Computational Linguistics
LING 409 · McGowan
Spring 2013
Final Exam (50 pts)

IMPORTANT: This exam contains 90 points worth of questions. **Please answer only 50 points** worth of questions. Which questions you choose to answer are entirely up to you.

Assigned: Wed, Apr 24, 2013 @ 9:00am
Due: by Wed, May 1, 2013 @ 12:00pm

This exam should not take you more than 3 hours, but you have the full exam week to complete it. You may consult your textbook, your notes, or anything I have uploaded or linked from our owlspace site. You may not, alas, collaborate.

Question 1 (10 points)

Consider the following regular grammar for NP:

(DT) (CD) JJ* (VBG) NN* NN

(Where NN is either NNS or NNP).
Write an equivalent CFG (no lexicon needed, just the phrase structure rules).

Question 2 (10 points)

Write a simple context-free grammar that can produce the following sentences:

1. Joe ate dinner.

2. Did Maureen eat dinner?

3. When did Joe eat dinner?

4. Eat Dinner!

Question 3 (10 points)

Give an example of each of the following instances of morphological derivation in English: verb → noun; adjective → noun; adjective → adverb; adjective → verb; noun → verb.

Question 4 (10 points)

Give an example for each of the following verb subcategorization frames.

1. NP
2. NP NP
3. Ø
4. V P_{to}
5. S

Question 5 (10 points)

Which of the columns (a) - (e) in Table 1 contains a grammar that correctly generates the following sentences?

-- The man ate fish.

-- The man ate fish with a fork.

while it DOES NOT generate the following sentences:

-- The man ate.

-- The man ate with a fork.

(a)	(b)	(c)	(d)	(e)
S → NP VP NP → ART N VP → V NP NP → N VP → VP PP PP → PREP NP ART → the N → man V → ate N → fish PREP → with ART → a N → fork	S → NP VP PP NP → ART N NP → N VP → V NP PP → PREP NP ART → the N → man V → ate N → fish PREP → with ART → a N → fork	S → NP VP NP NP → ART N NP → NP PP VP → V NP NP → N PP → PREP NP ART → the N → man V → ate N → fish PREP → with ART → a N → fork	S → NP VP NP → ART N VP → V VP → V NP NP → N VP → VP PP PP → PREP NP ART → the N → man V → ate N → fish PREP → with ART → a N → fork	S → NP VP PP NP → ART N NP → N VP → VP NP VP → V PP → PREP NP ART → the N → man V → ate N → fish PREP → with ART → a N → fork

Table 1: Grammars for Question 5.

Question 6 (20 points)

Please use the Centauri/Arcturan bitext on the left of this table to translate the three sentences that follow. Notice that you also have some monolingual Centauri text (right).

C1 ok-voon ororok sprok. A1 at-voon bichat dat.	ok-drubel anok ghirok farok . wiwok rarok nok zerok ghirok enemok . ok-drubel ziplok stok vok erok enemok kantok ok-yurp zinok jok yorok clok . lalok clok izok vok ok-drubel . ok-voon ororok sprok . ok-drubel ok-voon anok plok sprok . erok sprok izok hihok ghirok . ok-voon anok drok brok jok . wiwok farok izok stok . lalok sprok izok jok stok . lalok brok anok plok nok . lalok farok ororok lalok sprok izok enemok . wiwok nok izok kantok ok-yurp . lalok mok nok yorok ghirok clok . lalok nok crrrok hihok yorok zanzanok . lalok rarok nok izok hihok mok .
C2 ok-drubel ok-voon anok plok sprok. A2 at-drubel at-voon pippat rrat dat.	
C3 erok sprok izok hihok ghirok. A3 totat dat arrat vat hilat.	
C4 ok-voon anok drok brok jok. A4 at-voon krat pippat sat lat.	
C5 wiwok farok izok stok. A5 totat jjat quat cat.	
C6 lalok sprok izok jok stok. A6 wat dat krat quat cat.	
C7 lalok farok ororok lalok sprok izok enemok. A7 wat jjat bichat wat dat vat eneat.	
C8 lalok brok anok plok nok. A8 iat lat pippat rrat nnat.	
C9 wiwok nok izok kantok ok-yurp. A9 totat nnat quat sloat at-yurp.	
C10 lalok mok nok yorok ghriok clok. A10 wat nnat gat mat bat hilat.	
C11 lalok nok crrrok hihok yorok zanzanok. A11 wat nnat arrat mat zanzanat.	
C12 lalok rarok nok izok hihok mok. A12 wat nnat forat arrat vat gat.	

Table 2: Centauri & Arcturan bitext

Please translate the following sentences from Arcturan to Centauri (and try to get the Centauri order correct using the monolingual text):

(a) iat lat pippat eneat hilat oloat at-yurp .

(b) totat nnat forat arrat mat bat .

(c) wat dat quat cat uskrat at-drubel .

Question 7 (20 points)

Consider the five sentences shown in Figure 2. Think of them as your training data to build a probabilistic context-free grammar (PCFG).

Delta flight 411 leaves Toronto for Atlanta at 6 PM
This flight serves a light meal
When does this flight leave
Northwest flight 29 leaves Atlanta for Detroit at 10 AM
When does the Delta flight arrive

Figure 2: Training sentences for Question 6.

The parse trees for these sentences are shown in Figure 3

```
(S (NP (NNP Delta)
       (NN flight)
       (CD 411))
   (VP (VBZ leaves)
       (NP (NNP Toronto))
       (PP (IN for)
           (NP (NNP Atlanta)))
       (PP (IN at)
           (NP (QP (CD 6)
                   (RB PM))))))

(S (NP (DT This)
       (NN flight))
   (VP (VBZ serves)
       (NP (DT a)
           (JJ light)
           (NN meal))))

(S (WH (WRB When))
   (S (AUX does)
      (NP (DT this)
          (NN flight))
      (VP (VB leave))))

(S (NP (NNP Northwest)
       (NN flight)
       (CD 29))
   (VP (VBZ leaves)
       (NP (NP (NNP Atlanta))
           (PP (IN for)
               (NP (NNP Detroit))))
       (PP (IN at)
           (NP (CD 10)
               (RB AM)))))

(S (WH (WRB When))
   (S (AUX does)
      (NP (DT the)
          (NNP Delta)
          (NN flight))
      (VP (VB arrive))))

(S (NP (DT That)
       (NN flight))
   (VP (MD may)
       (VP (VB serve)
           (NP (DT a)
               (NN meal)))))
```

Figure 3: Parse trees for the sentences in Figure 2.

(a) Build a PCFG using the training data. For each non-lexical rule (e.g. S → NP VP), indicate its probability.

(b) Build a probabilistic lexicon for the PCFG. Give the probability of each lexical rule (e.g., NN → flight).

(c) Smoothing (reserving probability mass for unobserved rules is very important when building PCFGs. Redo parts (a) and (b) above using 10% of the probability mass for each non-terminal or lexical category to cover unknown words. Example: if A → B C has a probability of .6 and A → B D has a probability of .4, you need to create a new rule A → α with a probability of .1 and readjust downward the probabilities of the other two rules that have A on the left hand side.

(d) For each of the following two sentences "When does Northwest flight 77 leave for Milwaukee" and "Does this flight leave for Milwaukee", draw **one** parse tree according to the (smoothed) grammar in part (c). If you are getting any zero probabilities, return to part (c) and fix your grammar appropriately. What are the final probabilities for each of these two sentences? For this question, you don't need to find all possible parses of a given sentence. One parse per sentence will be enough.