

# Deeper Linguistic Processing with HPSG and WordNet

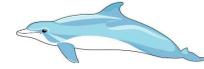
Francis Bond

**Division of Linguistics and Multilingual Studies**

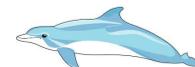
Nanyang Technological University

`<bond@ieee.org>`

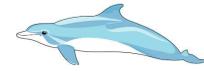
2011-06-25



- Combine lexical semantics (WN) and structural semantics (MRS)
- Why
- How
- When
- Who



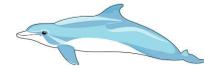
- lexical semantics helps in parse ranking structural semantics  
Fujita et al, Egirre et al.
- Only with back-off up the semantic hierarchy
- structural semantics helps WSD  
Stevenson, Tanaka et al
- semantic classes useful in transfer rules  
nomu(medicine) → take  
nomu(liquid) → drink  
nomu(condition) → accept



- 
- Make each predicate a type
  - Include WN as a type hierarchy
  - . . .
  - Profit

# Advantages

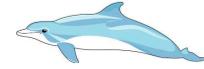
---



- Conceptually simple
- Easy to experiment with



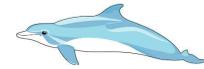
- Slow with current machinery (loading WN as a TFS > 10 min)
- Many senses increase ambiguity a lot
  - but eminently packable
  - can super-sense-tag the input to prune
- MWE/concept differences  
Quite a lot of work to harmonize (well worth doing)
- Unknown words/NEs



➤ First idea (replace *dog* with 0203459-n)  $\left[ \begin{array}{l} \text{PRED } 0203459\text{-n} \\ \text{ARG0 } x \end{array} \right]$

➤ Add concept as a separate property of the MRS  $\left[ \begin{array}{l} \text{PRED } \text{dog} \\ \text{CONCEPT } 0203459\text{-n} \\ \text{ARG0 } x \end{array} \right]$

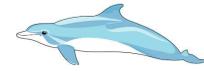
➤ Add concept as a separate property of the MRS  $\left[ \begin{array}{l} \text{PRED } \text{dog} \\ \text{ARG0 } x \left[ \text{CONCEPT } 0203459\text{-n} \right] \end{array} \right]$



- Large Free WordNets
  - English, Japanese, Spanish
- Large Freeish WordNets
  - Chinese, Portuguese, German
- Tagged WN corpora
  - English, Japanese, Chinese

# When

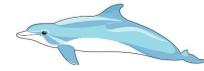
---



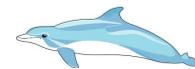
➤ Hopefully in the next year or so

# Who

---



➤ ???



- Better not to unify semantics (that way lies madness)
- The right interface is the SEM-I
- We can decorate the index without troubling the grammar
- Several applications would like this information
- Better not to commit to a single ontology