

Department of Computer Science
CSC 247/447
Spring 2003

Assignment 4

Due: 5pm, Tuesday, Apr. 8

Problem 1 - Long-distance dependencies

Draw a phrase structure tree for each of the following sentences, showing relevant features such as *subcat*, *vform*, and *wh*; you can omit agreement features. For (d) assume a coordination rule

$X \rightarrow X \text{ COORD } X,$

where X is any category and both X 's on the RHS are head daughters. (COORD expands as *and* or *or*.)

- (a) With what shall I fill it?
- (b) Where was the key you mislaid?
- (c) Where she goes, he goes.
(Treat *where she goes* as a topicalized adverbial, and having a structure like that of a *wh*-nominal clause.)
- (d) Alec bought more jewelry than Fran wanted or he could afford

Problem 2: Exercise 3 (p.154), ch.5 of NLU, 1st sentence only. Corrections to Figure 5.12: the AUX1 constituent should have feature COMPFORM *base* rather than SUBCAT *base* (see VP rule page 124); the VP1 constituent should have VFORM *base*; and the empty box for “*see*” should have constituent V1, with features VFORM *base* and SUBCAT *-np*.

Problem 3, grads

- (a) Exercise 4(a) (p.184), ch.6 of NLU. Represent the oracle by drawing the transition graph (as in Fig. 6.5) and labelling *reduce*-states in the graph with the phrase structure rule to be used for the reduction. (There is no need to represent the oracle as a table, as in Fig. 6.6.) In the construction of the graph, treat the *to* in the INF rule just as if it were a lexical category.¹

¹There's a slight inconsistency in the book concerning how words in the input stack are processed: according to Figs. 6.5 and 6.6, the parser never expects to see words, only lexical and phrasal symbols. Fig. 6.7 seems to allow for direct processing of words, but in the illustration starting on p.166, words are tacitly converted to lexical categories when they are shifted from the input stack to the parse stack. Assume that such a conversion is done here.

- (b) Do enough of part (b) of Exercise 4 to show that guesswork would be required in the conversion of words to lexical categories in a shift-reduce parse of the sentences shown, using the oracle in (a).
- (c) Explain how the oracle can be modified to deal with this ambiguity deterministically.

Problem 4

- (a) The *odds* in favor of an event A, $O(A)$, is defined as the ratio of its probability of occurrence to its probability of non-occurrence:

$$O(A) = \frac{P(A)}{P(\neg A)}.$$

Similarly one can define conditional odds:

$$O(A|B) = \frac{P(A|B)}{P(\neg A|B)}.$$

Show from first principles, using only the definitions of conditional probability and odds, that

$$O(A|B) = O(A) \frac{P(B|A)}{P(B|\neg A)}.$$

- (b) For each of the following cases, determine whether the events A and B are conditionally independent, given event C. Justify your answer in each case.
- (i) $P(A|C) = .5, P(B|A, C) = .6, P(\neg B|C) = .4$
- (ii) $P(A|C) = .3, P(B|C) = .5, P(A \wedge B|C) = .15, P(A \vee B|C) = .65$
- (iii) $P(C) = .7, P(A \wedge C) = P(B \wedge C) = .5, P(A \wedge B \wedge C) = .3$

Problem 5: Exercise 3, p.221, ch.7 of NLU, but change the sentence to “*Flowers flower like flowers*”.

Problem 6 - lambda-conversion

Ex. 1, ch. 9 (p. 291) of text

Problem 7 - LFs

Show phrase structure trees annotated with logical forms (LFs) at all non-word nodes for the following sentences. You can omit all syntactic features except verb subcategorization features; but include tense in the logical forms. For (a), give both a compact predicate logic based LF and a thematic role based LF. For the remaining sentences you may use either form. For names, you may use the simple approach of treating them as individual constants here, rather than using the “NAME” operator of the JFA text.

In (b) and (d), use coordination rule

$X \rightarrow X \text{ COORD } X; \langle \text{COORD}' X' X' \rangle,$

producing an unscoped coordinator and its operands as an LF. For *No-one* in (c) assume a quantifier **No**, restricted by **person**. In (d) assume that the complement of the copula (*is*) must have a **+pred** feature, and this can appear on both APs and NPs. NPs normally are **-pred**, and we have a special rule for forming an NP[**+pred**]:

$\text{NP}[\text{+pred}] \rightarrow \text{Det}[\text{a}\{\text{n}\}] \text{N}; \text{N}';$

essentially this ignores the determiner, using the noun to supply the meaning of the entire NP (so that the NP is interpreted as a predicate).

- (a) Mary found the missing file
- (b) John insisted that he saw Elvis or Marilyn
- (c) No-one must leave.
- (d) John is clever and a charlatan
- (e) Sagan talked on Mars (*supply two analyses*)
- (f) What has Mary read?

Problem 8: Scoping Quantifiers and Connectives

Produce all readings of 2(b) and (c) resulting from “lifting” the unscoped operators to sentential positions.

Problem 9: Thematic Roles

Ex. 6, ch. 8 (p. 259) of text

Note to grads: There’s one month left in the semester. Please be sure to move ahead on your project/essay. You should have a clear idea of what you’ll do at this point, and should have some key references at hand, and an outline of the contents. Recall that the project could tie in with the class project, or could be separate. Some possible topics are: robust parsing; syntactic or lexical disambiguation; learning subcategorization patterns from corpora; learning grammars from corpora; learning lexical meanings from corpora; learning taxonomies from lexicons; kinds of inference (e.g., abductive) needed for language understanding; semantics of beliefs, intentions, etc.; determining referents of referring expressions; interpreting noun-noun compounds; interpreting tense and aspect; machine translation (or some subtopic thereof); vagueness; metaphor; scoping quantifiers and other operators; neural net approaches to language processing (e.g., parsing or disambiguation); aspects of language generation; morphological analysis; speech act or plan recognition; etc. (The book, including chapters not covered yet, provides starting points for many other topics.)