Interconnecting lexicographic resources. A first proof of Michael’s model

Dan Cristea and Andrei-Liviu Scutelnicu
“Alexandru Ioan Cuza” University of Iași
Institute of Computer Science of the Romanian Academy

dcristea@info.uaic.ro
liviu.scutelnicu@info.uaic.ro
Investigation

• What to do when one resource does not bring us to the expected result?
  – mix resources of different types

• Could a mix of resources be more successful?
  – doing for seeing…
Investigation

• What kind of resources could we mix?
  – for the time being: a WN, an explanatory dictionary
  – in the future: a corpus (access to contexts), Wikipedia/DBpedia, a collection of dictionaries, etc.

• At what price?
  – ad-hoc programming right now => expensive
  – big expectation for a nice generalisation => cheap
How is the WN connected with the Dictionary?

- A **tight** alignment: WN lexicals in synsets aligned with senses of entries in the explanatory dictionary (ExDi)
  - a WN synset:
    $$\text{pos (def, ex, } w_{1}^{s1} \ldots w_{k}^{sk} \ldots w_{n}^{sn})$$
  - an explanatory dictionary entry:
    $$w_{k}, \text{pos, } <w_{k}^{s1}, \text{def}_{1}, \text{ex}_{1}> \ldots <w_{k}^{sk}, \text{def}_{k}, \text{ex}_{k}> \ldots <w_{k}^{sm}, \text{def}_{m}, \text{ex}_{m}>$$
How is the WN connected with the Dictionary?

• In the actual implementation, a **light** alignment: WN lexicals in synsets aligned with title words in ExDi

  – a WN synset:
    
    \[
    \text{pos (def, ex, } w_1 \ldots w_k \ldots w_n) \]

  – an explanatory dictionary entry:
    
    \[
    w_k, \text{ pos, } <w_k^{s1}, \text{def}_1, \text{ex}_1> \ldots <w_k^{sk}, \text{def}_k, \text{ex}_k> \ldots <w_k^{sm}, \text{def}_m, \text{ex}_m> \]

3rd COST-ENeL meeting, Vienna, 11-13 February 2015
Algorithm

• A source word $w$
  
  $\Rightarrow$ extract definitions of $w$ from ExDi: $defs$
  
  $\Rightarrow$ search $w$ in WN and get its domain: $d$
  
  $\Rightarrow$ filter the words belonging to $defs$ and $d$: $f$
  
  $\Rightarrow$ merge all literals belonging to the synsets of $f$ in WN: $m$
  
  $\Rightarrow$ cluster $m$: replace the list $m$ with their hypernyms
  
  $\Rightarrow$ let the user choose among these clusters: $c$
  
  $\Rightarrow$ display the cluster $c$
  
  $\Rightarrow$ iterate if the looked for word in not in the list $c$

3rd COST-ENeL meeting, Vienna, 11-13 February 2015
En example...

• Start word: *abate* (EN: *superior* – religion)
• DEX-online: collection of definitions in many dictionaries: 7 entries for *abate*
• End of Step 1: 601 words, the majority in the religious domain (among them - the goal word *vicar*)
Results till now

• Do be done
  – clusterisation: take hypernyms of all these words => #hypernyms – less than the initial list.
  – at a certain level up, the category "human" should pop up.
  – the intersection of all its hyponyms with the expansion initial list should be much smaller in size and, of course, among them – the target word.

• Problem: out of what level up should we take the hypernyms? Is it simply a matter of quantity?

3rd COST-ENeL meeting, Vienna, 11-13 February 2015
Thank you!