

Department of Computer Science
CSC 247/447 and BCS 235/535
Fall 2006

Assignment 2 (non-Lisp portion)

Due: 3:15pm (hand in at TA's office), Tuesday, Oct. 10

Problem 1 - Chart Parsing

Consider the following CFG and lexical entries:

S → NP VP	The: Det
NP → Pron	man: N
NP → Det N	saw: V, N
VP → V	it: Pron
VP → V NP	

Show the additions to the set of active arcs and to the chart that would be produced by the parser of section 3.4 in the text (a bottom-up, left-corner chart parser) operating on the sentence

₁ *The* ₂ *man* ₃ *saw* ₄ *it* ₅

Draw active arcs beneath the sentence and chart entries above it, in the approximate order in which they are generated. (*Note:* As indicated in class, there is some leeway concerning the order in which the actions corresponding to each completed constituent are carried out; it's not important that you do this in exactly the order specified in the book or in class. Also it's worth noting that the way left-corner instantiation of a rule is done in the book is to add an *empty* active arc that is still "expecting" all of its constituents, including the first, and this is then extended in the arc extension step along with other arcs. You don't need to show these empty arcs.)

Problem 2 - Auxiliary Verb Sequencing

- (a) Using the grammar rules and lexical assumptions given in class for adding auxiliary verbs to a verb phrase, give phrase structure trees for the VPs '*may have left*', '*should be sleeping*', and '*is being given away*'. (*Note:* Be careful here, making especially sure to get the VFORM and SUBCAT features right – this is the key to auxiliary verb ordering. Just showing branches and nodes of the right type in the right order is not nearly enough.)
- (b) Explain why the following are not possible VPs, in terms of the possible SUBCAT and VFORM values of the verbs involved: *'*may has left*', *'*may should leave*', *'*has may leave*', *'*is having left*', and *'*is being leaving*'.

Note: When diagnosing a problem in something like "has may leave", you can't just say something like "The modal auxiliary should come first". You're not being asked for a correction (it's a given fact that it's wrong), or for your intuitions (we

all have those, when it comes to language). You need to say *why it won't be accepted by the grammar that was given in class, i.e., why an attempt to parse it with those rules will fail*. For that you have to consider the possible lexical categories of the words involved, and the SUBCAT features.

Problem 3 - Parsing with Features

Exercise 3, ch. 4 of NLU (p. 119). Use Allen's parser – you can get it from Allen's web page at

<http://www.cs.rochester.edu/u/www/u/james/>

which points you to a basic parser intended to support the text. Use the parser, rather than taking the option of drawing charts by hand. Familiarity with Allen's parser will be needed for later assignments as well.

In writing down PSGs, if you rely on feature principles/conventions, indicate briefly where they apply and what effect they have. (But given the need to implement the rules, it probably makes more sense to handle feature propagation/agreement explicitly via variables.)

Be sure not to neglect the part that involves showing that you can handle a range of similar cases.

Problem 4 - Grammars with Features

- (a) Exercise 5, ch. 4 of NLU (p. 119). Use a CASE feature with values subj and obj.
- (b) (*Grads only*) Exercise 4, ch. 4 of NLU (p. 119). Just to get you started, here are some possible rules:

```
(NP HOUR ?h MINUTES 0) -> (A hour ?h) o'clock
(A HOUR 1) -> one
(NP HOUR ?h MINUTES ?m) -> (NP MINUTES ?m) (PP AFTER + HOUR ?h)
```

The +AFTER feature would be carried by prepositions *after* and *past*. Note that for something like *ten to two* or *ten minutes to two* we'd like to get HOUR = 1, MINUTES = 50 ...

Problem 5 - Yes-No Question Syntax

Draw syntactic trees (using concise notation, i.e., with [...] -features) for the following questions

- (a) Will John have arrived?
- (b) Has Mary been brought to the airport?

Once again, be sure to include SUBCAT and VFORM features. It also makes sense to include agreement features, at least where these play a role in agreement. Note that the "inv" feature is crucial for questions.

Problem 6 - RTNs – *Grads only*

Exercise 10, ch. 3 of NLU (p. 78), *but with the added CFG rule*

$NP \rightarrow \text{ART } N.$

(Without this rule the CFG generates nothing at all, because then every NP has a postmodifying PP, which contains an NP, which contains a postmodifying VP, etc.) Remember this rule in doing part (c)!

For part (b), also write a “generalized CFG” equivalent to the RTN, i.e., one that can use the Kleene star “*” and optional constituents. (More generally we could – but won’t – allow regular expressions on the RHS of grammar rules.) Make it as close as possible to the RTN shown.

Note: Past experience indicates that students tend to forget any additions/modifications to questions taken from the book. To avoid losing marks inadvertently in this way, *pencil such changes briefly into the margin, as reminders to yourself! Also re-read the problems as posed here before assuming you’re done with them.*