



NL-Soar update

Deryle Lonsdale
BYU Linguistics
lonz@byu.edu

The logo consists of a black crosshair centered over a square. The square is divided into four quadrants: top-left is yellow, top-right is white, bottom-left is red, and bottom-right is blue.

BYU Soar Research Group

- 1 faculty, 6-10 students/year
- Weekly meetings
 - Literature, development, planning
- Goals:
 - Expand Soar knowledge and explore possible uses on-campus
 - Provide and support an NL capability to the Soar research community
 - Toolkits, resources, knowledge repositories
 - Carry out research into the cognitive modeling of linguistic performance



NL-Soar (overview)

- Soar for natural-language modeling
- Comprehension (NLC)
 - Parsing sentences into syntax/semantics
- Discourse (NLD)
 - Discourse moves, turns, participants
- Generation (NLG)
 - Tactical sentence generation
- Various language-related applications
 - Parsing, acquisition, translation, task integration



What we've accomplished

- Several students have learned to use NL-Soar (and some even write sp's)
- Substantial improvements to coverage
 - Still a lot remains, though...
- Testing (including the patience of those who want to use it...)
- Bug tracking via in-house version of SourceForge
- LG-Soar (Data Extraction RG)



What we've accomplished (2)

- Updated website (biblio, mini-tutorials)
- Two-agent conversation over sockets
 - Involves NLC, NLD, NLG
- Workable solution(s) to massive ambiguity problems due to WordNet
- Tutorial at this workshop



The new release (2003)

- Soar 7.3, Tcl 8.x, WordNet 1.6
- Some new documentation
- Hasn't gone out yet...
 - What's the best vehicle?
- Works pretty well, though semantics coverage needs to be increased



What we'll be working on

- Increasing linguistic coverage
 - Including an interface to morphology engine
- Releasing and documenting new release
- Chunking/learning (including WSD)
- WordNet 1.6 → 1.7.1
- Soar 8.x (works already, but...)
- More interesting conversational discourse
- Other languages (French)
- Integration with other NLP tools
 - Other lexicons (e.g. CELEX)
 - Other grapher (CLIG)



Coals

- Complete linguistics coverage is hard
- It'll be a difficult transition from in-house
- The NLP pendulum is not currently favorable
- We're a strictly volunteer team again

Nuggets

- We're doing real language (and gaining associated insights)
- The architecture still works (and we're still trying to listen...)
- Less ad-hoc, more principled development
- Interest is picking up