Alignment of Concepts and the Hierarchies

Virach Sornlertlamvanich
virach@nwg.nectec.or.th
NECTEC
Bangkok, Thailand.

22 Feb 1999
Why need alignment of concepts and the hierarchies?

- Only 20%-30% of the concepts in each language are shared pairwise, and 10%-20% of them are shared among the languages, in CICC MMT project.

- Less than 10% of the concepts in EDR Japanese and English word dictionaries are shared.

⇒ Though we are aiming at creating a common set of concepts.
Necessity of a concept hierarchy

- Semantic restriction:

\[(c\#\text{boy} \leftarrow \text{supc} - c\#\text{human}) \]
\[(c\#\text{girl} \leftarrow \text{supc} - c\#\text{human}) \]
\[+\]
\[(c\#\text{boy} \leftarrow \text{agt} - c\#\text{speak}) \]
\[(c\#\text{girl} \leftarrow \text{agt} - c\#\text{speak}) \]
\[\vdots\]
\[\downarrow\]
\[(c\#\text{human} \leftarrow \text{agt} - c\#\text{speak})\]

- Alternative interpretation:

\[(c\#\text{sashimi} \leftarrow \text{supc} - c\#\text{fish}) \]
\[+\]
\[(c\#\text{sashimi} \leftarrow \text{obj} - c\#\text{eat}) \]
\[\downarrow\]
\[(c\#\text{fish} \leftarrow \text{obj} - c\#\text{eat})\]
Diversity of the concept definitions

“tired”

• EDR concept description
  - “having or displaying a need for rest or an exhaustion of physical or mental strength”
  - “having lost interest”
  - “revealing a dearth of imaginativeness or originality”

• Wordnet 1.5
  - A1: tired (vs. rested)
  - A2: bromidic, commonplace, hackneyed, shopworn, threadbare, timeworn, tired, trite, well-worn
  - V1: tire, pall, grow weary, weary, fatigue, get tired, jade
  - V2: tire, wear upon, tire out, wear, weary, jade, wear out, outwear, wear down, fag out, fag, fatigue
  - V3: run down, exhaust, sap, tire, use up
  - V4: bore, tire

• UW
  - “tired”
  - “tired(agt > use)”
  - “tired(aoj > joke)”
  - “tired(aoj > thing)”
  - “tired(aoj > volitionalthing)”
  - “tired(gol > activity)”
  - “tired(idl > #state)”
  - “tired(idl > bodycondition)”
  - “tired(idl > do)”
  - “tired(idl > occur)”
  - “tired(idl > tiredness)”
Concept alignment in MMTS

Word dictionary

- w1 - e1
- e2
- e3

EDR concept dictionary

- e1 - id11:d11
- id12:d12
- id13:d13
- e2 - id21:d21
- id22:d22
- e3 - id31:d31
- id32:d32

Merged dictionary

- w1 - e1 - id11:d11
- id12:d12
- id13:d13
- e2 - id21:d21
- id22:d22
- e3 - id31:d31
- id32:d32

Aligned dictionary

- w1 - e1 - id12:d12
- w1 - e3 - id31:d31
Degrees of concept alignment in MMTS

1. The sense of the word is equivalent to the assigned concept. \((s(w_i) \equiv c_i)\)

2. The sense of the word is wider than the assigned concept. \((s(w_i) \supset c_i)\)

3. The sense of the word is narrower than the assigned concept. \((s(w_i) \subset c_i)\)

4. The sense of the word has some relations with the assigned concept. \((s(w_i) \sim c_i)\)

5. The original word sense. \((s(w_i))\)
(1) Concept Composition

- A word concept in language A corresponding to a composite concept in language B.

\[
\begin{align*}
(c\#\text{break}) \\
(c\#\text{cause}[\neg \text{obj} \rightarrow c\#\text{nil}] & \leftarrow a-\text{obj}-c\#\text{broken}) \\
(c\#\text{do} - \text{goal} & \rightarrow c\#\text{broken})
\end{align*}
\]

IL1:  
\[
\text{c\#break} \\
\text{agent} \quad \text{object} \\
\text{c\#he} \quad \text{c\#glass}
\]

IL2:  
\[
\text{c\#cause} \\
\text{agent} \quad \text{object} \\
\text{c\#he} \quad \text{c\#glass} \\
\text{a-object}
\]

IL3:  
\[
\text{c\#do} \\
\text{agent} \quad \text{goal} \quad \text{object} \\
\text{c\#he} \quad \text{c\#glass} \quad \text{c\#broken}
\]
(2) Concept Divergency

- A word concept in language A corresponding to a meta-concept in language B.

\[(c\#crow - \text{agt} \rightarrow c\#cock)\]
\[(c\#sing - \text{agt} \rightarrow c\#bird)\]
\[(c\#niwatori \sim c\#bird)\]

\[\downarrow\]

\[(c\#sing - \text{agt} \rightarrow c\#niwatori) \cdots ?\]
\[(c\#crow - \text{agt} \rightarrow c\#niwatori)\]
(3) Concept Granularity

- A word concept in language A corresponding to multiple concepts in language B.
- A gap between defining the word senses.

<table>
<thead>
<tr>
<th>Japanese</th>
<th>English</th>
</tr>
</thead>
<tbody>
<tr>
<td>sakura (tree)</td>
<td>cherry tree</td>
</tr>
<tr>
<td>(flower)</td>
<td>cherry blossom</td>
</tr>
<tr>
<td>sakuranbo (fruit)</td>
<td>cherry</td>
</tr>
<tr>
<td></td>
<td>(fruit)</td>
</tr>
<tr>
<td></td>
<td>(tree)</td>
</tr>
</tbody>
</table>
Conclusion

Concepts and the hierarchies are dynamically changed. To keep the lexical knowledge and the hierarchy in a manageable size:

⇒ **Flexible (dynamic) concept hierarchy.**
⇒ **Expressive concept.**
⇒ **Concept insertion/deletion, composition/decomposition.**