COMP 6721 Applied Artificial Intelligence (Winter 2022)

Worksheet #9: Introduction to Natural Language Processing

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

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

$P(I\ want\ to\ eat\ British\ food)$

$=$ .....

$=$ .....

$=$ .....

Part-of-Speech Tagging. Given the following lexicon, assign a *part-of-speech* (POS) tag to each word for the sentence below:

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
```

I	prefer	a	direct	flight	to	Chigaco.

Parsing. Now, given the following context-free grammar:

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 ??
```

create a *parse tree* for the sentence, *"I prefer a direct flight to Chicago."* using the POS tags you assigned above:

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

- P(*the*|BANK1) = (5+.5) / (30+.5V)
- P(*world*|BANK1) = (1+.5) / 55
- P(*and*|BANK1) = (1+.5) / 55
- …
- P(*off*|BANK1) = (0+.5) / 55
- P(*Potomac*|BANK1) = (0+.5) / 55
- P(BANK1) = 5/7

- P(*the*|BANK2) = (3+.5) / (12 + .5V)
- P(*world*|BANK2) = (0+.5) / 37
- P(*and*|BANK2) = (0+.5) / 37
- …
- P(*off*|BANK2) = (1+.5) / 37
- P(*Potomac*|BANK2) = (1+.5) / 37
- P(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."*:

1. Score(BANK1) = ...
2. Score(BANK2) = ...

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