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.

$\Rightarrow$ Though we are aiming at creating a common set of concepts.

Necessity of a concept hierarchy

- Semantic restriction:

$$(c\#boy \leftarrow supc - c\#human)$$
$$(c\#girl \leftarrow supc - c\#human)$$
$$+$$
$$(c\#boy \leftarrow agt - c\#speak)$$
$$(c\#girl \leftarrow agt - c\#speak)$$
$$\vdots$$
$$\Downarrow$$
$$(c\#human \leftarrow agt - c\#speak)$$

- Alternative interpretation:

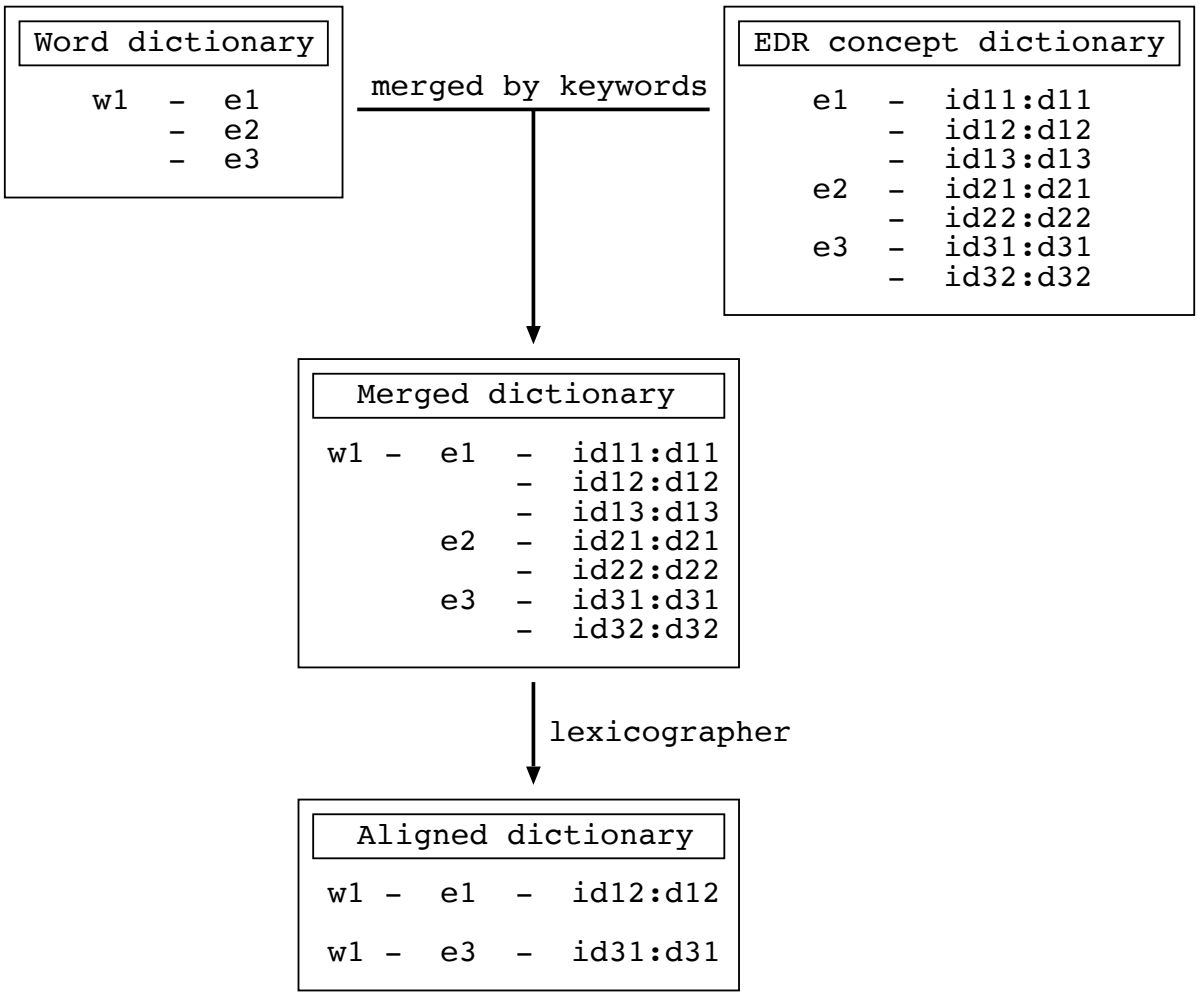
$$(c\#sashimi \leftarrow supc - c\#fish)$$
$$+$$
$$(c\#sashimi \leftarrow obj - c\#eat)$$
$$\Downarrow$$
$$(c\#fish \leftarrow obj - c\#eat)$$

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($icl > \#state$)"
  - "tired($icl > bodycondition$)"
  - "tired($icl > do$)"
  - "tired($icl > occur$)"
  - "tired($icl > tiredness$)"

Concept alignment in MMTS

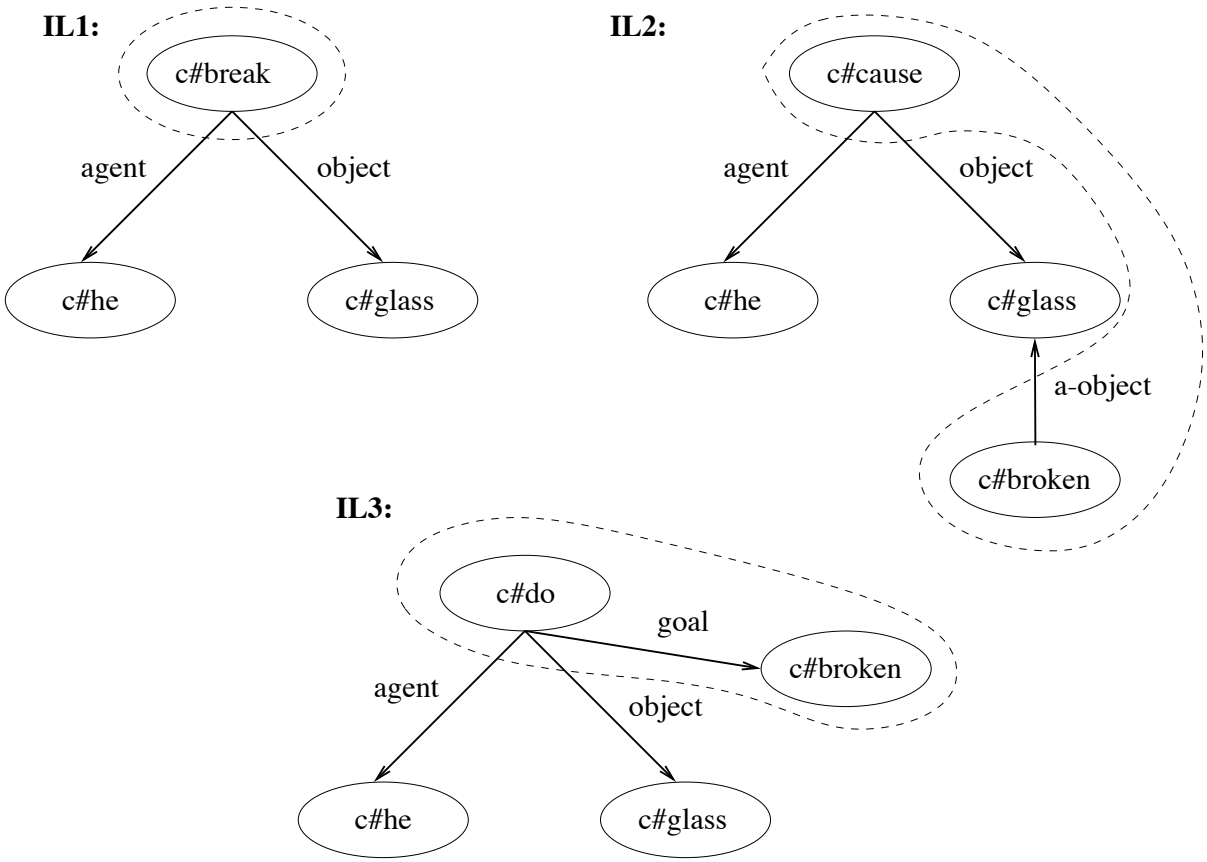


Degrees of concept alignment in MMTS

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.
  $(c\#break)$
  $(c\#cause[-obj \rightarrow c\#nil] \leftarrow a-obj-c\#broken)$
  $(c\#do - gol \rightarrow c\#broken)$



(2) Concept Divergency

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

  $(c\#crow - agt \rightarrow c\#cock)$

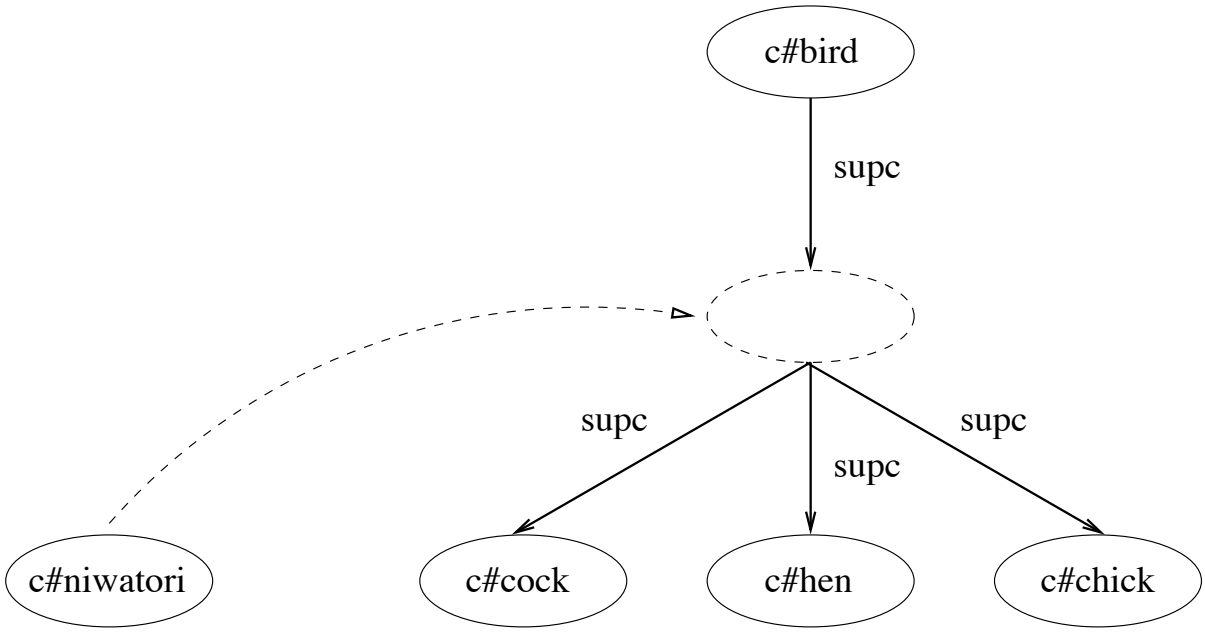
  $(c\#sing - agt \rightarrow c\#bird)$

  $(c\#niwatori \sim c\#bird)$

  $\Downarrow$

  $(c\#sing - agt \rightarrow c\#niwatori) \cdots ?$

  $(c\#crow - agt \rightarrow c\#niwatori)$



(3) Concept Granularity

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

Japanese		English	
sakura	(tree)	cherry tree	
	(flower)	cherry blossom	
sakuranbo	(fruit)	cherry	(fruit)
			(tree)

Conclusion

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

- $\Rightarrow$ Flexible (dynamic) concept hierarchy.
- $\Rightarrow$ Expressive concept.
- $\Rightarrow$ Concept insertion/deletion, composition/decomposition.