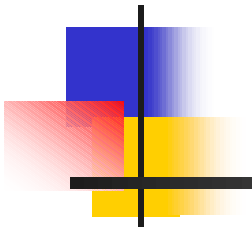
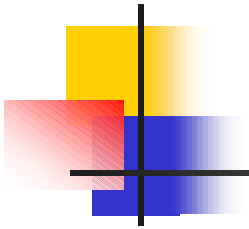


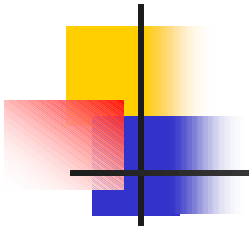
NL-Soar and LG-Soar: ongoing work

Deryle Lonsdale
BYU Linguistics
lonz@byu.edu



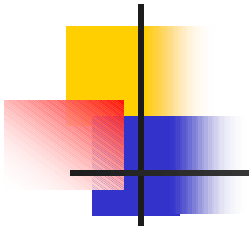
BYU Soar Research Group

- 1 faculty, 5 undergraduates, 3 graduates from across campus
- 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



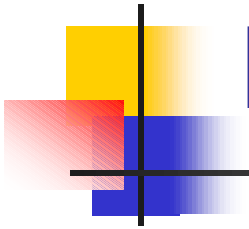
NL-Soar (review)

- Soar for natural-language modeling
- Comprehension (NLC)
  - Syntactic and semantic structures
  - Soar 7.0.4 → 7.3 (reported in 2001)
  - WordNet integration (reported in 2000)
- Discourse (NLD)
  - Discourse moves, turns, participants
- Generation (NLG)
  - Tactical sentence generation
- Various language-related applications
  - Parsing, acquisition, translation, task integration



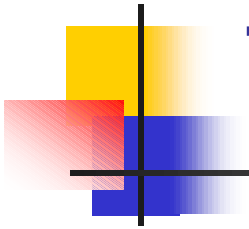
LG-Soar (review)

- Robust, rapid, versatile text processing engine
- Useful for difficult-to-handle input (e.g. IE, IR)
- No NL-Soar architectural constraints
- Shallow syntax: the Link Grammar parser
- Integration via TSI



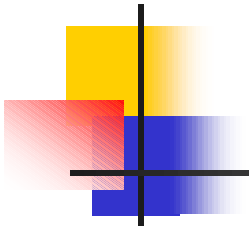
Project support

- Won competitive university-wide bid for project support
  - "Mentoring environments": ugrad research
- Deliverables: public releases of NL-Soar, LG-Soar
- Positions: documenter, website developer, linguistic developers and testers
- Student mentors helping students under faculty supervision
- New machine for lab, research work



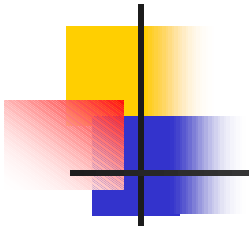
Training

- Tutorials, coloring book are useful
- Scripts, traces to follow
  - Demos, example run-throughs
  - Annotated operator traces
- Code walkthroughs
- Big-picture context
  - Workshop proceedings very helpful
- Even NL-Soar is learnable
  - LG-Soar is good first step for linguistics students



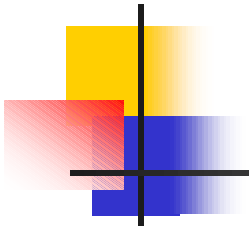
LG-Soar developments

- Refsoar package
  - Interprets LG output, creates predicate-argument structures corresponding to deep semantic properties (hinted at in 2001)
- Displaying syntactic, semantic output
  - Discourse representation structures
  - CLIG grapher (mentioned in 2001)
- Question-answering front-end
  - Positing traces in questions
  - Interfacing with knowledge sources (corpora, www) mostly via Tcl/C
- Port to 8.2 (so far), release this year



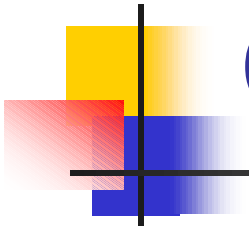
LG-Scorer

- LG-Soar for grading ESL essays
  - Robust parsing, highly ungrammatical input
- LG output, scoring operators based on error types
- Performed about as well as other ESL rating systems, especially for lower-fluency students



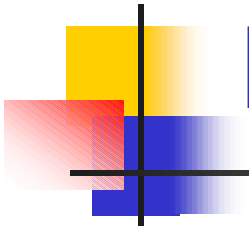
Discourse/dialogue

- NLD running in 7.3
- Work with TrindiKit
  - Possible inspiration, crossover, influence
- WordNet integration
  - Adapt NLD discourse interpretation for WordNet output
- More dialogue plans (beyond TACAIR)



Generation

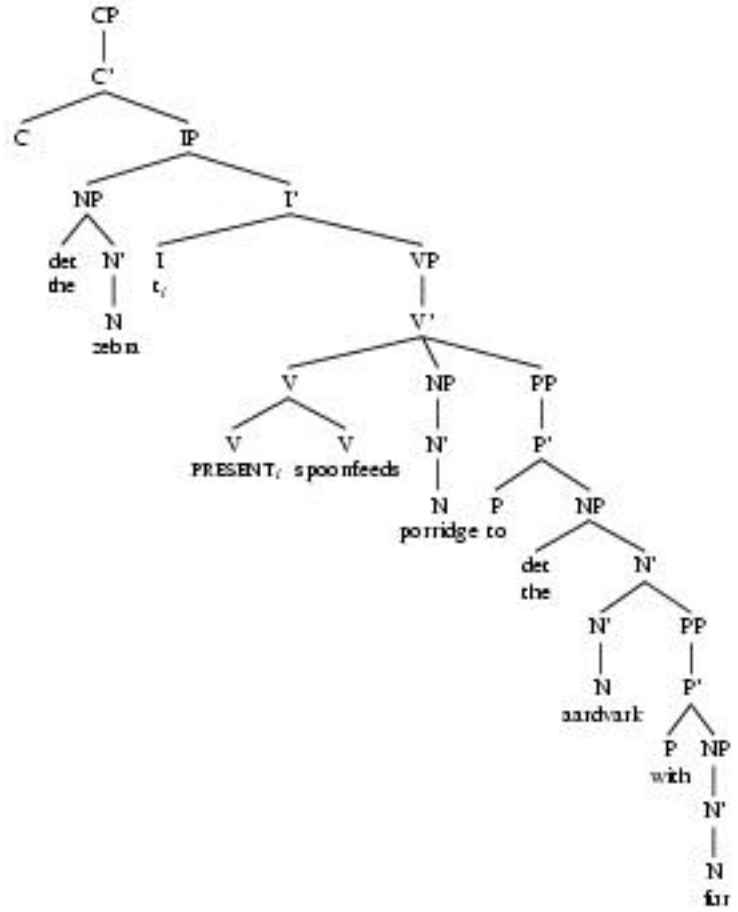
- NLG running in 7.3
- Wider repertoire of lexical selection operators
- WordNet integration
- Serious investigation into chunking behavior

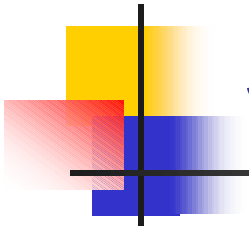


PP-attachment

- How to decide which PP attaches to which constituent
- WordNet provides some clues
  - Verb subcategorization frames
- Now: we can leverage what information WordNet does give us
- So far: syntactic constraints only
  - Ongoing: semantic constraints as well

Sample pp-attachment



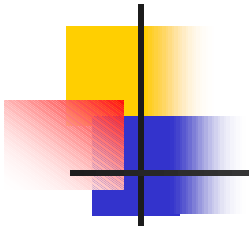


Semantic complexity

- WordNet word-sense complexity is astounding
- Has resulted in severe performance problems in NL-Soar
  - Some (simple!) sentences not possible
- New: user-selectable threshold
- Result: possible to avoid bogging down of system

Sample WordNet ambiguity

head 30	break 63
line 29	make 48
point 24	give 45
cut 19	run 42
case 18	cut 41
base 17	take 41
center 17	carry 38
place 17	get 37
play 17	hold 36
shot 17	draw 33
stock 17	fall 32
field 16	go 30
lead 16	play 29
pass 16	catch 28
break 15	raise 27
charge 15	call 26
form 15	check 26
light 15	cover 26
position 15	charge 25
roll 15	pass 25
slip 15	clear 24



Future

- WordNet 1.6 → 1.7
- Soar 8.x
- NL-Soar on Windows
- LG-Soar Q/A front end
- LG-Soar + WordNet
- Soar + Prolog/Trindi (?)
- Other languages (French)
- NL tools for NL-Soar, LG-Soar
  - Interfaces to other lexicons (e.g. CELEX)