

The XNL-Soar Sandbox

Deryle Lonsdale, Jeremiah McGhee,
Nathan Glenn, and Tory Anderson

Why?

- Language involves interesting cognition
- Language/task integrations compelling
- Lots of theories, microtheories
- Linguists aren't always grounded
- Gives students experience
- Carrying forward prior work
- Little CL/NLP interaction with modeling

From NL-Soar...

- NL-Soar: language modeling system
 - Natural language parsing: syntax
 - Problematic and difficult structures
 - Semantics (lexical conceptual structure)
 - Generation (concepts → text)
 - Discourse (discourse recipes)
 - Language translation (English ↔ French)
 - Task integrations
 - WordNet



- NL-Soar versions
 - Soar 6: mid/late 1990's
 - Soar 7: early 2000's
 - Soar 8: late 2000's
- Extensive chunking
- Extremely deep stacks
- Many different operators
- Soar 8: ran in Soar 7 mode

... to XNL-Soar

- XNL-Soar: upgrade, re-do NL-Soar
- More current syntax (Minimalist Program)
- Soar 9 architecture (finally!)
- Soar 9.3

Migrating WordNet

- Reduction of total # productions by an order of magnitude
- C interface
- Tcl /C interface
- Java interface (JAWS interface)
- Several intermediate versions (1.6-3.0)
- SMem version too

Graphing the parses

- X-windows code
 - Low-level, tightly integrated with agent
- CLIG: computational linguistics interactive grapher
 - Nice primitives, good export capability
 - Not well maintained/disseminated, Tcl
- GraphViz
 - Very versatile
 - Various output formats

Benefits from the rewrite

- Several thousand productions to a few hundred
- Updated syntax
- Soar architectural improvements
- Easier system installation
- Eclipse, SVN
- Students writing code

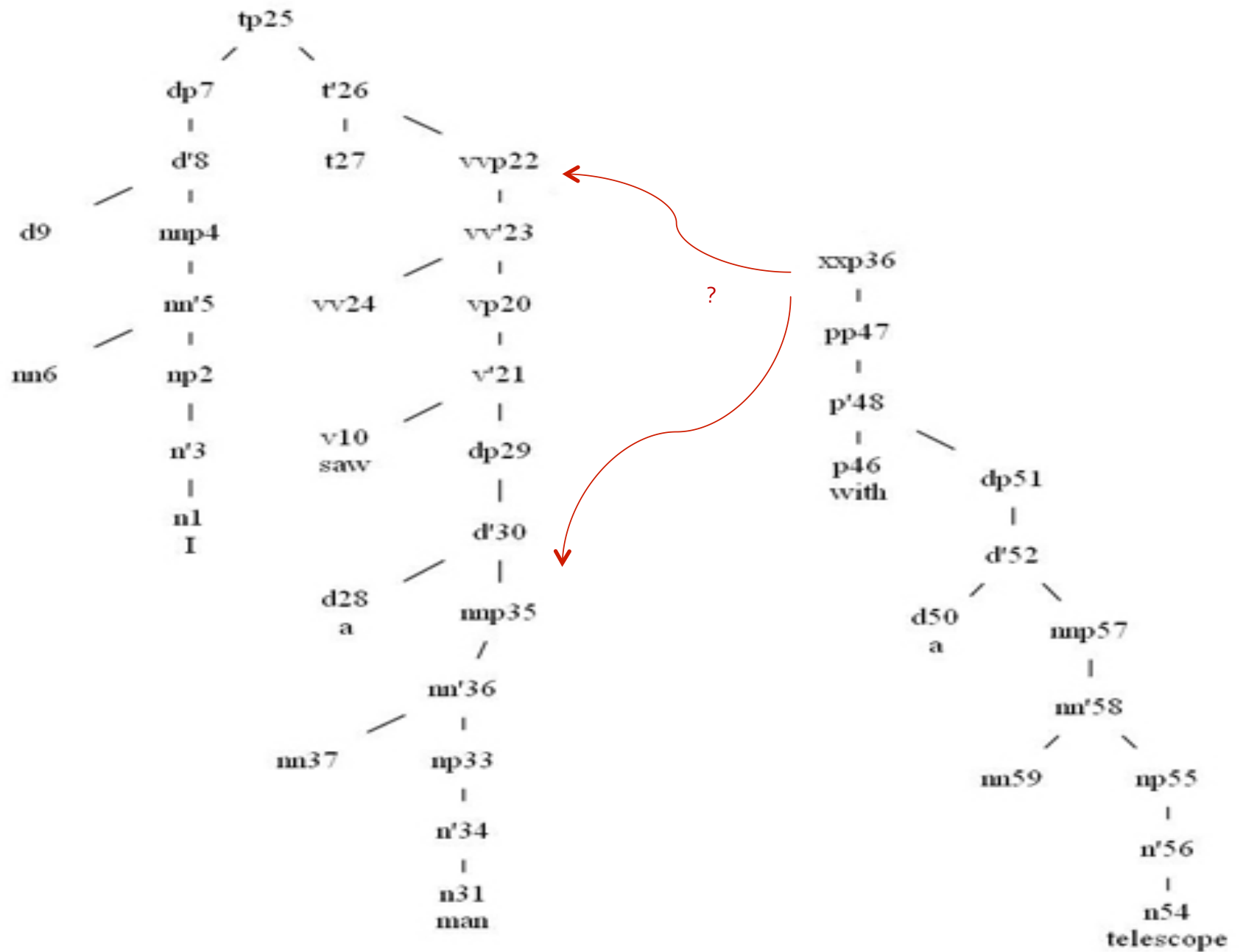
Semantic memory

- Ambiguity is pervasive in natural language; need to:
 - Avoid it when possible (strategies)
 - Detect it when present
 - Disambiguate it when necessary
- Two types of ambiguity
 - PP attachment
 - Present participle / gerund

PP attachment ambiguity

- I saw a man with a moustache. (N attach)
- I saw a man with my own eyes. (V attach)
- I saw a man with a telescope. (N/V attach)

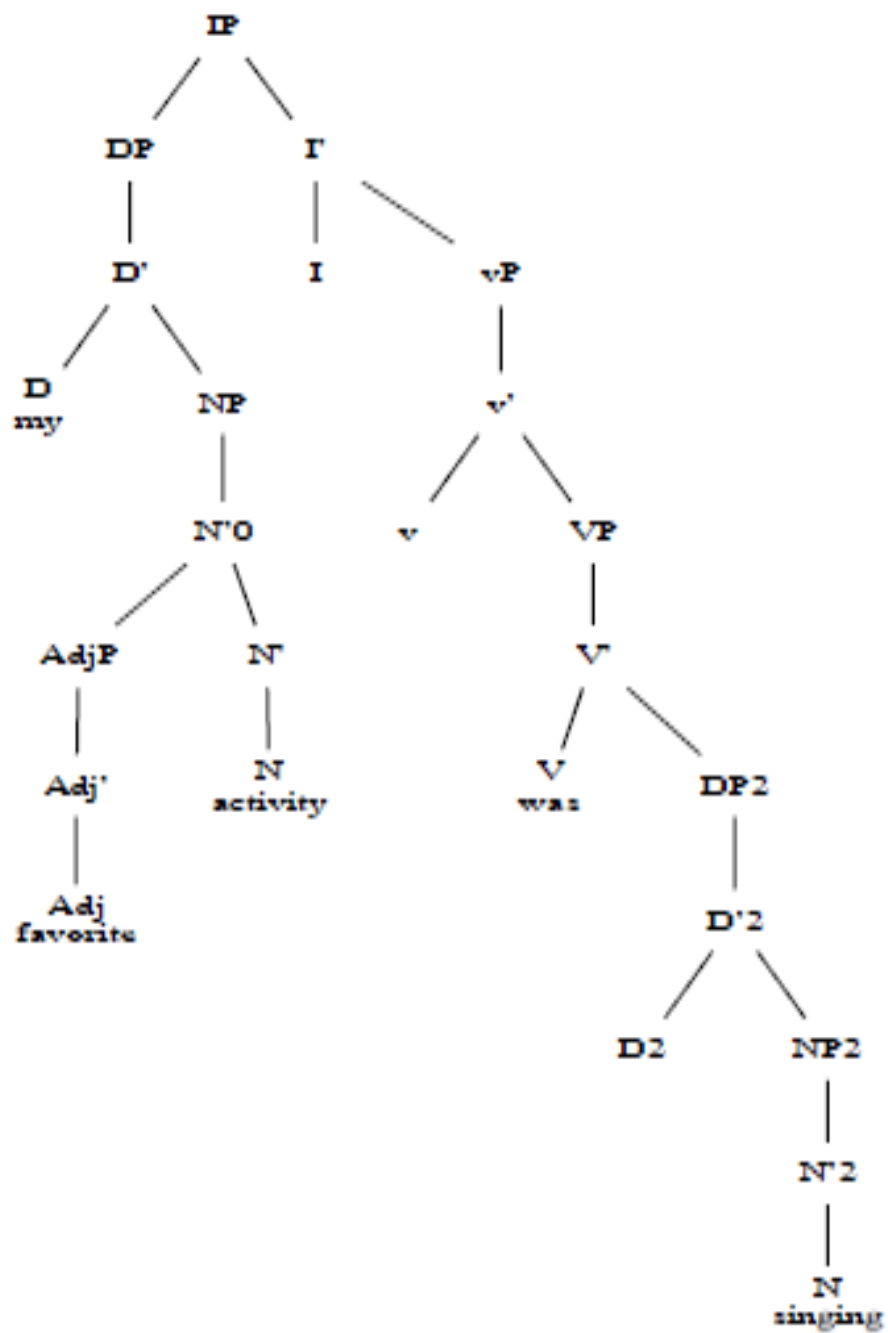
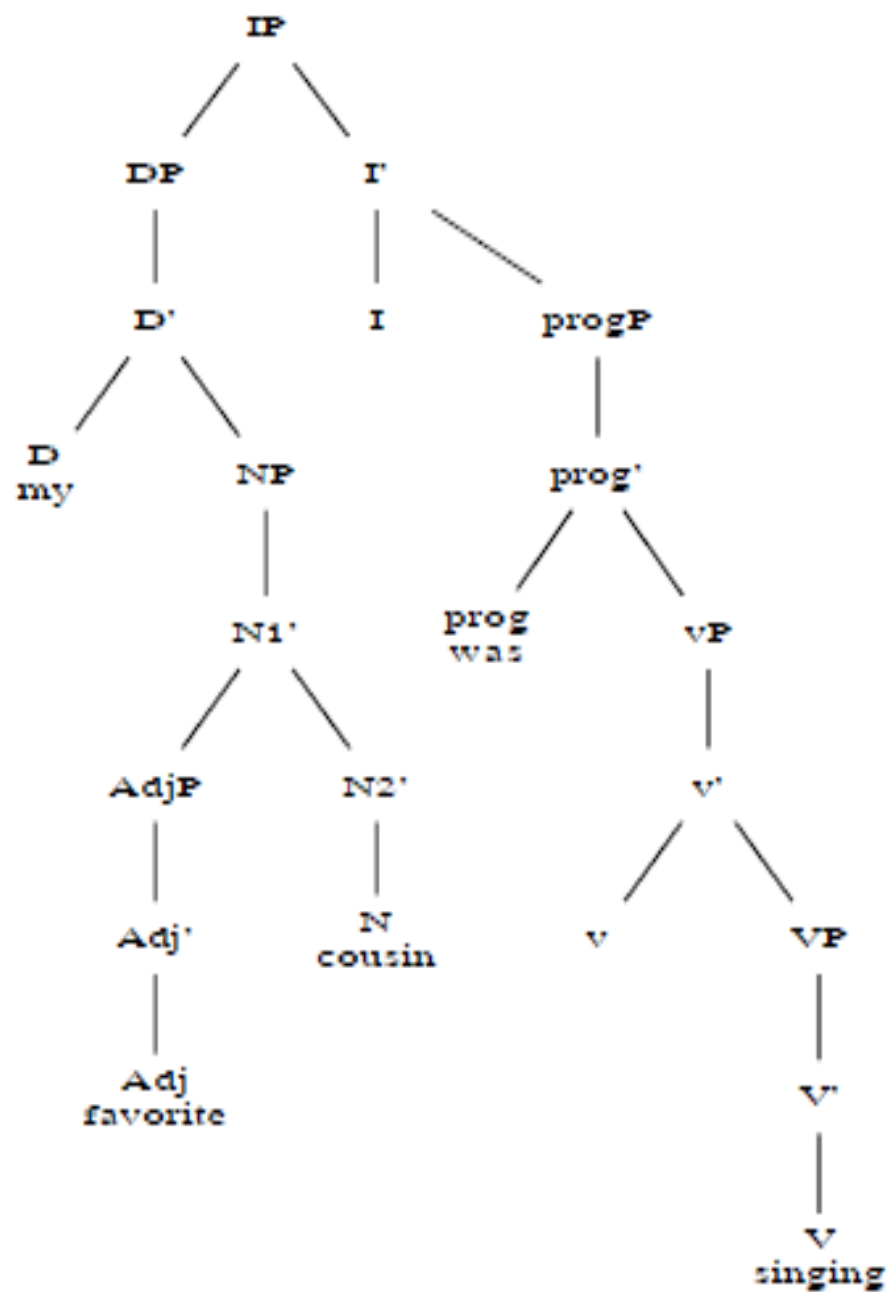
- Training: corpus of WSJ PP attachments
- Test: annotated Brown Corpus sentences
- 6 different strategies
- SMem boosts, generalizes



Participle/gerund ambiguity

- My favorite cousin was singing. (participle)
- My favorite activity was singing. (gerund)
- His business is advertising. (both possible)

- Exemplars mined from Penn treebank
- SMem helps, need more semantics



Underway

- Snips: undo erroneous hypotheses
 - Semantics
 - Semsnips: semantic snips
 - Finish semantic graphing
-
- French
 - Japanese

Why sandbox?

- No learning/chunking
- Walled off from other tasks
- Learning how to code Soar
- Getting used to Soar 9

To do

- Translation
- Discourse
- Generation
- Chunking?

Conclusions

COAL

- Still not public-ready
- Some more linguistics to implement
- No chunking yet
- No activation, decay

NUGGETS

- Up-to-date after years
- More easily deployable
- SMem and EpMem
- More streamlined tools
- Evaluation with real-language use cases
- Hundreds of sentences parsed