

COMP 6721 Applied Artificial Intelligence (Winter 2021)

Worksheet #9: Introduction to Natural Language Processing

Sentence Probability. Given the following *bigram* probabilities, compute the probability for the sentence:

$P(\text{on} \text{eat}) = .16$	$P(\text{want} \text{I}) = .32$	$P(\text{eat} \text{to}) = .26$
$P(\text{some} \text{eat}) = .06$	$P(\text{would} \text{I}) = .29$	$P(\text{have} \text{to}) = .14$
$P(\text{British} \text{eat}) = .001$	$P(\text{don't} \text{I}) = .08$	$P(\text{spend} \text{to}) = .09$
...
$P(\text{I} \text{<s>}) = .25$	$P(\text{to} \text{want}) = .65$	$P(\text{food} \text{British}) = .6$
$P(\text{I'd} \text{<s>}) = .06$	$P(\text{a} \text{want}) = .5$	$P(\text{restaurant} \text{British}) = .15$
...

$P(\text{I want to eat British food})$

=
 =
 =

Parsing. Given the following lexicon and context-free grammar, create a *parse tree* for the sentence, “I prefer a direct flight to Chicago.”:

Lexicon:

```

N --> flights | trip | breeze | morning    // noun
V --> is | prefer | like                    // verb
Adj --> direct | cheapest | first           // adjective
Pro --> me | I | you | it                   // pronoun
PN --> Chicago | United | Los Angeles      // proper noun
D --> the | a | this                        // determiner
Prep --> from | to | in                     // preposition
Conj --> and | or | but                     // conjunction

```

Grammar:

```

S --> NP VP                                // I + prefer United
NP --> Pro | PN | D N | D Adj N            // I, Chicago, the morning
VP --> V | V NP | V NP PP                  // is, prefer + United,
PP --> Prep NP                               // to Chicago, to I ??

```

Step 1: Assign a *part-of-speech* (POS) tag to each word:

I	prefer	a	direct	flight	to	Chigaco.

Step 2: Draw a legal *parse tree*:

Word Sense Disambiguation. Using the following probabilities you obtained from a training corpus:

```

□  $P(\text{the}|\text{BANK1}) = (5+.5) / (30+.5V)$ 
□  $P(\text{world}|\text{BANK1}) = (1+.5) / 55$ 
□  $P(\text{and}|\text{BANK1}) = (1+.5) / 55$ 
□ ...
□  $P(\text{off}|\text{BANK1}) = (0+.5) / 55$ 
□  $P(\text{Potomac}|\text{BANK1}) = (0+.5) / 55$ 
□  $P(\text{BANK1}) = 5/7$ 

 $P(\text{the}|\text{BANK2}) = (3+.5) / (12 + .5V)$ 
 $P(\text{world}|\text{BANK2}) = (0+.5) / 37$ 
 $P(\text{and}|\text{BANK2}) = (0+.5) / 37$ 
 $P(\text{off}|\text{BANK2}) = (1+.5) / 37$ 
 $P(\text{Potomac}|\text{BANK2}) = (1+.5) / 37$ 
 $P(\text{BANK2}) = 2/7$ 

```

Using 0.5 smoothing as shown above, using a context window of ± 3 , find the correct sense for *bank* in the sentence, “I like the Potomac bank.”:

- Score(BANK1) =
- Score(BANK2) =

(Words not shown above have an *unsmoothed* probability of 0.)