BabelNet and Word Sense Disambiguation

Overview:
• Original BabelNet
• BabelNet 2.5 (‘today’)
• Extrinsic Evaluations (SemEval-2007 T#16, SemEval-2007 T#7)
• SemEval-2010 T#3, 2013

Next episode (preview):
• Babelfy (an online, unified graph-based approach to EL and WSD)
Large and wide coverage multilingual semantic network;
Integrates lexicographic and encyclopaedic knowledge;
Further enriched by Machine Translation;
Coverage for 50 languages;
+9 million entries;

How? With the automatic integration of:
WordNet, OMW (☺), Wikipedia, OmegaWiki, Wiktionary, and Wikidata + SMT of senses across languages
An “Encyclopaedic Dictionary” by merging:

**Wordnet**
- concepts = sets of synonyms (synsets, ss);
- POS marking and word polysemy (1 word, many ss);
- Synset definitions/glosses;
- Synset example sentences;
- Lexical and semantic relations (e.g. *is-a, is-part-of, antonym, in-domain-of, etc.*);

+ **Gloss relations**;

**Gloss relations**

Given a Synset, $S$, and the set of disambiguated word in its gloss($S$),

$$s_i \in \text{gloss}(S) = \{s_1, \ldots, s_k\}, \quad i = 1, \ldots, k.$$ 

There is a relation between $S$ and all the synsets contained in its disambiguated gloss;
Lexicographic & Encyclopaedic Knowledge

An “Encyclopedic Dictionary” by merging:

**Wordnet**
- concepts = sets of synonyms (synsets, ss);
- POS marking and word polysemy (1 word, many ss);
- Synset definitions/glosses;
- Synset example sentences;
- Lexical and semantic relations (e.g. is-a, is-part-of, antonym, in-domain-of, etc.);

+ **Gloss relations**;

**Wikipedia**
- 1 article/page = 1 concept;
- Title of article = lemma;
- (opt.) Title label to help disambiguate the lemma (e.g. ‘play (activity)’ vs. ‘play (theatre)’);
- partly structured text (e.g. gloss is provided in the 1st sentence, info boxes with summarised info);
- Article relations (e.g. redirect pages, disambiguation pages, internal links, …);
Lexicographic & Encyclopaedic Knowledge

An “Encyclopedic Dictionary” by merging:

**Wikipedia**

- 1 article/page = 1 concept;
- Title of article = lemma;
- (opt.) Title label to help disambiguate the lemma (e.g. ‘play (activity)’ vs. ‘play (theatre)’);
- partly structured text (e.g. gloss is provided in the 1st sentence, info boxes with summarised info);
- **Article relations** (e.g. redirect pages, disambiguation pages, internal links, …);

- **Redirect pages** ≈ synonymity relations;
- **Disambiguation pages** ≈ word polysemy;
- **Inter-language links** ≈ synset keys (cross lingual);
- **Internal links** ≈ related synsets;
- **Categories** ≈ related synsets
Both can be viewed as graphs (w/ articles and synsets as nodes and relations and hyperlinks as edges).

It is evident that the two graphs complement each other.
**BabelNet - More Formally**

- **Labeled directed graph** with a set of nodes $V$ (concepts & named entities) and set of labeled edges $E \subseteq V \times R \times V$, that connect two nodes with a semantic relation from $R$, i.e., \{is-a, part-of, ..., $\varepsilon$\}; ($\varepsilon = \text{unspecified semantic relation}$)

- Each node $v \in V$ contains a set of lexicalizations in multiple languages > referred to as **Babel synsets**;

- One unified resource in **three steps**:
  - Combine WordNet and Wikipedia;
  - Harvest multilingual lexicalizations;
  - Harvest relations between Babel synsets;

Automatically acquiring a mapping between WordNet senses and Wikipages:

\[ \mu(w) = \begin{cases} 
  s \in \text{Senses}_{\text{WN}}(w) & \text{if a link can be established,} \\
  \varepsilon & \text{otherwise,} 
\end{cases} \]

- Treat mapping as a **disambiguation problem** - use disambiguation context to decide mapping;
- Mapping Algorithm - given \( w \), **finds s that maximizes the probability of s providing** an adequate corresponding concept for \( w \);
- Estimate the mapping conditional probability with two methods - simple bag-of-words (BoW), and graph based approach;

Pseudocode of the mapping algorithm:

```
for each
  for each
    if
      then : \( \mu(w) := w \)
  for each
    if \( \mu(w) = \varepsilon \) then :
      if \( \mu(d) \neq \varepsilon \) and \( \mu(d) \) is in a synset of \( w \) then :
        \( \mu(w) := \) sense of \( w \) in synset of \( \mu(d) \); break
  for each
    if
      if no tie occurs then :
        \( s \in \text{SensesWN}(w) \)
      \( s \in \text{SensesWN}(w) \)
  return \( \mu \)
```

\( P(s, w) \) - The joint probability of a WordNet sense and Wikipage, or “the probability of a WordNet sense and Wikipage referring to the same concept”:

- Similar to WSD
- The disambiguation context for each of the two concepts is the set of words with some semantic relation to each concept (from the corresponding resource).
  - labels, links, redirections and categories - WikiSenses \((w)\)
  - synonymy, hypernymy/hyponymy and gloss - WNSenses \((s)\)

\[
\text{Ctx}(w) , \text{Ctx}(\text{Play (theatre)}) = \{ \text{‘theatre’, ‘literature’, ‘comedy’, ‘drama’, ‘character’, … } \} \\
\text{Ctx}(s) , \text{Ctx}(\text{play\#01}^n) = \{ \text{‘drama’, ‘composition’, ‘work’, ‘intend’, ‘actor’, ‘stage’, … } \}
\]

Back do the probability estimation:

\[ p(s, w) = \frac{\text{score}(s, w)}{\sum_{s' \in \text{Senses}_{\text{WN}}(w), \ w' \in \text{Senses}_{\text{Wiki}}(w)} \text{score}(s', w')} \]

Two methods for computing \(\text{score}(s, w)\):

- **Bag-of-words method**
  \[ \text{score}(s, w) = |\text{Ctx}(s) \cap \text{Ctx}(w)| + 1 \] (smoothing)

- **Graph-based method**
  Transforms \(\text{Ctx}(w)\) into a subgraph of the Wordnet containing all the word in context and all edges and intermediate senses along all paths of a maximal length \(L\).

The scoring function is then defined as:

\[ \text{score}(s, w) = \sum_{cw \in \text{Ctx}(w)} \sum_{s' \in \text{Senses}_{\text{WN}}(cw)} \sum_{p \in \text{paths}_{\text{WN}}(s, s')} e^{-(\text{length}(p) - 1)} \]
2. Translating Babel synsets

So far, **Babel Synsets** are $S \cup W$, and where $W$ includes:

- $w$ - WikiSense;
- the set of redirections to $w$;
- all inter-language links;
- the redirections to the inter-language links found in the Wikipedia of the target language.

Two issues:

- **Unlinked concepts** between Wiki and WN
- Even if linked, Wiki may **not provide any/all translations**

To guarantee **coverage for all languages**, this was also tackled with **automated processes**.
2. Translating Babel synsets

They wanted **full coverage for 6 languages! > SMT!**

**For each polysemous WNsense and WikiSense**, SemCor and Wikipedia were mined for sentences (respectively) - ‘**BabelCor’**.

- min. 3 sentences/sense (for precision);
- max. 10 sentences/sense (for time saving);
- **excluded** WikiSenses recognised as **Named Entities** - assumed they are kept the same across languages (didn’t account for transliterations);

**NEs simple heuristic**: titles which contained at least two tokens starting with an uppercase letter were NEs - 94% on a validation sample of 100 pages. (e.g. William Shakespeare)
2. Translating Babel synsets

**Applied SOA SMT**, and identified top-scored translations as fit lexical entries for Babel Synsets. (Google Translate)

- monosemous senses were translated contextless;

As a result, **translated 324,137 WikiSenses** (reduced from over 3 million).
3. Harvesting semantic relations

- All lexical and semantic relations from WordNet ( + gloss relations) are inherited by BabelNet.
- All hyperlink relations from Wikipedia are collected and assigned an unspecified semantic relation $\varepsilon$.
  - including relations from other languages

**Weighted edges:**

- WN edges based on the Dice coefficient $\left(\frac{2 \times |S \cap S'|}{|S| + |S'|}\right)$
  - overlap between synonyms + gloss’s content words
- WikiPages uses a co-occurrence based method also applied to a Dice coefficient $\left(\frac{2 \times f_{w,w'}}{f_w + f_{w'}}\right)$
  - co-occurrence context of 40 words, by the total number of hyperlinks.
In Vitro Evaluation - Mapping

- Gold Standard set of 1000 WikiPages hand linked to WNSenses (w/ inter-annotator agreement of 0.9);

- Evaluation by replicating the BoW and the graph-based methods to estimate mapping probabilities;

- Explored different disambiguation contexts for WN;

- Disambiguation context for Wikipages is ‘everything’ they have
In Vitro Evaluation - Mapping

- Varied the maximum depth of graph search (exploring bigger portions of WordNet when building the disambiguation graphs)

- Baselines: Most Frequent Sense + Random Sense

- Removed the 100 most frequent linked sense labels Wikipages from the WordNet-Wikipedia intersection (avoid mapping WordNet senses to pages belonging to domains which are typically found in Wikipedia only - as ‘bands’, ‘movies’, etc.);
In Vitro Evaluation - Mapping

<table>
<thead>
<tr>
<th>Mapping method</th>
<th>P</th>
<th>R</th>
<th>F₁</th>
<th>A</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>BoW</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>taxonomic</td>
<td>89.7</td>
<td>47.8</td>
<td>62.3</td>
<td>72.6</td>
</tr>
<tr>
<td>gloss</td>
<td>87.6</td>
<td>51.8</td>
<td>65.1</td>
<td>74.0</td>
</tr>
<tr>
<td>taxonomic + gloss</td>
<td>87.5</td>
<td>65.6</td>
<td>75.0</td>
<td>80.9</td>
</tr>
<tr>
<td><strong>Graph</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>taxonomic relations</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>max depth @ 2</td>
<td>87.2</td>
<td>60.8</td>
<td>71.6</td>
<td>77.9</td>
</tr>
<tr>
<td>max depth @ 3</td>
<td>81.6</td>
<td>65.0</td>
<td>72.4</td>
<td>78.7</td>
</tr>
<tr>
<td>max depth @ 4</td>
<td>78.3</td>
<td>69.5</td>
<td>73.6</td>
<td>79.4</td>
</tr>
<tr>
<td>gloss relations</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>max depth @ 2</td>
<td>80.5</td>
<td>60.6</td>
<td>69.1</td>
<td>77.0</td>
</tr>
<tr>
<td>max depth @ 3</td>
<td>77.5</td>
<td>65.2</td>
<td>70.9</td>
<td>78.2</td>
</tr>
<tr>
<td>max depth @ 4</td>
<td>72.4</td>
<td>67.1</td>
<td>69.6</td>
<td>78.0</td>
</tr>
<tr>
<td>taxonomic + gloss relations</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>max depth @ 2</td>
<td>81.2</td>
<td>74.6</td>
<td>77.7</td>
<td>82.7</td>
</tr>
<tr>
<td>max depth @ 3</td>
<td>72.8</td>
<td>77.4</td>
<td>75.1</td>
<td>80.1</td>
</tr>
<tr>
<td>max depth @ 4</td>
<td>64.3</td>
<td>76.2</td>
<td>69.8</td>
<td>75.0</td>
</tr>
<tr>
<td>MFS baseline</td>
<td>25.4</td>
<td>49.2</td>
<td>33.5</td>
<td>25.4</td>
</tr>
<tr>
<td>Random baseline</td>
<td>24.2</td>
<td>46.9</td>
<td>31.9</td>
<td>24.2</td>
</tr>
</tbody>
</table>

- MFS and Random baselines are virtually the same thing!
- Richer disambiguation context helps,
- Graph based methods give a much higher recall
- Depth > 3 seems to hurt F₁ (noisy gloss-derived relations)

* table obtained from [1]
### On source of translation:

- **Translations came from multiple sources**;
- **SemCor was too small** to provide a substantial number of translations; (remember that a min. of 3 sentences was required)
- Combined translations from all sources (+ large contribution of sourceless monosemous translations);
- **Translate a substantial portion of WordNet**:
  - 83.4% of 82,115 nominal **synsets**
  - 79.0% of 146,312 nominal **senses**

---

**Table 4.2.3.1. Datasets.**

<table>
<thead>
<tr>
<th></th>
<th>SemCor</th>
<th>Wikipedia</th>
<th>SemCor ∪ Wikipedia</th>
<th>monosemous</th>
<th>all</th>
</tr>
</thead>
<tbody>
<tr>
<td># synsets</td>
<td>3,901</td>
<td>31,308</td>
<td>33,359</td>
<td>62,259</td>
<td>68,554</td>
</tr>
<tr>
<td># senses</td>
<td>6,852</td>
<td>35,372</td>
<td>40,504</td>
<td>101,853</td>
<td>115,606</td>
</tr>
</tbody>
</table>

---

*table obtained from [1]*
Coverage against gold-standard Wordnets:

- 5 golden standard Wordnets; (Spanish, Catalan, French, German, Italian)
- All linked to PWN which is contained in BabelNet - so it’s easy to quantify their overlap;
- **Synset Coverage** is measured by the % of synsets of the gold-standard WN that shares at least one term with BabelNet;
- **Word Coverage** is measured by the ration of word senses in the gold-standard WN that overlap with each Babel Synset;
- **Extra Synset Coverage** and **Extra Word Coverage** measures the PWN synsets and word ratio that are not covered by gold-standard WNs, but that are covered by BabelNet;
### Size of the gold-standard wordnets.

<table>
<thead>
<tr>
<th>Resource</th>
<th>Catalan</th>
<th>French</th>
<th>German</th>
<th>Italian</th>
<th>Spanish</th>
</tr>
</thead>
<tbody>
<tr>
<td>Word senses</td>
<td>64,171</td>
<td>44,265</td>
<td>15,762</td>
<td>57,255</td>
<td>83,114</td>
</tr>
<tr>
<td>Synsets</td>
<td>40,466</td>
<td>31,742</td>
<td>9,877</td>
<td>32,156</td>
<td>55,365</td>
</tr>
</tbody>
</table>

### Coverage against gold-standard wordnets (percentages).

<table>
<thead>
<tr>
<th>Resource</th>
<th>Method</th>
<th>Wiki</th>
<th>Transl.</th>
<th>Transl.</th>
<th>All</th>
<th>Wiki</th>
<th>Transl.</th>
<th>Transl.</th>
<th>All</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Catalan</td>
<td>Links</td>
<td></td>
<td></td>
<td>20.3</td>
<td></td>
<td></td>
<td></td>
<td>25.6</td>
</tr>
<tr>
<td></td>
<td>French</td>
<td>70.0</td>
<td>69.6</td>
<td>16.3</td>
<td>86.0</td>
<td>72.4</td>
<td>79.6</td>
<td>19.4</td>
<td>92.9</td>
</tr>
<tr>
<td></td>
<td>German</td>
<td>39.6</td>
<td>42.6</td>
<td>21.0</td>
<td>57.6</td>
<td>50.7</td>
<td>58.2</td>
<td>28.6</td>
<td>73.4</td>
</tr>
<tr>
<td></td>
<td>Italian</td>
<td>28.1</td>
<td>39.9</td>
<td>19.7</td>
<td>52.9</td>
<td>40.0</td>
<td>58.0</td>
<td>28.7</td>
<td>73.7</td>
</tr>
<tr>
<td></td>
<td>Spanish</td>
<td>34.4</td>
<td>47.9</td>
<td>25.2</td>
<td>66.4</td>
<td>40.7</td>
<td>56.1</td>
<td>30.0</td>
<td>76.6</td>
</tr>
</tbody>
</table>

### Extra coverage against gold-standard wordnets (percentages).

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Catalan</td>
<td>Links</td>
<td></td>
<td></td>
<td>100</td>
<td></td>
<td></td>
<td></td>
<td>142</td>
</tr>
<tr>
<td></td>
<td>French</td>
<td>255</td>
<td>223</td>
<td>92</td>
<td>514</td>
<td>63</td>
<td>102</td>
<td>67</td>
<td>159</td>
</tr>
<tr>
<td></td>
<td>German</td>
<td>1349</td>
<td>940</td>
<td>367</td>
<td>2298</td>
<td>506</td>
<td>668</td>
<td>303</td>
<td>902</td>
</tr>
<tr>
<td></td>
<td>Italian</td>
<td>160</td>
<td>234</td>
<td>83</td>
<td>419</td>
<td>87</td>
<td>153</td>
<td>68</td>
<td>213</td>
</tr>
<tr>
<td></td>
<td>Spanish</td>
<td>214</td>
<td>158</td>
<td>56</td>
<td>384</td>
<td>48</td>
<td>74</td>
<td>30</td>
<td>102</td>
</tr>
</tbody>
</table>
On the precision of the extra coverage:

- Manual validation of 3000 random Babel Synsets:
  (check whether lexical entries fit the Babel Synset glosses translations, SemCor translations and wiki sentences translations)
  
  ▶ 600 synsets x 5 languages (exc. English);
  ▶ 200 synsets present only in PWN;
  ▶ 200 synsets derived from Wikipedia only;
  ▶ 200 synsets derived from intersection of PWN and Wiki;

Precision of BabelNet on synonyms in WordNet (WN), Wikipedia (Wiki) and their intersection (WN $\cap$ Wiki): percentage and total number of words (in parentheses) are reported.

<table>
<thead>
<tr>
<th>Resource</th>
<th>Catalan</th>
<th>French</th>
<th>German</th>
<th>Italian</th>
<th>Spanish</th>
</tr>
</thead>
<tbody>
<tr>
<td>WordNet</td>
<td>75.58 (258)</td>
<td>67.16 (268)</td>
<td>73.76 (282)</td>
<td>72.32 (271)</td>
<td>69.45 (275)</td>
</tr>
<tr>
<td>Wiki</td>
<td>92.71 (398)</td>
<td>96.44 (758)</td>
<td>97.74 (709)</td>
<td>99.09 (552)</td>
<td>92.46 (703)</td>
</tr>
<tr>
<td>WordNet $\cap$ Wiki</td>
<td>82.98 (517)</td>
<td>77.43 (709)</td>
<td>78.37 (777)</td>
<td>80.83 (574)</td>
<td>78.53 (643)</td>
</tr>
</tbody>
</table>

* table obtained from [1]
Number of monosemous and polysemous words by part of speech (verbs, adjectives and adverbs are the same as in WordNet 3.0).

<table>
<thead>
<tr>
<th>POS</th>
<th>Monosemous words</th>
<th>Polysemous words</th>
</tr>
</thead>
<tbody>
<tr>
<td>Noun</td>
<td>22,763,265</td>
<td>1,134,857</td>
</tr>
<tr>
<td>Verb</td>
<td>6,277</td>
<td>5,252</td>
</tr>
<tr>
<td>Adjective</td>
<td>1,503</td>
<td>4,976</td>
</tr>
<tr>
<td>Adverb</td>
<td>3,748</td>
<td>733</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>22,789,793</strong></td>
<td><strong>1,145,818</strong></td>
</tr>
</tbody>
</table>

Number of lemmas, synsets and word senses in the 6 languages currently covered by BabelNet.

<table>
<thead>
<tr>
<th>Language</th>
<th>Lemmas</th>
<th>Synsets</th>
<th>Word senses</th>
</tr>
</thead>
<tbody>
<tr>
<td>English</td>
<td>5,938,324</td>
<td>3,032,406</td>
<td>6,550,579</td>
</tr>
<tr>
<td>Catalan</td>
<td>3,518,079</td>
<td>2,214,781</td>
<td>3,777,700</td>
</tr>
<tr>
<td>French</td>
<td>3,754,079</td>
<td>2,285,458</td>
<td>4,091,456</td>
</tr>
<tr>
<td>German</td>
<td>3,602,447</td>
<td>2,270,159</td>
<td>3,910,485</td>
</tr>
<tr>
<td>Italian</td>
<td>3,498,948</td>
<td>2,268,188</td>
<td>3,773,384</td>
</tr>
<tr>
<td>Spanish</td>
<td>3,623,734</td>
<td>2,252,632</td>
<td>3,941,039</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>23,935,611</strong></td>
<td><strong>3,032,406</strong></td>
<td><strong>26,044,643</strong></td>
</tr>
</tbody>
</table>

* tables obtained from [1]
### Original BabelNet Stats

Number of lexico-semantic relations harvested from WordNet, WordNet glosses and the 6 wikipedias.

<table>
<thead>
<tr>
<th></th>
<th>English</th>
<th>Catalan</th>
<th>French</th>
<th>German</th>
<th>Italian</th>
<th>Spanish</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>WordNet</td>
<td>364,552</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>364,552</td>
</tr>
<tr>
<td>WordNet glosses</td>
<td>617,785</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>617,785</td>
</tr>
<tr>
<td>Wikipedia</td>
<td>50,104,884</td>
<td>978,006</td>
<td>5,613,873</td>
<td>5,940,612</td>
<td>3,602,395</td>
<td>3,411,612</td>
<td>69,651,382</td>
</tr>
<tr>
<td>Total</td>
<td>51,087,221</td>
<td>978,006</td>
<td>5,613,873</td>
<td>5,940,612</td>
<td>3,602,395</td>
<td>3,411,612</td>
<td>70,633,719</td>
</tr>
</tbody>
</table>

Glosses for the Babel synset referring to the concept of play as 'dramatic work'.

<table>
<thead>
<tr>
<th>English</th>
<th>WordNet</th>
<th>Wikipedia</th>
</tr>
</thead>
<tbody>
<tr>
<td>A dramatic work intended for performance by actors on a stage.</td>
<td>A play is a form of literature written by a playwright, usually consisting of scripted dialogue between characters, intended for theatrical performance rather than just reading.</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Catalan</th>
</tr>
</thead>
<tbody>
<tr>
<td>El drama en termes generals és una obra literària o una situació de la vida real que resulta complexa i difícil però amb un final favorable o feliç.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>French</th>
</tr>
</thead>
<tbody>
<tr>
<td>Le drame (du latin drama, emprunté au grec ancien δράμα, qui signifie action (théâtrale), pièce de théâtre) désigne étymologiquement toute action scénique.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>German</th>
</tr>
</thead>
<tbody>
<tr>
<td>Drama (altgriechisch δράμα ‘Handlung’) ist ein Oberbegriff für Texte mit verteilten Rollen.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Italian</th>
</tr>
</thead>
<tbody>
<tr>
<td>Un dramma, dal greco “drama” (azione, storia; da δρας, fare), è una forma letteraria che include parti scritte per essere interpretate da attori.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Spanish</th>
</tr>
</thead>
<tbody>
<tr>
<td>Drama (del griego δράμα, hacer o actuar) es la forma de presentación de acciones a través de su representación por actores.</td>
</tr>
</tbody>
</table>

**+ BabelCor - Sense-tagged corpus** with almost 2 million sentences (46,155 from SemCor and 1,940,402 from Wikipedia) - 330,993 annotated senses

* tables obtained from [1]
New BabelNet 2.5

Integrates data from:

- WordNet3.0
- Wikipedia, Wiktionary, Wikidata
- OmegaWiki
- Open Multilingual WordNet 😊
- DBpedia

Now also includes:

- Translations for all open-class POS;
- Links to Categories;
- Images;
- etc.;
New BabelNet 2.5 Stats

<table>
<thead>
<tr>
<th>Category</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of languages</td>
<td>50</td>
</tr>
<tr>
<td>Total number of Babel synsets</td>
<td>9,348,287</td>
</tr>
<tr>
<td>Total number of Babel senses</td>
<td>67,873,191</td>
</tr>
<tr>
<td>Total number of concepts</td>
<td>3,684,512</td>
</tr>
<tr>
<td>Total number of Named Entities</td>
<td>5,663,775</td>
</tr>
<tr>
<td>Total number of lexico-semantic relations</td>
<td>262,687,848</td>
</tr>
<tr>
<td>Total number of glosses (textual definitions)</td>
<td>21,771,854</td>
</tr>
<tr>
<td>Total number of images</td>
<td>7,764,270</td>
</tr>
<tr>
<td>Total number of RDF triples</td>
<td>1,138,337,378</td>
</tr>
</tbody>
</table>

* data and charts obtained from [http://babelnet.org/stats](http://babelnet.org/stats)
Let’s go online for a second…

We’ll continue shortly after with…

Extrinsic Evaluation
SemEval-2007 T#16 - Evaluating wide-coverage Knowledge Resources (KBEval)

Knowledge bases were assessed by first generating so-called topic signatures + monolingual WSD;

Task:

1. Given a concept, generate a topic signature (e.g. word vector)
2. Unsupervised monolingual WSD:
   - given a word in context, compute the topic signature for each word in context,
   - compute a simple overlap score (with test sentence),
   - word sense with max score is selected;
Test data:

- Two sets from previous Senseval and SemEval tasks;
- Sense annotated with PWN senses;

BabelNet-1 & BabelNet-2 Systems

1. Collect all synsets where word appears as synonym in a WN ‘Babel enriched’ synset;
2. Topic signature is all the english lexicalizations reachable by a distance of 1 and 2 (respectively);
3. Output the PWN synset associated with the winning BabelNet Synset;
Results on the SemEval-2007 task 16: Evaluation of wide coverage knowledge resources.

(a) Senseval-3 English Lexical Sample task:

<table>
<thead>
<tr>
<th>Knowledge base</th>
<th>P</th>
<th>R</th>
<th>F₁</th>
<th>Avg. size</th>
</tr>
</thead>
<tbody>
<tr>
<td>TRAIN</td>
<td>65.1</td>
<td>65.1</td>
<td>65.1</td>
<td>450</td>
</tr>
<tr>
<td>TRAIN-MFS</td>
<td>54.5</td>
<td>54.5</td>
<td>54.5</td>
<td>–</td>
</tr>
<tr>
<td>WN-MFS</td>
<td>53.0</td>
<td>53.0</td>
<td>53.0</td>
<td>–</td>
</tr>
<tr>
<td>SEMCOR-MFS</td>
<td>49.0</td>
<td>49.1</td>
<td>49.0</td>
<td>–</td>
</tr>
<tr>
<td>TSSEM</td>
<td>52.5</td>
<td>52.4</td>
<td>52.4</td>
<td>103</td>
</tr>
<tr>
<td>BabelNet-1</td>
<td>44.3</td>
<td>44.3</td>
<td>44.3</td>
<td>119</td>
</tr>
<tr>
<td>BabelNet-2</td>
<td>35.0</td>
<td>35.0</td>
<td>35.0</td>
<td>2,128</td>
</tr>
<tr>
<td>KnowNet-20</td>
<td>44.1</td>
<td>44.1</td>
<td>44.1</td>
<td>610</td>
</tr>
<tr>
<td>RANDOM</td>
<td>19.1</td>
<td>19.1</td>
<td>19.1</td>
<td>–</td>
</tr>
</tbody>
</table>

(b) SemEval-2007 English Lexical Sample (task 17):

<table>
<thead>
<tr>
<th>Knowledge base</th>
<th>P</th>
<th>R</th>
<th>F₁</th>
<th>Avg. size</th>
</tr>
</thead>
<tbody>
<tr>
<td>TRAIN</td>
<td>87.6</td>
<td>87.6</td>
<td>87.6</td>
<td>450</td>
</tr>
<tr>
<td>TRAIN-MFS</td>
<td>81.2</td>
<td>81.2</td>
<td>81.2</td>
<td>–</td>
</tr>
<tr>
<td>WN-MFS</td>
<td>66.2</td>
<td>59.9</td>
<td>62.9</td>
<td>–</td>
</tr>
<tr>
<td>SEMCOR-MFS</td>
<td>42.4</td>
<td>38.4</td>
<td>40.3</td>
<td>–</td>
</tr>
<tr>
<td>WN + XWN + KN-20</td>
<td>53.0</td>
<td>53.0</td>
<td>53.0</td>
<td>627</td>
</tr>
<tr>
<td>BabelNet-1</td>
<td>52.2</td>
<td>46.3</td>
<td>49.1</td>
<td>130</td>
</tr>
<tr>
<td>BabelNet-2</td>
<td>56.9</td>
<td>53.1</td>
<td>54.9</td>
<td>2,352</td>
</tr>
<tr>
<td>KnowNet-20</td>
<td>49.5</td>
<td>46.1</td>
<td>47.7</td>
<td>561</td>
</tr>
<tr>
<td>RANDOM</td>
<td>19.1</td>
<td>19.1</td>
<td>19.1</td>
<td>–</td>
</tr>
</tbody>
</table>

F₁ = 49.9 and 43.3 on SemEval-2007 for WordNet and Wikipedia-only relations at distance 2, respectively.

* tables obtained from [1]
SemEval-2007 - Task#7

SemEval-2007 T#7 - Coarse-grained all-words WSD task

Is granularity of WN senses an obstacle for WSD?

A coarse sense inventory is obtained semi-automatically by clustering WN senses via a mapping to the Oxford Dictionary of English;

Task:

1. ~6,000 words ‘coarse + grained’ sense tagged,
2. Participants have access to a lemma and a POS for each content word;
3. They have to output a coarse sense (optionally a cluster replaces the fine-grained sense choice) for each word.
BabelNet in SemEval-2007 T#7:

**Hypothesis**: the meanings of Wikipages are intuitively coarser than those in WordNet, so it should be better at coarser WSD.

1. Edge filtering: filtered paths connecting different senses of the same word + removed edges from the graph whose weight is below a certain threshold;

2. Defined a general framework for transforming an input context into a graph; (same as for estimating mapping probabilities)

3. Applied 4 algorithms for graph-based lexico-semantic disambiguation;

4. Best results with MFS is assigned when no sense assignment it attempted; (weakly supervised)
Performance on SemEval-2007 coarse-grained all-words WSD with MFS as a back-off strategy when no sense assignment is attempted. The differences between the results in bold in each column of the table are not statistically significant at $p < 0.05$ based on a $\chi^2$ test.

<table>
<thead>
<tr>
<th>Resource</th>
<th>Algorithm</th>
<th>Nouns only P/R/F</th>
<th>All words P/R/F</th>
</tr>
</thead>
<tbody>
<tr>
<td>WordNet</td>
<td>Degree</td>
<td>80.1</td>
<td>79.7</td>
</tr>
<tr>
<td></td>
<td>PLength</td>
<td>80.3</td>
<td>79.8</td>
</tr>
<tr>
<td></td>
<td>SProbability</td>
<td>79.5</td>
<td>79.3</td>
</tr>
<tr>
<td></td>
<td>PageRank</td>
<td>79.7</td>
<td>79.4</td>
</tr>
<tr>
<td>BabelNet</td>
<td>Degree</td>
<td><strong>84.7</strong></td>
<td><strong>82.3</strong></td>
</tr>
<tr>
<td></td>
<td>PLength</td>
<td><strong>85.4</strong></td>
<td><strong>82.7</strong></td>
</tr>
<tr>
<td></td>
<td>SProbability</td>
<td><strong>84.6</strong></td>
<td><strong>82.1</strong></td>
</tr>
<tr>
<td></td>
<td>PageRank</td>
<td>82.1</td>
<td>80.1</td>
</tr>
<tr>
<td></td>
<td>SUSSX-FR</td>
<td>81.1</td>
<td>77.0</td>
</tr>
<tr>
<td></td>
<td>TreeMatch</td>
<td>N/A</td>
<td>73.6</td>
</tr>
<tr>
<td></td>
<td>NUS-PT</td>
<td>82.3</td>
<td><strong>82.5</strong></td>
</tr>
<tr>
<td></td>
<td>SSI</td>
<td><strong>84.1</strong></td>
<td><strong>83.2</strong></td>
</tr>
<tr>
<td></td>
<td>MFS BL</td>
<td>77.4</td>
<td>78.9</td>
</tr>
<tr>
<td></td>
<td>Random BL</td>
<td>63.5</td>
<td>62.7</td>
</tr>
</tbody>
</table>

* Best results were found for a maximum depth of 3, and a minimum edge weight of 0.01

BabelNet beats the MFS baseline on nouns (a notably difficult competitor for unsupervised and knowledge-rich systems) - even without back-off strategy.
SemEval-2010 T#3: Cross-lingual WSD

In this task, lexical disambiguation is operationalized as a word translation task;

Task:

1. Given a predefined sense inventory in a MLCorpus (where all necessary sense distinctions are available for every language);

2. 20 target words x 50 sentences each (1,000 test instances), for each word in context, participants disambiguate the target word by translating it into a different language;

3. The meaning preservingness of the translations are evaluated from a list of weighted/ranked, gold-standard translations;
BabelNet in SemEval-2010 T#3:
Same ‘turn input into graph’ framework from an input + context (max. depth of 3) + 4 algorithms for monolingual WSD;

» **Standard Setting:** return MFTranslation from winning synset (as ordered by frequency of alignment in Europarl); back-off to MFAlignment in Europarl - if no sense assigned

For better insight: remove from selected synset lexicalizations not seen in the gold-standard corpus; then MFTranslation; back-off to MFAlignment in Europarl; (+Oracle Transl.)

Upper bounds:
- **BabelNet:** return all gold-standard translations found in the known BabelSS of the test instance;
- **Task:** since evaluation metrics are not in %, this gives the highest ranking translations chosen by humans;
### Results on the SemEval-2010 task 3: Cross-lingual Word Sense Disambiguation.

<table>
<thead>
<tr>
<th></th>
<th>French P</th>
<th>French R</th>
<th>German P</th>
<th>German R</th>
<th>Italian P</th>
<th>Italian R</th>
<th>Spanish P</th>
<th>Spanish R</th>
</tr>
</thead>
<tbody>
<tr>
<td>Baseline</td>
<td>21.25</td>
<td>21.25</td>
<td>13.16</td>
<td>13.16</td>
<td>15.18</td>
<td>15.18</td>
<td>19.74</td>
<td>19.74</td>
</tr>
<tr>
<td>UvT-v</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>23.39</td>
<td>23.39</td>
</tr>
<tr>
<td>UvT-g</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>19.83</td>
<td>19.64</td>
</tr>
<tr>
<td>T3-COLEUR</td>
<td>21.97</td>
<td>21.75</td>
<td>13.18</td>
<td>13.05</td>
<td>14.82</td>
<td>14.67</td>
<td>19.83</td>
<td>19.64</td>
</tr>
<tr>
<td>Degree</td>
<td>22.94</td>
<td>22.94</td>
<td>17.15</td>
<td>17.15</td>
<td>18.03</td>
<td>18.03</td>
<td>22.48</td>
<td>22.48</td>
</tr>
<tr>
<td>PLength</td>
<td>23.42</td>
<td>23.42</td>
<td>17.72</td>
<td>17.72</td>
<td>18.19</td>
<td>18.19</td>
<td>22.76</td>
<td>22.76</td>
</tr>
<tr>
<td>SProbability</td>
<td>23.27</td>
<td>23.27</td>
<td>17.61</td>
<td>17.61</td>
<td>18.14</td>
<td>18.14</td>
<td>22.69</td>
<td>22.69</td>
</tr>
<tr>
<td>+ ORACLE TRANSLATIONS</td>
<td>25.85</td>
<td>25.85</td>
<td>20.50</td>
<td>20.50</td>
<td>21.74</td>
<td>21.74</td>
<td>25.48</td>
<td>25.48</td>
</tr>
<tr>
<td>BabelNet upper bound</td>
<td>30.21</td>
<td>30.21</td>
<td>25.39</td>
<td>25.39</td>
<td>27.67</td>
<td>27.67</td>
<td>30.73</td>
<td>30.73</td>
</tr>
<tr>
<td>Task upper bound</td>
<td>39.44</td>
<td>100.00</td>
<td>34.36</td>
<td>100.00</td>
<td>40.00</td>
<td>100.00</td>
<td>39.54</td>
<td>100.00</td>
</tr>
</tbody>
</table>

**Precision:** number of correct system translations over the total number of translations returned by the system, each weighted by their score. (higher ranked would receive higher points)

**Recall:** number of correct translations given by the system over the total number of items in the test set, each weighted by their score.
SemEval-2013 - Task#12

Multilingual WSD using BabelNet Senses

Task Preparation:

1. 13 articles were selected from WSMT, all existed in 4 langs. (English, French, German and Spanish) + translation in Italian;
2. Due to the automatic integration, all the mappings of 8306 synsets (for 978 lemmas appearing in the corpus), were manually checked (delete, add mapping or merge);
3. Manual correction of POS, NE and MWE tagging;
4. Manual sense annotation for English, projected across to other languages;
5. Evaluated for precision and recall on BabelNet, Wordnet, (against WN MFS) and Wikipedia senses (against pseudo MFS for WikiSenses - WN frequency or lexical ordering).
Cross Language Sense Projection:

1. English dataset was manually annotated; (1+ senses allowed)
2. Other datasets were sentence aligned, and lemmas compared to that lang’s lexical entries in used english senses;
   - if a match occurred, that english sense would be projected;
   - labelled 50%-70% of non-english datasets;
3. Manually completed, corrected and later reviewed;
   - only 22-37% needed correction; (simple but efficient)

<table>
<thead>
<tr>
<th>Language</th>
<th>Projected instances</th>
<th>Valid projections</th>
<th>Invalid projections</th>
</tr>
</thead>
<tbody>
<tr>
<td>French</td>
<td>1016</td>
<td>791</td>
<td>225</td>
</tr>
<tr>
<td>German</td>
<td>592</td>
<td>373</td>
<td>219</td>
</tr>
<tr>
<td>Italian</td>
<td>1029</td>
<td>774</td>
<td>255</td>
</tr>
<tr>
<td>Spanish</td>
<td>911</td>
<td>669</td>
<td>242</td>
</tr>
</tbody>
</table>

* table obtained from [2]
## Statistics for sense annotated data

<table>
<thead>
<tr>
<th>Language</th>
<th>Instances</th>
<th>Single-words</th>
<th>Multiword expressions</th>
<th>Named Entities</th>
<th>Mean senses per instance</th>
<th>Mean senses per lemma</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>BabelNet</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>English</td>
<td>1931</td>
<td>1604</td>
<td>127</td>
<td>200</td>
<td>1.02</td>
<td>1.09</td>
</tr>
<tr>
<td>French</td>
<td>1656</td>
<td>1389</td>
<td>89</td>
<td>176</td>
<td>1.05</td>
<td>1.15</td>
</tr>
<tr>
<td>German</td>
<td>1467</td>
<td>1267</td>
<td>21</td>
<td>176</td>
<td>1.00</td>
<td>1.05</td>
</tr>
<tr>
<td>Italian</td>
<td>1706</td>
<td>1454</td>
<td>211</td>
<td>41</td>
<td>1.22</td>
<td>1.27</td>
</tr>
<tr>
<td>Spanish</td>
<td>1481</td>
<td>1103</td>
<td>129</td>
<td>249</td>
<td>1.15</td>
<td>1.19</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Wikipedia</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>English</td>
<td>1242</td>
<td>945</td>
<td>102</td>
<td>195</td>
<td>1.15</td>
<td>1.16</td>
</tr>
<tr>
<td>French</td>
<td>1039</td>
<td>790</td>
<td>72</td>
<td>175</td>
<td>1.18</td>
<td>1.14</td>
</tr>
<tr>
<td>German</td>
<td>1156</td>
<td>957</td>
<td>21</td>
<td>176</td>
<td>1.07</td>
<td>1.08</td>
</tr>
<tr>
<td>Italian</td>
<td>977</td>
<td>869</td>
<td>85</td>
<td>41</td>
<td>1.20</td>
<td>1.18</td>
</tr>
<tr>
<td>Spanish</td>
<td>1103</td>
<td>758</td>
<td>107</td>
<td>248</td>
<td>1.11</td>
<td>1.10</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>WordNet</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>English</td>
<td>1644</td>
<td>1502</td>
<td>85</td>
<td>57</td>
<td>1.01</td>
<td>1.10</td>
</tr>
</tbody>
</table>

* Table obtained from [2]
Participants:

- 7 systems participated; (6 for BabelNet; 4 to WN; 3 to Wiki)
- All of them used graph-based approaches for WSD;
  - **DAEBAK!** (1 BabelNet) - ±5 sentence window around the target word; sense selection based on measuring connectivity to the synsets of neighboring lemmas; MFS as back off;
  - **GETALP** (2x BabelNet, 1 WN) - all based on the ant-colony algorithm (tuned differently); BN1 optimizes from the trial data; BN2 and WN1 are completely unsupervised;
  - **UMCC-DLSI** (3 x BabelNet, 1 WN, 1 Wiki) - all based on ISR-WN (resource that enriches WN semantic network from multiple lexical resources); WSD performed by an extension of the Personalized PageRank; RUN-1 uses all noun instances in the sentence as context, RUN-2 all noun instances in the document, and RUN-3 all words in the sentence;
### F1 score per language on BabelNet senses

<table>
<thead>
<tr>
<th>Team</th>
<th>System</th>
<th>English</th>
<th>French</th>
<th>German</th>
<th>Italian</th>
<th>Spanish</th>
</tr>
</thead>
<tbody>
<tr>
<td>DAEBAK!</td>
<td>PD</td>
<td>0.604</td>
<td>0.538</td>
<td>0.591</td>
<td>0.613</td>
<td>0.600</td>
</tr>
<tr>
<td>GETALP</td>
<td>BN-1</td>
<td>0.263</td>
<td>0.261</td>
<td>0.404</td>
<td>0.324</td>
<td>-</td>
</tr>
<tr>
<td>GETALP</td>
<td>BN-2</td>
<td>0.266</td>
<td>0.257</td>
<td>0.400</td>
<td>0.324</td>
<td>0.371</td>
</tr>
<tr>
<td>UMCC-DLSI</td>
<td>RUN-1</td>
<td>0.677</td>
<td>0.605</td>
<td>0.618</td>
<td>0.657</td>
<td>0.705</td>
</tr>
<tr>
<td>UMCC-DLSI</td>
<td>RUN-2</td>
<td>0.685</td>
<td>0.605</td>
<td>0.621</td>
<td>0.658</td>
<td>0.710</td>
</tr>
<tr>
<td>UMCC-DLSI</td>
<td>RUN-3</td>
<td>0.680</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>MFS</td>
<td></td>
<td>0.665</td>
<td>0.453</td>
<td>0.674</td>
<td>0.575</td>
<td>0.645</td>
</tr>
</tbody>
</table>

### Systems’ performance on Wordnet senses (English Only)

<table>
<thead>
<tr>
<th>Team</th>
<th>System</th>
<th>Precision</th>
<th>Recall</th>
<th>F1</th>
</tr>
</thead>
<tbody>
<tr>
<td>GETALP</td>
<td>WN-1</td>
<td>0.406</td>
<td>0.406</td>
<td>0.406</td>
</tr>
<tr>
<td>UMCC-DLSI</td>
<td>RUN-1</td>
<td>0.639</td>
<td>0.635</td>
<td>0.637</td>
</tr>
<tr>
<td>UMCC-DLSI</td>
<td>RUN-2</td>
<td>0.649</td>
<td>0.645</td>
<td>0.647</td>
</tr>
<tr>
<td>UMCC-DLSI</td>
<td>RUN-3</td>
<td>0.642</td>
<td>0.639</td>
<td>0.640</td>
</tr>
<tr>
<td>MFS</td>
<td></td>
<td>0.630</td>
<td>0.630</td>
<td>0.630</td>
</tr>
</tbody>
</table>

* tables obtained from [2]
## F1 score per instance type, averaged across all languages

<table>
<thead>
<tr>
<th>Team</th>
<th>System</th>
<th>Single term</th>
<th>Multiword expression</th>
<th>Named Entity</th>
</tr>
</thead>
<tbody>
<tr>
<td>DAEBAK!</td>
<td>PD</td>
<td>0.502</td>
<td>0.801</td>
<td>0.910</td>
</tr>
<tr>
<td>GETALP</td>
<td>BN-1</td>
<td>0.232</td>
<td>0.724</td>
<td>0.677</td>
</tr>
<tr>
<td>GETALP</td>
<td>BN-2</td>
<td>0.235</td>
<td>0.740</td>
<td>0.656</td>
</tr>
<tr>
<td>UMCC-DLSI</td>
<td>RUN-1</td>
<td>0.582</td>
<td>0.806</td>
<td>0.865</td>
</tr>
<tr>
<td>UMCC-DLSI</td>
<td>RUN-2</td>
<td><strong>0.584</strong></td>
<td><strong>0.809</strong></td>
<td>0.864</td>
</tr>
<tr>
<td>MFS</td>
<td></td>
<td>0.511</td>
<td>0.853</td>
<td>0.920</td>
</tr>
</tbody>
</table>

* table obtained from [2]
No system used cross lingual WSD - did not used bitext nor multilingual structure of BabelNet;

The task organisers tweaked the submitted systems to check the utility of simple multilingual sense analysis;

- Sense assignments were only kept iff at least two other aligned sentences have the same sense assigned to some word;

### Post Hoc: scores for outputs where at least two other aligned sentences (cross lingual) were used

<table>
<thead>
<tr>
<th>Team</th>
<th>System</th>
<th>English</th>
<th>French</th>
<th>German</th>
<th>Italian</th>
<th>Spanish</th>
</tr>
</thead>
<tbody>
<tr>
<td>DAEBAK</td>
<td>PD</td>
<td>0.769</td>
<td>0.364</td>
<td>0.494</td>
<td>0.747</td>
<td>0.387</td>
</tr>
<tr>
<td>GETALP</td>
<td>BN-2</td>
<td><strong>0.793</strong></td>
<td>0.111</td>
<td>0.195</td>
<td><strong>0.623</strong></td>
<td>0.130</td>
</tr>
<tr>
<td>UMCC-DLSI</td>
<td>RUN-1</td>
<td>0.787</td>
<td>0.421</td>
<td>0.549</td>
<td>0.754</td>
<td>0.441</td>
</tr>
<tr>
<td>UMCC-DLSI</td>
<td>RUN-2</td>
<td>0.791</td>
<td>0.419</td>
<td>0.548</td>
<td>0.760</td>
<td>0.436</td>
</tr>
</tbody>
</table>

* tables obtained from [2]
An excerpt of the semantic interpretation graph automatically built for the sentence *Thomas and Mario are strikers playing in Munich* (the edges connecting the correct meanings are in bold).

* graph obtained from [3]
[EN] The queen of England was eating a delicious hot dog while she was waiting for her driver by the bank.

[PT] A rainha de Inglaterra estava a comer um delicioso cachorro quente enquanto esperava pelo seu motorista junto ao banco.

[IT] La regina di Inghilterra stava mangiando un hot dog delizioso mentre aspettava il suo autista alla banca.

[CH1] 英国女王在银行等待她的司机的时候吃了一个美味的热狗。

[CH2] 当英格兰女王在银行等待她的司机的时候，她正在吃着一个美味的热狗。

[JP] イギリスの女王様が銀行の手前で運転手を待っていながら美味しいホットドッグを食べていた。

[KO] 영국 여왕은 은행 옆에서 운전수를 기다리고 있을 때 맛있는 핫도그를 먹고 있었다.
Some considerations

- BabelNet is ‘grand’ but a bit noisy;
- The WN disambiguation context for mapping Wiki to WN is too weak… (did they mentioned everything?)
- And from a cognitive perspective, weighted relations for WSD also seem the right choice for me; but also hierarchical (if a threshold is reached by stronger relations, disregard weaker)
- Should we have the gloss relations explicit in the OMW?
  - and could we ask ILI to try to provide WSD’ted definitions?
- This would be even more interesting with a stronger cross lingual disambiguation when mapping WikiSenses to WNSenses; (it may have happened in V2.5)
References

