

December 9, 2003

Wrap-up/review

Notes on the exam/final projects

- Padelford B202 12/15 (Monday) 4:30-6:20
- Open book, open notes, closed web
- Cumulative – with an emphasis on topics covered since the midterm.
- Projects/papers are due by 4:30 on Monday 12/15, either electronically, to my box, or in person to me.
- Remember to document your projects!
- No more schedule office hours this quarter; appointments welcome.

Notes on the review

- Remind ourselves where we've been.
- Look for connections and generalizations across the different topics we've covered.
- Congratulate ourselves on how much we've learned.

Topics covered before the midterm

- Regular expressions
- Finite state automata
- Finite state transducers
- Morphology & morphological parsing
- CFG
- Syntactic parsing
- Feature structures
- Unification
- Parsing with unification

Topics covered since the midterm (1/2)

- Probabilistic parsing:
 - PCFGs, probabilistic chart parsing, acquiring probabilities, probabilistic lexicalized PCFGs
- Evaluating parsers: precision and recall
- Pronoun resolution:
 - Hard and soft constraints
 - Algorithms: Saliency factors, tree search, centering theory
- Text coherence, coherence resolution

Topics covered since the midterm (2/2)

- Lexical semantics:
 - homonymy v. synonymy, hyponymy
 - WordNet
 - theta roles and selectional restrictions
- Dialogue managers, turn taking, grounding
- Phonetic alphabets, FSTs for phonological rules, two-level phonology
- TTS, pronunciation dictionaries/lexica

By request: Lexicalized PCFGs (1/3)

- PCFGs:
 - $G = (N, \Sigma, P, S, D)$
 - N : A set of non-terminal symbols
 - Σ : A set of terminal symbols (disjoint from N)
 - P : A set of productions (or phrase structure rules)
 $A \rightarrow \beta$ where $A \in N$ and $\beta \in (\Sigma \cup N)^*$
 - S : A designated start symbol, selected from N .
 - D : a function assigning probabilities to each rule in P .

By request: Lexicalized PCFGs (2/3)

- Each node encodes lex item at bottom of its head path.
- Model rule-head and head-head dependencies:

$$P(T) = \prod_{n \in T} p(r(n) \mid n, h(n)) \times p(h(n) \mid n, h(m(n)))$$

- Given that the head is *dumped*, what is the probability of expanding this VP as V NP PP?
- Given that the mother's head is *dumped*, what is the probability that the head of this NP is *sacks*?
- Estimating these probabilities requires smoothing and back-off techniques to deal with sparse data.

By request: Lexicalized PCFGs (3/3)

- How to lexicalized PCFGs improve on plain PCFGs?
- What other kinds of information might we want to include?

Synthesis

- Computational linguistics research results usually fall into one of three categories:
 - Algorithms (often associated with software)
 - Resources
 - Methods of evaluations
- Examples?

Synthesis

- The algorithms we've seen this quarter seem to fall into two classes:
 - Those that stress conceptual elegance/conciseness (and possibly efficiency)
 - Those that seem to consist of a baroque set of constraints/procedures carefully put together
- Examples?
- Why might this be?

Synthesis

- How is computational linguistics research influenced by practical applications?
- How is computational linguistics research influenced by practical considerations?
- Are these influences positive or negative?
- Should they be extended to other areas of linguistics?

Topics covered before the midterm

- Regular expressions
- Finite state automata
- Finite state transducers
- Morphology & morphological parsing
- CFG
- Syntactic parsing
- Feature structures
- Unification
- Parsing with unification

Topics covered since the midterm (1/2)

- Probabilistic parsing:
 - PCFGs, probabilistic chart parsing, acquiring probabilities, probabilistic lexicalized PCFGs
- Evaluating parsers: precision and recall
- Pronoun resolution:
 - Hard and soft constraints
 - Algorithms: Saliency factors, tree search, centering theory
- Text coherence, coherence resolution

Topics covered since the midterm (2/2)

- Lexical semantics:
 - homonymy v. synonymy, hyponymy
 - WordNet
 - theta roles and selectional restrictions
- Dialogue managers, turn taking, grounding
- Phonetic alphabets, FSTs for phonological rules, two-level phonology
- TTS, pronunciation dictionaries/lexica

To learn more...

- Compling lab (“TreeHouse”) meetings – subscribe to compling mailing list on mailman for information about next quarter.
- Spring 04: Grammar Engineering, Statistical NLP seminar
- Colloquia and conferences