CPSC 477/577 NATURAL LANGUAGE PROCESSING
FINAL EXAM
SPRING 2021

Part 1: Short Answer Questions with Textual Answers

7 items * 3 points = 21 points

Q1. What property (properties) of Tree Adjoining Grammars makes them (strictly) more powerful than Context Free Grammars?

Q2. What WordNet node (synset) is likely to be the Lowest Common Subsumer (LCS) for *eagle* and *goose* (in their most frequent senses as nouns).

Q3. Consider the sentence *Fruit flies like a banana*.

a. This sentence is ambiguous. Paraphrase the two possible interpretations
b. Consider the following CCG parses of the sentence, which have some of their labels removed (indicated with [ ]). For each parse, fill in the missing labels: [1.], [2.], … [6.]
c. Note: a “fruit fly” is a tiny insect.

Parse 1:

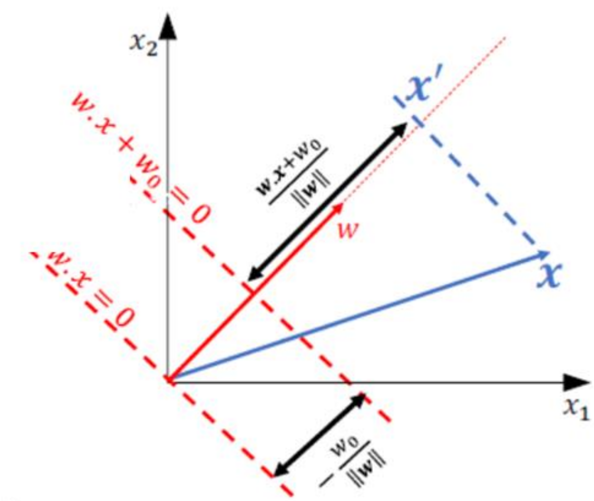
<table>
<tr><td>Fruit</td><td>flies</td><td>like</td><td>a</td><td>banana</td></tr>
<tr><td>N</td><td>S\NP</td><td>[1.]</td><td>NP/N</td><td>N</td></tr>
<tr><td>[2.]</td><td></td><td></td><td colspan="2">NP</td></tr>
<tr><td></td><td></td><td colspan="3">[3.]</td></tr>
<tr><td></td><td colspan="4">S\NP</td></tr>
<tr><td colspan="5">S</td></tr>
</table>

Parse 2:

Fruit	flies	like	a	banana
N	N\N	**[4.]**	NP/N	N
N			NP	
[5.]		**[6.]**		
		S		

c. For each parse, indicate which of the two meanings the given parse corresponds to.

Q4. Explain the meaning of x_1, x_2, x, x', w, w_0, and ||w|| in the image below.



Q5. Fill in the missing semantic types. Hint: one of the types is <e, t>

```
S   -> NP VP    {VP.Sem(NP.Sem)}      t
VP  -> V NP     {V.Sem(NP.Sem)}       ......
NP  -> N        {N.Sem}               ......
V   -> likes    {λ x,y likes(x,y)     ......
N   -> Javier   {Javier}              e
N   -> pizza    {pizza}               ......
```

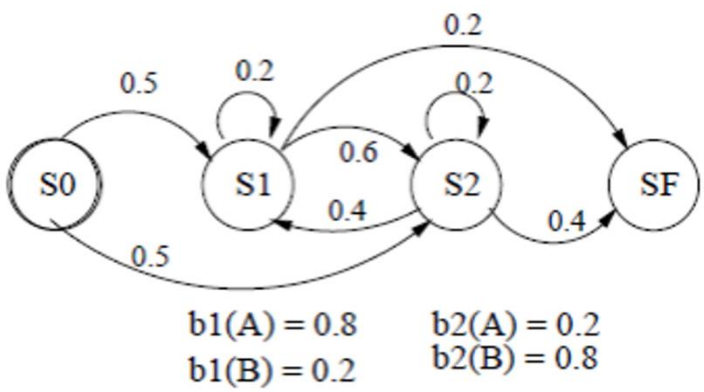
Q6. Given the word embeddings (V_w) for the words w in {Paris, London, France, English, French, Euro, UK, etc.}, what is the value of ($V_{London} – V_{Paris} + V_{Euro}$)?

Q7. Which of the following words start(s) with a voiced consonant?
tear, cat, three, gate, part, fuse, deer

Part 2: Short Answer Questions with Numerical Answers
6 items * 4 points = 24 points

Q1. Using a chart, compute the edit distance between APPLE and PIE. Assume equal costs for the three basic edit operations.

Q2. Consider the HMM below where the transition probabilities are shown in the graph and the observation probabilities (where V = {A,B}) are in the tables below each state.



Use the forward algorithm (or some other method) to compute the probability of generating the short output string "A B", starting at node S0.

Q3. Write down a PCFG such that:

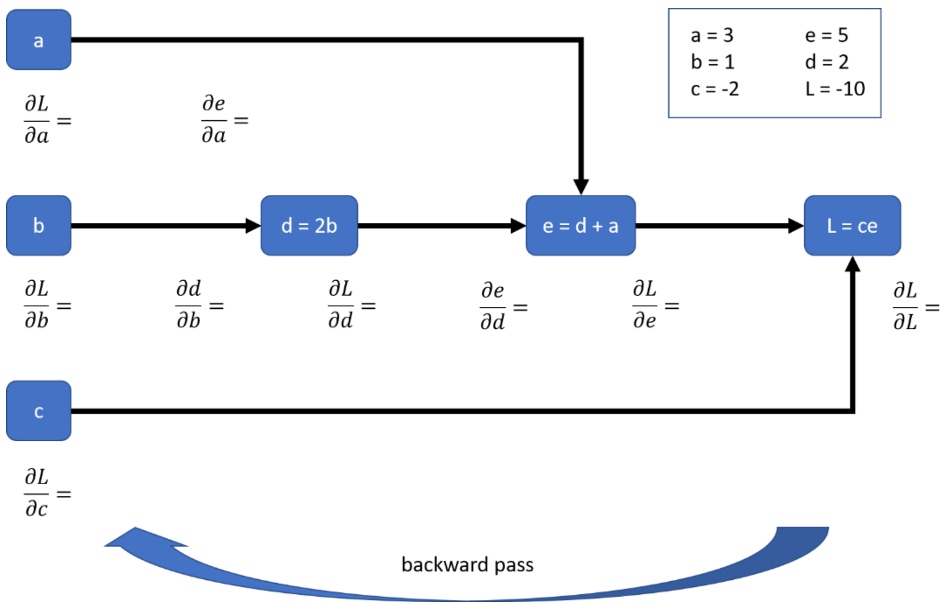
1. Any sentence consisting of the word "the" n times in a row, where $n \geq 1$, has probability $0.4 \times 0.6^{n-1}$
2. Any other sentence has probability 0.

Q4.

<table>
<tr><td colspan="2" rowspan="2">P(A|B)</td><td colspan="2">A = "dog"</td></tr>
<tr><td>Yes</td><td>No</td></tr>
<tr><td rowspan="2">B = NN</td><td>Yes</td><td>0.1</td><td>0.9</td></tr>
<tr><td>No</td><td>0.01</td><td>0.99</td></tr>
</table>

Suppose *p(NN)* = 0.4. Use the Bayes rule to write the expression for *p (NN| "dog")*. Write it first as a function of *p(dog)* and then, also give the actual probability value.

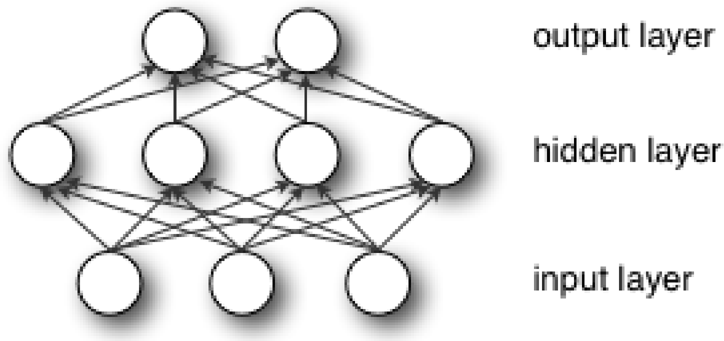
Q5. Compute all the missing partial derivatives.



Q6. A corpus contains 1,000,000 tokens, including 6,000 instances of the word "the". The next four most frequent words are "of", "and", "to", and "a". What are the expected counts for each of these four words in the corpus?

Part 3: Neural network sentiment classification
5 points

Q1. We use a 3-layer neural network for sentiment classification of words. The architecture of the neural network is shown in the picture. Words are represented by 3-dimensional embeddings as inputs, and the network outputs a probability distribution over the positive and negative classes as a 2-dimensional vector.



The hidden layer and the output layer use Rectified Linear Units (ReLU) as the activation function.

$$ReLU(t) = \begin{cases} t, if\ t \geq 0 \\ 0, otherwise \end{cases}$$

Suppose your parameters are

$$W_1 = \begin{bmatrix} -1 & 2 & 0 \\ 3 & 4 & 2 \\ 1 & -2 & -1 \\ 0 & 1 & 0 \end{bmatrix}$$

$$b_1 = \begin{bmatrix} 3 \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

$$W_2 = \begin{bmatrix} 2 & -1/3 & 0 & 2 \\ 1 & 0 & 1 & 4 \end{bmatrix}$$

$$b_2 = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

1) Suppose the input word has an embedding of:

$$x = \begin{bmatrix} 3 \\ -1 \\ 2 \end{bmatrix}$$

Calculate the activation of the hidden layer.

2) Following (1), what is the probability distribution after we apply softmax over the output layer activations?

We use cross-entropy loss as our objective function defined as:

$$L = -\frac{1}{N}\sum_{n=1}^{N}[y_n \log p_n + (1 - y_n)\log(1 - p_n)]$$

where N is the batch size, and we label positive words with y = 1, and negative words with y = 0.

3) Now suppose the current mini-batch contains four words: {good, bad, excellent, poor}. The network outputs a probability distribution as the following:

	good	bad	excellent	poor
P (y=1)	0.7	0.3	0.6	0.2

What is the value of the objective function for the current mini-batch?

Part 4: Multiple Choice Questions
25 items * 2 points = 50 points

Q1. The following sentences show examples of what linguistic phenomenon?

I had a coffee this morning (meaning "I had one cup of coffee")
I tried two wines last night (meaning "I had two types of wine")
I had fish for dinner (meaning "I had some fish", not "I had a fish")

a. type coercion
b. selectional restriction
c. non-projectivity
d. backoff

Q2. Consider the tweet below. It is funny, partially because it involves an ambiguity. What type of ambiguity is it?

a. part of speech
b. referential
c. morphological
d. syntactic

Q3. What is the formula for tanh(x) as a function of the sigmoid function s(x).

a. $\tanh(x)=2s(2x)-1$
b. $\tanh(x)=s(2x)$
c. $\tanh(x)=s(x)$
d. $\tanh(x)=2s(x)-1$
e. $\tanh(x)=2s(2x)$

Q4. What is this formula used for?

$$d = \arg\max_i \frac{(x_b - x_a + x_c)^T x_i}{||x_b - x_a + x_c||}$$

a. continuous bag of words (CBOW)
b. wordnet-based semantic similarity
c. dimensionality reduction
d. word analogy computation
e. none of the above

Q5. Which of these languages is closest to English from a historical, evolutionary perspective?

a. Turkish
b. Norwegian
c. Russian
d. Korean
e. Spanish

Q6. A dependency tree for a sentence with N words includes this many dependencies:

a. N/2
b. N-1
c. N
d. 2N
e. N(N-1)/2

Q7. If f(x) is coffee beans and f'(x) is ground coffee, what is f''(x)?

a. cup of water
b. cup of green tea
c. cup of black coffee
d. cup of venti non-fat ice caramel macchiato with extra foam, to go

Q8. Which of the following areas of linguistics deals with the derivation of the word "computer" from the word "compute"?

(a) Inflectional morphology.
(b) Derivational morphology.
(c) Lexical semantics.
(d) Compositional semantics.
(e) None of the above.

Q9. The cosine similarity between the vectors (1,2,0) and (1,1,2) is:

$$\cos(\theta) = \frac{\mathbf{A} \cdot \mathbf{B}}{\|\mathbf{A}\|_2 \|\mathbf{B}\|_2} = \frac{\sum\limits_{i=1}^{n} A_i B_i}{\sqrt{\sum\limits_{i=1}^{n} A_i^2} \sqrt{\sum\limits_{i=1}^{n} B_i^2}}$$

(a) 0.5
(b) 0.32
(c) 1
(d) 0.55
(e) 0.80

Q10. A collection includes 1,000,000 documents. The word "saline" appears in 1,000 of these documents: in 200 of them it appears once in each document, in the next 500 of them it appears twice in each document, and in the final 300 of them it appears nine times in each document. What is the IDF (inverse document frequency) of the word "saline", rounded to the nearest integer?

(a) 3
(b) 11
(c) 5
(d) 9
(e) 7

Q11. How would you represent the sentence "Exactly one student passed the test" in First Order Logic (FOL)?

(a) ∃x: student(x) ∧ passed (x, test) ∧ [∀y: (student (y) ∧ passed(y, test)) ==> x=y]
(b) ∀x : student (x) ==> ¬ passed (x, test)
(c) ∃x: student (x) ∧ ¬ passed (x, test)
(d) ¬ [∃x : student (x) ∧ ¬ passed (x, test)]
(e) ∃x : student (x) ∧ passed (x, test) ∧ [∃y : student (y) ∧ passed (y, test) ∧ x = y]

Q12. Consider the following two sequences of parts of speech (each sequence corresponds to a sentence).

DT JJ NN PRP VBP TO VB DT NN IN VB IN DT NNS
DT NN IN DT NN NN RB VBZ PRP VB RP NNS

What is the maximum likelihood estimate (MLE) for the probability of bigram "NN IN"?

(a) 0
(b) .6
(c) .25
(d) .4
(e) .1

Q13. What is the complex type for a preposition in CCG?

(a) (NP\NP)/NP
(b) (S\NP)/NP
(c) NP/N
(d) NP\NP
(e) S/NP

Q14. Consider the following segment of a Wordnet-like ontology, augmented with subtree probabilities. What is the "Lin" similarity between chicken and dog?

```
                                   |
                              entity (1.00)
                            /                \
         animal (0.50)                        building (0.50)
       /                    \
    mammal (0.375)            bird (0.125)
  /                \                       \
cat (0.125)          dog (0.25)              chicken (0.125)
```

(a) 0.125
(b) 0.200
(c) 0.250
(d) 0.400
(e) 0.500

Q15. What is "acerola"?

(a) a fast train connecting New York to Boston
(b) a fruit with soft pulp
(c) a roughly straight-line configuration of three or more celestial bodies in a gravitational system
(d) a deep learning library
(e) a Pokemon character

Q16. According to the rules of ITG (Inversion Transduction Grammar), how many reorderings are allowed for the production

```
NP -> ART CARD JJ NN?
```

(a) 1
(b) 2
(c) 4
(d) 16
(e) 24

Q17. True or False: Given the feature structures FS1 and FS2,

FS1

```
[ agr  = [ number = 'singular' ] ]
[        [ person = 1          ] ]
[                                ]
[ type = 'NP'                    ]
```

FS2

```
[ agr  = [ number = ?n ] ]
[                        ]
[ subj = [ number = ?n ] ]
```

the output of their unification Unify (FS1, FS2) is correctly shown below.

FS3

```
[ agr  = [ number = 'singular' ] ]
[        [ person = 1          ] ]
[                                ]
[ subj = [ number = 'singular' ] ]
[                                ]
[ type = 'NP'                    ]
```

(a) TRUE
(b) FALSE
(c) no idea

Q18. IBM Models 1, 2, and 3 are used for:

(a) statistical machine translation

(b) neural machine translation
(c) constituent parsing
(d) dependency parsing
(e) none of the above

Q19. In order to prevent the possibility of an artificially high machine translation score, BLEU includes the following component:

(a) counting crossing brackets
(b) finetuning
(c) “forget” gates
(d) brevity penalty
(e) subcategorization

Q20. What grammatical formalism is specifically designed to handle sentences like this one: “I bought a hat for my son and a book for my daughter”?

(a) Head Driven Phrase Structure Grammar (HPSG)
(b) Dependency Grammar
(c) Regular Grammar
(d) Context Sensitive Grammar (CSG)
(e) Combinatory Categorial Grammar (CCG)

Q21. Given the sentence “When you get home, I will have fixed the sink”, in what order do the utterance (U), reference point (R), and event (E) occur? The “<” symbol below means “precedes”.

(a) E < R < U
(b) U < E < R
(c) U < (R,E)
(d) (U,R) < E
(e) E < U < R

Q22. Which of the following techniques is *not* used by BERT:

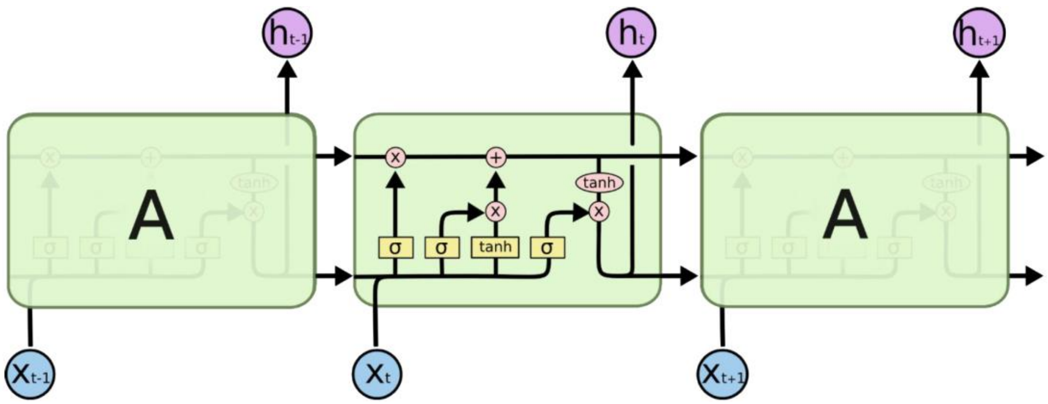
(a) bidirectional encoding
(b) attention
(c) parallel corpora
(d) stacked encoders
(e) position embeddings

Q23. What type of neural network uses the equations below?

$$h_t = \sigma(W_h h_{t-1} + W_x x_t)$$
$$y_t = softmax(W_y h_t)$$

(a) recurrent neural network
(b) recursive neural tensor network
(c) convolutional network
(d) gated recurrent unit
(e)long short term memory network

Q24. What type of neural network is represented here (image from Chris Olah)?



(a) recurrent neural network
(b) recursive neural tensor network
(c) convolutional network
(d) gated recurrent unit
(e)long short term memory network

Q25. In the sentence “She is eating a red apple”, which of the following pairs of words would exhibit the *highest* attention scores?

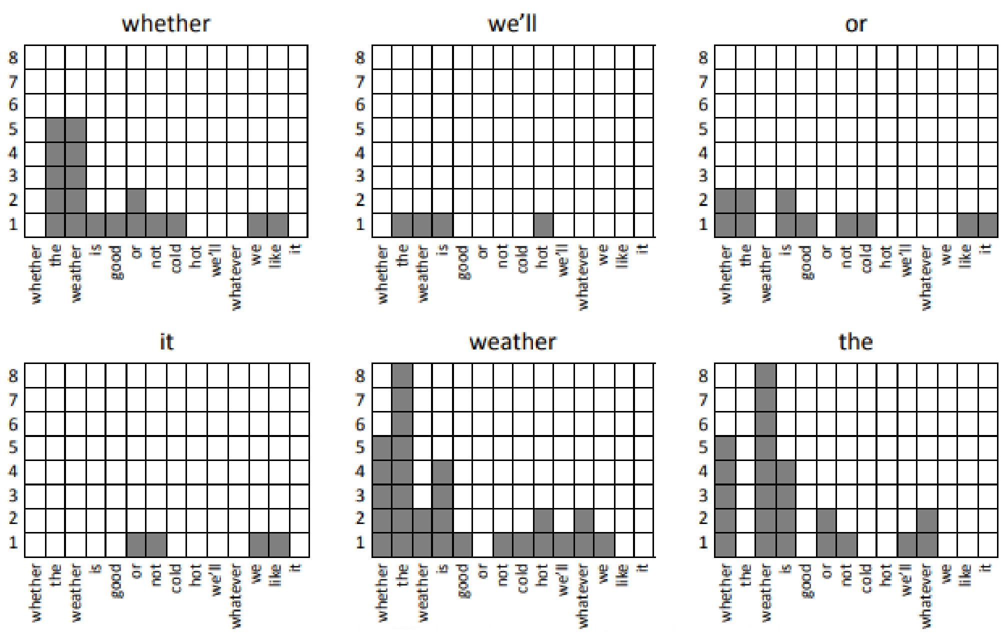
(a) she, red
(b) is, a
(c) eating, red
(d) she, apple
(e) eating, apple

Part 5: Bonus Question
5 points

Q1. There are many ways to represent a word's meaning in terms of its distribution. For this question, we will build count vectors for words using the following poem:

> *Whether the weather is good, or whether the weather is not,*
> *Whether the weather is cold, or whether the weather is hot,*
> *We'll weather the weather—whatever the weather—*
> *Whether we like it or not.*

The representations of some words from this poem are shown below as obtained by counting how often each other word occurs in a certain window around the word in question.



1. Write down a vector representation for the word *is* in the same scheme as the graphs above. For example, *whether* would be encoded as <0, 5, 5, 1, 1, 2, 1, 1, 0, 0, 0, 1, 1, 0>.

Below are 33 word count vectors. These were obtained from a different sample text and show the counts of 15 words (word A through word O), but the identities of these 15 words are not given. Your task: Study these 33 word graphs and then answer the questions that follow.

Word	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
queen	1	0	2	0	8	0	0	5	0	0	0	0	3	0	0
neigh	0	8	0	0	0	6	0	0	0	0	1	0	1	0	0
man	9	0	1	0	2	0	9	0	3	0	0	0	7	3	0
kings	6	0	3	0	1	0	0	0	0	5	0	4	0	5	0
woman	3	0	4	0	10	0	9	0	3	0	0	0	10	0	0
horse	0	2	0	0	0	7	0	0	0	0	5	0	0	0	1
queens	0	0	5	0	8	0	0	0	0	6	0	3	3	2	0
uncle	9	0	0	1	0	0	0	2	0	0	4	0	1	4	0
king	7	0	0	0	1	0	0	6	0	0	0	0	0	3	0
ugliest	0	0	0	1	0	0	6	0	0	3	1	0	0	0	2
X1	1	3	3	0	0	4	0	0	0	5	10	4	0	2	0
X2	3	10	9	8	9	9	7	10	7	10	10	10	9	9	8
X3	1	9	0	0	0	3	0	6	0	0	6	0	1	0	0
X4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
X5	2	1	2	0	8	0	0	1	0	0	5	0	4	2	0

2. The 11 mystery words (X1-X5) have the following definitions (but not in this order):

a) antismartnessesquely ______
b) aunt ______
c) cats ______
d) meow ______
e) the ______

Match each definition above to its mystery word. Hint: remember the algebraic properties of word embeddings. That is, *queens* - *queen* = *kings* - *king*. These kinds of analogies can be applied in sequence to build approximate equations for words.