Linking Thesauri and Glossaries
Case Study 0: linking a fake resource
Roberto Navigli

The Luxembourg BabelNet Workshop – Session 6
• Linking Glossaries and Thesauri to BabelNet and Babelfy

• Case study 0: linking a minimal fake resource
Linking glossaries and thesauri to BabelNet

Your (linguistic) resource
Ambiguity: a recurring issue

- Words in natural languages are ambiguous, though.
- In what meanings words are used in your thesaurus or glossary depends on its generality, its domain specificity, etc.

“The quick brown fox jumps over the lazy dog”
Ambiguity: a recurring issue

- Words in natural languages are ambiguous, though.
- In what meanings words are used in your thesaurus or glossary depends on its generality, its domain specificity, etc.

“The quick brown fox jumps over the lazy dog”
Linking to BabelNet
Linking to BabelNet
Linking to BabelNet
Linking to BabelNet

A key proxy for multilinguality!
Case study 0: Hello world (a minimal fake resource)

• Let's see how to perform a simple mapping between BabelNet and a minimal fake resource
Case study 0: Hello world (a minimal fake resource)

• Let's see how to perform a simple mapping between BabelNet and a minimal fake resource
Case study 0: Hello world (a minimal fake resource)

- Open your browser
Case study 0: Hello world (a minimal fake resource)

• Open your browser
• Type: babelnet.org/data/fake_resource.zip
Case study 0: Hello world (a minimal fake resource)

- Open your browser
- Type: babelnet.org/data/fake_resource.zip
Case study 0: Hello world (a minimal fake resource)

- Open your browser
- Type: babelnet.org/data/fake_resource.zip
- Unpack your zip file
Case study 0: Hello world (a minimal fake resource)

<table>
<thead>
<tr>
<th>ID</th>
<th>Lemma</th>
<th>POS (Part Of Speech)</th>
<th>Glosses</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Exhaust</td>
<td>VERB</td>
<td>To draw or let out wholly; to drain off completely; as, to exhaust the water of a well; the moisture of the earth is exhausted by evaporation.</td>
</tr>
<tr>
<td>2</td>
<td>Orb</td>
<td>NOUN</td>
<td>A period of time marked off by the revolution of a heavenly body.</td>
</tr>
<tr>
<td>3</td>
<td>Conduct</td>
<td>NOUN</td>
<td>Act; action; construction; manner of development.</td>
</tr>
<tr>
<td>4</td>
<td>Cop</td>
<td>NOUN</td>
<td>A tube or quill upon which ink is wound.</td>
</tr>
<tr>
<td>5</td>
<td>Sheep</td>
<td>NOUN</td>
<td>The female of the sheep, and of sheeplike animals.</td>
</tr>
<tr>
<td>6</td>
<td>Drill</td>
<td>VERB</td>
<td>To pierce or bore with a drill, or with a drill; to perforate; as, to drill a hole in a rock; to drill a piece of metal.</td>
</tr>
<tr>
<td>7</td>
<td>Play</td>
<td>VERB</td>
<td>To play, as a horse.</td>
</tr>
<tr>
<td>8</td>
<td>Superlative</td>
<td>NOUN</td>
<td>The superlative.</td>
</tr>
<tr>
<td>9</td>
<td>Evidence</td>
<td>NOUN</td>
<td>That which is legally submitted to competent tribunal, as a means of ascertaining the truth of any alleged matter of fact under investigation.</td>
</tr>
<tr>
<td>10</td>
<td>Hopscotch</td>
<td>NOUN</td>
<td>A game played on a yard or alley.</td>
</tr>
<tr>
<td>11</td>
<td>Manis</td>
<td>NOUN</td>
<td>A genus of entedantes, covered with large, hard, triangular scales, with sharp edges that overlap each other like tiles on a roof. They inhabit tropical and subtropical areas.</td>
</tr>
<tr>
<td>12</td>
<td>Currency</td>
<td>NOUN</td>
<td>The state or quality of being current; general acceptance or reception; a passing from person to person, or from hand to hand; circulation.</td>
</tr>
<tr>
<td>13</td>
<td>Verdict</td>
<td>NOUN</td>
<td>A decision or judgment of a court of law; an authoritative declaration on a point of law.</td>
</tr>
<tr>
<td>14</td>
<td>Indifference</td>
<td>NOUN</td>
<td>The quality or state of being indifferent, or not making a difference; want of sufficient importance to constitute a difference.</td>
</tr>
<tr>
<td>15</td>
<td>Solution</td>
<td>NOUN</td>
<td>The act or process by which a body (whether solid, liquid, or gaseous) is absorbed into a liquid, and, remaining or becoming fluid, passes into the solution.</td>
</tr>
<tr>
<td>16</td>
<td>Tor</td>
<td>NOUN</td>
<td>A tower; a turret.</td>
</tr>
<tr>
<td>17</td>
<td>Hog</td>
<td>NOUN</td>
<td>A rough, flat scrubbing broom for scrubbing a ship's bottom under water.</td>
</tr>
<tr>
<td>18</td>
<td>Pica</td>
<td>NOUN</td>
<td>A service-book. See Ple.</td>
</tr>
<tr>
<td>19</td>
<td>Met</td>
<td>NOUN</td>
<td>A pithy or witty saying; a witticism.</td>
</tr>
<tr>
<td>20</td>
<td>Zone</td>
<td>NOUN</td>
<td>An area or part of a region characterized by uniform or similar animal and plant life; a life zone; as, Littoral zone, Austral zone, etc.</td>
</tr>
<tr>
<td>21</td>
<td>Marshal</td>
<td>NOUN</td>
<td>The chief officer of arms, whose duty it was, in ancient times, to regulate contests in the lists.</td>
</tr>
<tr>
<td>22</td>
<td>Hasp</td>
<td>NOUN</td>
<td>An implement of iron or steel, consisting of pieces of timber or metal crossing each other, and set with iron or wooden teeth. It is driven into the wood and used as a hinge.</td>
</tr>
<tr>
<td>23</td>
<td>Requisition</td>
<td>NOUN</td>
<td>A demand by the invaders upon the people of an invaded country for supplies, as of provision, forage, transportation, etc.</td>
</tr>
<tr>
<td>24</td>
<td>Damnation</td>
<td>NOUN</td>
<td>A sin deserving of everlasting punishment.</td>
</tr>
<tr>
<td>25</td>
<td>Atmosphe</td>
<td>NOUN</td>
<td>The density or weight of the air at the sea level, on a unit of surface, or about 14 7 lbs. to the sq. inch.</td>
</tr>
<tr>
<td>26</td>
<td>Dictionary</td>
<td>NOUN</td>
<td>A book containing the words belonging to any system or province of knowledge, arranged alphabetically; as, a dictionary of Greek, Latin, or other languages; a dictionary of the Bible; a dictionary of popular science.</td>
</tr>
<tr>
<td>27</td>
<td>Sculpture</td>
<td>NOUN</td>
<td>The art of carving, cutting, or hewing wood, stone, metal, etc., into statues, ornaments, etc., or into figures, as of men, or of animals.</td>
</tr>
<tr>
<td>28</td>
<td>Ciliary</td>
<td>ADJECTIVE</td>
<td>Pertaining to or connected with the cilia in animal or vegetable organisms; as, ciliary motion.</td>
</tr>
<tr>
<td>29</td>
<td>Side</td>
<td>NOUN</td>
<td>Any outer portion of a thing considered apart from, and yet in relation to, the rest; as, the upper side of a sphere; also, any part or portion of a thing.</td>
</tr>
<tr>
<td>30</td>
<td>Sympathy</td>
<td>NOUN</td>
<td>An agreement of affections or dispositions, or a conformity of natural temperament, which causes persons to be pleased, or in accord with each other.</td>
</tr>
<tr>
<td>31</td>
<td>Wash</td>
<td>NOUN</td>
<td>Substances collected and deposited by the action of water; as, the wash of a sewer, a river, etc.</td>
</tr>
<tr>
<td>32</td>
<td>Root</td>
<td>NOUN</td>
<td>The underground portion of a plant, whether a true root or a tuber, a bulb or rootstock, as in the potato, the onion, or the sweet flag.</td>
</tr>
<tr>
<td>33</td>
<td>Justification</td>
<td>NOUN</td>
<td>The act of justifying or the state of being justified; a showing or proving to be just or conformable to law, justice, right, right, etc.</td>
</tr>
<tr>
<td>34</td>
<td>Pacer</td>
<td>NOUN</td>
<td>One who, or that which, paces; especially, a horse that paces.</td>
</tr>
<tr>
<td>35</td>
<td>Bile</td>
<td>NOUN</td>
<td>Bitterness of feeling; choleric; anger; ill humor; as, to stir one's bile.</td>
</tr>
<tr>
<td>36</td>
<td>Jacob</td>
<td>NOUN</td>
<td>A Hebrew patriarch (son of Isaac, and ancestor of the Jews), who in a vision saw a ladder reaching up to heaven (Gen. xxviii. 12).</td>
</tr>
<tr>
<td>37</td>
<td>Repeater</td>
<td>NOUN</td>
<td>A device with a striking apparatus which, upon pressure of a spring, will indicate the time, usually in hours and quarters.</td>
</tr>
<tr>
<td>38</td>
<td>Baronet</td>
<td>NOUN</td>
<td>A company; an assembly of persons, especially of ladies.</td>
</tr>
<tr>
<td>39</td>
<td>Crown</td>
<td>NOUN</td>
<td>The person entitled to wear a regal or imperial crown; the sovereign.</td>
</tr>
<tr>
<td>40</td>
<td>Miner</td>
<td>NOUN</td>
<td>One who mines; a digger for metals, etc.; one engaged in the business of getting ore, coal, or precious stones, out of the earth; one who works in a mine.</td>
</tr>
<tr>
<td>41</td>
<td>Chivalry</td>
<td>NOUN</td>
<td>Exploit.</td>
</tr>
<tr>
<td>42</td>
<td>Tinker</td>
<td>NOUN</td>
<td>A young maker about two years old.</td>
</tr>
<tr>
<td>43</td>
<td>Charge</td>
<td>NOUN</td>
<td>Thirty-six pigs of lead, each pig weighing about seventy pounds.</td>
</tr>
<tr>
<td>44</td>
<td>Virtuous</td>
<td>ADJECTIVE</td>
<td>Having moral excellence; characterized by morality; upright; righteous; pure; as, a virtuous action.</td>
</tr>
<tr>
<td>45</td>
<td>Visibility</td>
<td>NOUN</td>
<td>The quality or state of being visible.</td>
</tr>
<tr>
<td>46</td>
<td>Word</td>
<td>NOUN</td>
<td>To defend by walls, fortifications, etc.</td>
</tr>
</tbody>
</table>
| 47 | Tournament  | NOUN                 | An instrument for ascertaining hooperage. It consists essentially of a rod or compass, upon which pressure is made by a hand which...
Case study 0: Hello world (a minimal fake resource)

ARE YOU READY?
Case study 0: Hello world (a minimal fake resource)
Case study 0: Hello world (a minimal fake resource)
Case study 0: Hello world (a minimal fake resource)

- First, let's see how to perform a semi-automatic mapping (for **Browser Users**)

- Then we will create a simple program for automatically mapping via the BabelNet and Babelfy Java APIs (for **Java Programmers**)

The Luxembourg BabelNet Workshop - Session 6
Roberto Navigli
Case study 0: Hello world (a minimal fake resource)

- First, let's see how to perform a semi-automatic mapping (for Browser Users)
Case study 0: Hello world (a minimal fake resource)

Steps to follow to perform a **semi-automatic mapping**

<table>
<thead>
<tr>
<th>Code</th>
<th>Word</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>1084</td>
<td>Cow</td>
<td>NOUN A chimney cap; a cowl</td>
</tr>
<tr>
<td>1085</td>
<td>Press</td>
<td>NOUN An apparatus or machine by which any substance or body is pressed,</td>
</tr>
<tr>
<td>1086</td>
<td>Lap</td>
<td>NOUN To lay or place over anything so as to partly or wholly cover it;</td>
</tr>
<tr>
<td>1087</td>
<td>Lotus</td>
<td>NOUN The lote, or nettle tree. See Lote.</td>
</tr>
<tr>
<td>1088</td>
<td>Angle</td>
<td>NOUN The figure made by two lines which meet.</td>
</tr>
<tr>
<td>1089</td>
<td>Cypris</td>
<td>NOUN A lewd woman; a harlot.</td>
</tr>
<tr>
<td>1090</td>
<td>Uncertainty</td>
<td>NOUN That which is uncertain; something unknown.</td>
</tr>
<tr>
<td>1091</td>
<td>Base</td>
<td>NOUN A place or tract of country, protected by fortifications, or by na</td>
</tr>
</tbody>
</table>

- Copy a **lemma** defined in our resource to Babelfy.org
Case study 0: Hello world (a minimal fake resource)

- Copy a **lemma** defined in our resource to Babelfy.org
Case study 0: Hello world (a minimal fake resource)

- Also add the **Glosses** available for that lemma
Case study 0: Hello world (a minimal fake resource)

- Also add the **Glosses** available for that lemma

Click!!
Hello world (a minimal fake resource)

The Luxembourg BabelNet Workshop - Session 6

Roberto Navigli
The space between two lines or planes that intersect; the inclination of one line to another; measured in degrees or radians. The angle between lines A and B.
The space between two lines or planes that intersect; the inclination of one line to another; measured in degrees or radians. The angle between lines A and B.

The figure made by two lines which meet.
Case study 0: Hello world (a minimal fake resource)

- Babelfy disambiguates "Angle" with the synset bn:00004117n
- We add this answer to the annotation form created for this session
- Type this URI on your browser: http://goo.gl/forms/gNm7wRPJJX
- OR go to the workshop site (http://babelnet.org/lux) session 6 and click on "form".
- Answer the questions! :D
Case study 0: Hello world (a minimal fake resource)

The Luxembourg BabelNet Workshop - Session 6

Result of mapping using Babelfy on Browser

*Campo obbligatorio

ID FAKE_RESOURCE *

La tua risposta

ID BABELNET (e.g bn:00000001n or null) *

La tua risposta

IS CORRECT? *

- Yes!
- No!

Send
Case study 0: Hello world (a minimal fake resource)
Case study 0: Hello world (a minimal fake resource)

- Try yourself to map the fake resource to BabelNet
- Choose some random entries
- Apply the above steps
- Fill in the form for each mapped entry (type null if the target word cannot be disambiguated)
Case study 0: Hello world (a minimal fake resource)
Case study 0: Hello world (a minimal fake resource)

- We will create a simple program for automatically mapping through the BabelNet and Babelfy Java API (for Java Programmers)
Case study 0: Hello world (a minimal fake resource)

• Create a project containing
  – The BabelNet API jar
  – The Babelfy API jar

• Create a simple mapping class (Mapping) that:
  – Reads the fake resource (FR) line by line
  – Retrieves all the synsets for the target lemma with the BabelNet API
  – Disambiguates the lemma and its definitions using Babelfy
  – Checks if the target lemma has been disambiguated
  – Returns the synset provided by Babelfy for that lemma, if available; else, returns null
    • If you are also using the BabelNet API, you can retrieve more information about
Case study 0: Hello world (a minimal fake resource)

- Create a project containing
  - The BabelNet API jar
  - The Babelfy API jar

Like we did yesterday during the Babelfy tech session
Case study 0: Hello world (a minimal fake resource)

- Create a simple mapping class that:
  - Reads the fake resource (FR) line by line
Case study 0: Hello world (a minimal fake resource)

- Create a simple mapping class that:
  - Reads the fake resource (FR) line by line
    - Just neglect the greyed parts for the moment
- Code:
  - Open the file
  - Parse each line
  - Close the file

```java
private void mapping() {
    try {
        BabelNet bn = BabelNet.getInstance();
        BufferedReader reader = Files.getBufferedReader(PATH+"fake_resource.txt");
        BufferedWriter bufferedWriter = Files.getBufferedWriter(PATH+"mapping_fake_resource.txt");
        String line = null;
        //read file
        while (((line = reader.readLine()) != null) ) {
            String[] fields = line.split("\t");
            String id = fields[0];
            String lemma = fields[1];
            BabelPOS pos = BabelPOS.valueOf(fields[2]);
            String glosses = fields[3];
            HashSet<BabelSynsetID> candidates = new HashSet<BabelSynsetID>();
            //for(BabelSynset by : bn.getSynsets(lemma, Language.EN, pos))
            //{
            //    candidates.add(by.getId());
            //}
            //BabelSynsetID idMapping = applyBabelfy(lemma+", "+glosses, candidates);
            if(idMapping!=null) {
                BabelSynset babelSynset = idMapping.toBabelSynset();
                bufferedWriter.append(id+"\t"+lemma+"\t"+pos+"\t"+idMapping+"\n"
                        + "BABELNET:\"+babelSynset.getMainGloss(Language.EN)+"\n"
                        + "FAKE_RESOURCE:\"+glosses+"\n\n";
            }
        }
    }
    bufferedWriter.flush();
    bufferedWriter.close();
    reader.close();
    } catch (FileNotFoundException e) {
        e.printStackTrace();
    } catch (IOException e) {
        e.printStackTrace();
    }
}
```
Case study 0: Hello world (a minimal fake resource)

• Create a simple mapping class that:
  – Retrieves all synsets in BabelNet for the target lemma to be mapped
Case study 0: Hello world (a minimal fake resource)

- Create a simple mapping class that:
  - Retrieves all synsets in BabelNet for the target lemma to be mapped

- Just neglect the greyed parts for the moment

- Code:
  - Obtain the BabelNet object
  - Retrieve synsets that contain the candidate lemma in English
  - Add their IDs to a set called "candidates"

```java
private void mapping()
{
    try {
        BabelNet bn = BabelNet.getInstance();
        BufferedReader reader = Files.getBufferedReader(PATH+"fake_resource.txt");
        BufferedReader bufferedWriter = Files.getBufferedReader(PATH+"mapping_fake_resource.txt");
        String line = null;
        //read file
        while ((line = reader.readLine()) != null) {
            String[] fields = line.split("\t");
            String id = fields[0];
            String lemma = fields[1];
            BabelPOS pos = BabelPOS.valueOf(fields[2]);
            String glosses = fields[3];
            HashSet<BabelSynsetId> candidates = new HashSet<BabelSynsetId>();
            //
            for(BabelSynset by : bn.getSynsets(lemma, Language.EN, pos)) {
                candidates.add(by.getId());
            }
            BabelSynsetId idMapping = applyBabelify(lemma+, "+glosses, candidates);;
            if(idMapping!=null) {
                BabelSynset babelSynset = idMapping.toBabelSynset();
                bufferedWriter.append(id+"\t"+lemma+"\t"+pos+"\t"+idMapping+"\n"
                        + "BADELNET:\t"+babelSynset.getMainGloss(Language.EN)+"\n"
                        + "FAKE_RESOURCE:\t"+glosses+"\n\n");
            }
        }
    }
    bufferedWriter.flush();
    bufferedWriter.close();
    reader.close();
} catch (FileNotFoundException e) {
    e.printStackTrace();
} catch (IOException e) {
    e.printStackTrace();
}
```
Case study 0: Hello world (a minimal fake resource)

- Create a simple mapping class that:
  - Disambiguates the target lemma and its definitions using Babelfy
Case study 0: Hello world (a minimal fake resource)

- Create a simple mapping class that:
  - Disambiguates target lemmas using Babelfy

- Just neglect the greyed parts for the moment

- Code:
  - Retrieve the lemma and its definitions and pass them to the disambiguation method "applyBabelfy" together with the candidate synsets for the target lemma

```java
private void mapping()
{
    try {
        BabelNet bn = BabelNet.getInstance();
        BufferedReader reader = Files.getBufferedReader(PATH+"fake resource.txt");
        BufferedReader bufferedWriter = Files.getBufferedWriter(PATH+"mapping fake resource.txt");
        String line = null;
        //read file
        while (((line = reader.readLine()) != null)) {
            String[] fields = line.split("\t");
            String id = fields[0];
            String lemma = fields[1];
            BabelPOS pos = BabelPOS.valueOf(fields[2]);
            String glosses = fields[3];
            HashSet<BabelSynsetID> candidates = new HashSet<BabelSynsetID>();
            //
            for(BabelSynset by : bn.getSynsets(lemma, Language.EN, pos))
            {
                candidates.add(by.getId());
            }
            BabelSynsetID idMapping = applyBabelfy(lemma+, "+glosses, candidates
            {;
            }
            }
            bufferedWriter.flush();
            bufferedWriter.close();
            reader.close();
    } catch (FileNotFoundException e) {
        e.printStackTrace();
    } catch (IOException e) {
        e.printStackTrace();
    }
}
```
Case study 0: Hello world (a minimal fake resource)

- Create a simple mapping class that:
  - Disambiguates the target lemma and its definitions using Babelfy

  ```java
  private BabelSynsetID applyBabelfy(String glosses, HashSet<BabelSynsetID> ids) {
    Babelfy bfy = new Babelfy();
    List<SemanticAnnotation> bfyAnnotations = bfy.babelfy(glosses, Language.EN);
    HashSet<BabelSynsetID> synsetsInGloss = new HashSet<BabelSynsetID>();
    for (SemanticAnnotation annotation : bfyAnnotations) {
      synsetsInGloss.add(new BabelSynsetID(annotation.getBabelSynsetID()));
    }
    SetView<BabelSynsetID> intersection = Sets.intersection(ids, synsetsInGloss);
    if (intersection.isEmpty()) return null;
    if (intersection.size() > 1) return null;
    else return intersection.iterator().next();
  }
  ```

- Code:
  - Obtain the Babelfy object
  - Disambiguate the lemma with its context
  - Retrieve all the disambiguated concepts
  - Check if there is an intersection with the candidate synsets in input
  - If only a single concept is in the intersection, we return it; else, we return null
Case study 0: Hello world (a minimal fake resource)

- Create a simple mapping class that:
  - Returns the Babel synset corresponding to the input entry
Case study 0: Hello world (a minimal fake resource)

- Create a simple mapping class that:
  - Returns the Babel synset corresponding to the input entry

  ```java
  private void mapping() {
    try {
      BabelNet bn = BabelNet.getInstance();
      BufferedReader reader = Files.getBufferedReader(PATH+"fake_resource.txt");
      BufferedWriter bufferedWriter = Files.getBufferedWriter(PATH+"mapping_fake_resource.txt");
      String line = null;
      // read file
      while ((line = reader.readLine()) != null) {
        String[] fields = line.split("\t");
        String id = fields[0];
        String lemma = fields[1];
        String BabelPOS = BabelPOS.valueOf(fields[2]);
        String glosses = fields[3];
        HashSet<BabelSynsetId> candidates = new HashSet<BabelSynsetId>();
        //
        for (BabelSynset by : bn.getSynsets(lemma, Language.EN, pos)) {
          candidates.add(by.getId());
        }
        //
        BabelSynsetId idMapping = applyBabelty(lemma+"\t"+glosses, candidates);
        if (idMapping!=null) {
          BabelSynset babelSynset = idMapping.toBabelSynset();
          bufferedWriter.append(id+"\t"+lemma+"\t"+pos+"\n"+
                        +"BAELNET:\t"+babelSynset.getMainGloss(Language.EN)+"\n"
                        +"FAKE_RESOURCE:\t"+glosses+"\n\n");
        }
      }
    }
    bufferedWriter.flush();
    bufferedWriter.close();
    reader.close();
  } catch (FileNotFoundException e) {
    e.printStackTrace();
  } catch (IOException e) {
    e.printStackTrace();
  }
  }
  ```

- Codice:
  - Create the output file
  - Check if a valid candidate exists
  - Write the output file
Case study 0: Hello world (a minimal fake resource)

- Run the class:

```java
public static void main(String[] args) throws IOException {
    Mapping test = new Mapping();
    test.mapping();
}
```
Case study 0: Hello world (a minimal fake resource)

• In summary, we have just seen how to map a simple resource by using a manual and an automatic procedure, both based on Babelfy

• In the next sessions, we will see more complex strategies to map EU resources!