Knowledge Engineering for NLP

January 29, 2007

More Matrix tour (what happened?)

Case, Agreement

Debugging itsdb?

Overview

- What did the customization script do?
- Case
- Agreement
- Debugging [incr tsdb()]?

What did the customization script do?

- Created subtypes
- Added constraints to subtypes
- Created instances (lexical entries, grammar rules, lexical rules)
- $\rightarrow$ take a look

Case

- Who has a language with case?
- What is our general strategy going to be?
- What types need additional constraints?
- Do you need any additional types?
- Do you need any lexical rules?

Case: General strategy (1/2)

- If you language has inflectional case...
- Define a feature CASE appropriate for *noun* (and...?)
- Define a type *case* and appropriate subtypes.
- Define subtypes of *verb-lex* (may already be defined) and constrain CASE on ARG-ST elements.

Case: General strategy (2/2)

- Define a lexical rule to add case endings to nouns, and/or
- Define lexical entries for pronouns with CASE information, and/or
- Define lexical entries for determiners with CASE information (and constrain nouns to require dets with matching case)

What about case-marking adpositions?

- If your language always has an adposition:
- Verbs select for *adp* complements, with appropriate CASE values
- CASE is appropriate for *adp*
- Adpositions specify their CASE values
- Case-marking adpositions are semantically empty

And another kind of case marking adp

- If your language has argument positions that can either be filled by PP or NP (for the same verb):
- Verbs select for *+np* complements, with appropriate CASE values.
- CASE is appropriate for *adp*
- Nouns are probably underspecified for CASE.
- Adpositions specify an appropriate CASE value.
- Case-marking adpositions are semantically empty

Agreement: General strategy

- Distinguish semantic from syntactic agreement
- Pronoun-antecedent agreement is semantic
- Until we need distinct syntactic agreement, keep any agreement involving the same features semantic.
- (Semantic) agreement features live on the INDEX.

Agreement: General strategy

- Noun classes (inherited by specific lexical entries) specify inherent properties (like person or gender/noun class)
- Inflectional rules for nouns specify varying properties (number)
- Lexical rules for agreeing verbs constrain the INDEX.PNG of their arguments appropriately.
- Agreeing modifiers/determiners specify the INDEX.PNG of their MOD/SPEC appropriately (through lexical types or lexical rules)

Lab 5 planning

- Who has case and agreement?
- Who has just case and no agreement?
- Who has agreement and numeral classifiers?
- Anyone else?

Overview

- What did the customization script do?
- Case
- Agreement
- Debugging [incr tsdb()]