

CS4705 MIDTERM, FALL 2002

1. Give an example of each of the following: (15 points)
 - (a) a wh-word
 - (b) a determiner
 - (c) a pronoun
 - (d) a proper noun
 - (e) an auxiliary verb
 - (f) a modal verb
 - (g) a phrasal verb (verb plus particle)
 - (h) a manner adverb
 - (i) a temporal adverb
 - (j) a locative adverb
 - (k) a conjunction
 - (l) a relative clause
 - (m) an NP
 - (n) a PP
 - (o) a VP

2. Identification of terms: Give 1-3 sentence identifications for 5 of the following: (15 pts)
 - (a) Closed class words
 - (b) Unification parsing
 - (c) Treebank
 - (d) (linguistic) Head
 - (e) Long distance dependency
 - (f) Affixation
 - (g) Subcategorization frame
 - (h) Minimum redundancy hypothesis (for lexical representation)

3. Short answer: Answer 2 of the following. (20 pts)
 - (a) What is the difference between mass and count nouns? Give 2 examples of each.
 - (b) What is the difference between derivational and inflectional morphology, e.g. in English. Give 2 examples of each.
 - (c) What is the difference between a deterministic and a non-deterministic finite state automaton?

4. Short exercises: Do 2 of the following exercises. (20 pts)

- (a) Create a finite state transducer that translate the emphatic sheep language 'baa*!' into the quizzical cow language 'moo*?'
- (b) Write a grammar rule and an associated subcategorization frame to enforce subject-verb agreement on person and number.
- (c) What are left-recursive grammar rules? What type of parsers are they a problem for? Give an example of a left-recursive grammar fragment. Turn this into a grammar *without* left recursion.

5. Essay questions: Write a 2-3 paragraph answer to 2 of the following questions (30 pts)

- (a) Discuss at least three sources of ambiguity in natural language. Which do you think are the most difficult to deal with?
- (b) Describe the strengths and weaknesses of Bottom-Up vs. Top-Down parsing. How does the Earley algorithm combine the two? What is the role of the 'dot' in this parsing technique? Left corners?
- (c) How do probabilistic parsing approaches such as the PCYK parser improve over (non-probabilistic) CFG parsers such as the original Early algorithm? What are their drawbacks?

6. Extra credit: (10 pts each)

- (a) Describe the algorithm used in Brill tagging (TBL).
- (b) Describe Kimmo Koskenniemi-style Two Level Morphological parsing.