>65%</td>
<td>40%</td>
<td>54%</td>
</tr>
<tr>
<td>Wiktio+S3 $F_{0.5}max$</td>
<td>1</td>
<td>63%</td>
<td>40%</td>
<td>53%</td>
</tr>
<tr>
<td>Wiktio+S4 $F_{0.5}max$</td>
<td>0.5</td>
<td>66%</td>
<td>38%</td>
<td>53%</td>
</tr>
<tr>
<td>Wiktio+S5 $F_{0.5}max$</td>
<td>0.75</td>
<td>66%</td>
<td>38%</td>
<td>53%</td>
</tr>
<tr>
<td>Wiktio+S6 $F_{0.5}max$</td>
<td>1</td>
<td>70%</td>
<td>36%</td>
<td>55%</td>
</tr>
<tr>
<td>EuRADic</td>
<td></td>
<td>51%</td>
<td>93%</td>
<td>56%</td>
</tr>
<tr>
<td>EuRA+S1 $F_{0.5}max$</td>
<td>0.25</td>
<td>74%</td>
<td>34%</td>
<td>58%</td>
</tr>
<tr>
<td>EuRA+S2 $F_{0.5}max$</td>
<td>0.75</td>
<td>74%</td>
<td>34%</td>
<td>58%</td>
</tr>
<tr>
<td>EuRA+S3 $F_{0.5}max$</td>
<td>0.25</td>
<td>69%</td>
<td>51%</td>
<td>59%</td>
</tr>
<tr>
<td>EuRA+S4 $F_{0.5}max$</td>
<td>0.1</td>
<td>71%</td>
<td>46%</td>
<td>60%</td>
</tr>
<tr>
<td>EuRA+S5 $F_{0.5}max$</td>
<td>0.25</td>
<td>71%</td>
<td>46%</td>
<td>60%</td>
</tr>
<tr>
<td>EuRA+S6 $F_{0.5}max$</td>
<td>0.25</td>
<td>68%</td>
<td>55%</td>
<td>64%</td>
</tr>
<tr>
<td>Resource</td>
<td>Linear combination</td>
<td>All frames</td>
<td>Test set</td>
<td></td>
</tr>
<tr>
<td>-----------------------</td>
<td>--------------------</td>
<td>------------</td>
<td>----------</td>
<td></td>
</tr>
<tr>
<td></td>
<td>#LU-frames</td>
<td>#Frames</td>
<td>$P$</td>
<td>$R$</td>
</tr>
<tr>
<td>Berkeley FrameNet</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wi_nofilter</td>
<td>11,171</td>
<td>796</td>
<td>70%</td>
<td>33%</td>
</tr>
<tr>
<td>Wi_P095</td>
<td>$\frac{1}{4}.S2 + \frac{1}{4}.S5 + \frac{1}{2}.S6$</td>
<td>19,912</td>
<td>781</td>
<td>94%</td>
</tr>
<tr>
<td>Wi_F05max</td>
<td>$\frac{1}{4}.S1 + \frac{1}{2}.S4 + \frac{1}{4}.S6$</td>
<td>2,889</td>
<td>686</td>
<td>74%</td>
</tr>
<tr>
<td></td>
<td>15,720</td>
<td>781</td>
<td></td>
<td></td>
</tr>
<tr>
<td>EuRADic_nofilter</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>EuRADic_P095</td>
<td>$\frac{3}{4}.S2 + \frac{1}{4}.S6$</td>
<td>57,787</td>
<td>795</td>
<td>58%</td>
</tr>
<tr>
<td>EuRADic_F05max</td>
<td>$\frac{1}{4}.S2 + \frac{1}{4}.S6$</td>
<td>616</td>
<td>210</td>
<td>100%</td>
</tr>
<tr>
<td></td>
<td>24,885</td>
<td>767</td>
<td></td>
<td></td>
</tr>
<tr>
<td>FrameNet.fr_nofilter</td>
<td>6,659</td>
<td>480</td>
<td>77%</td>
<td>23%</td>
</tr>
<tr>
<td>Union</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wi $\cup$ EuRADic</td>
<td>65,488</td>
<td>796</td>
<td>57%</td>
<td>92%</td>
</tr>
<tr>
<td>$W_{-P_{0.95}} \cup E_{-P_{0.95}}$</td>
<td>3,256</td>
<td>695</td>
<td>94%</td>
<td>12%</td>
</tr>
<tr>
<td>$W_{-F_{0.5\max}} \cup E_{-F_{0.5\max}}$</td>
<td>34,121</td>
<td>793</td>
<td>70%</td>
<td>59%</td>
</tr>
<tr>
<td>Intersection</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wi $\cap$ EuRADic</td>
<td>12,211</td>
<td>773</td>
<td>82%</td>
<td>25%</td>
</tr>
<tr>
<td>$W_{-F_{0.5\max}} \cap E_{-F_{0.5\max}}$</td>
<td>6,484</td>
<td>724</td>
<td>95%</td>
<td>15%</td>
</tr>
<tr>
<td>$Wi_{-F_{0.5\max}} \cap Eu_{-F_{0.5\max}} \cap FN.fr$</td>
<td>7,814</td>
<td>742</td>
<td>95%</td>
<td>18%</td>
</tr>
</tbody>
</table>