Samuele Serri 7069839
Ege Mert Balcik 7071632

1 Multiple Choice Questions (4 Points)

For each of the following questions, multiple of the given answers can be correct. State whether each option is correct or wrong.

You receive points for correctly identified statements, while you lose points for wrongly identified statements. However, you cannot get negative points for a question. You can also "skip" an option. Then you will neither gain nor lose any points for that. For example, you can submit your answers like: "correct, wrong, skip, skip"

1. The probability of a sentence based on a PCFG (2 points)
   - ☐ can consider more than one part of speech category that could be assigned to the words of the sentence.
   - ☐ depends on the context surrounding the sentence.
   - ☐ is the sum of the probabilities of all possible parsed trees based on the PCFG.
   - ☐ could never be zero as long as the PCFG includes the part of speech categories of all the words in the sentence.
2. Which statement(s) about neural language models is/are true? (2 points)
   - ☐ Autoregressive language models predict the next token based on previous tokens.
   - ☐ Masked language models predict a token x based on tokens to the left of x and tokens to the right of x.
   - ☐ Autoregressive language models use self-attention while masked language models don't.
   - ☐ Neural language models assign non-zero probabilities only to sequences of tokens that appear in the training data.

2 PCFG language modelling (12 points)

Consider the following sentence and the probabilistic context free grammar (PCFG) provided below.

He enjoyed pizzas with mushrooms.

Probabilistic context free grammar:
$G = \langle V, \Sigma, P, S \rangle$
$V = \{S, VP, PP, PR, NP, V, N, IN\} \bigcup \Sigma$
$\Sigma = \{$He, enjoyed, pizzas, with, mushrooms, friends$\}$
S=S

P =

rule	probability
S →NP VP	0.45
S →VP	0.39
VP →VP NP	0.33
VP →VP PP	0.24
VP →V	0.07
PP →IN NP	0.56
NP →PR NP	0.22
NP →PR	0.38
NP →NP PP	0.71
NP →NN	0.13
PR →"He"	0.46
V →"enjoyed"	0.04
NN →"pizzas"	0.013
NN →"mushrooms"	0.019
NN →"friends"	0.091
IN →"with"	0.075

1. This sentence can be analyzed into two different syntactic trees based on the PCFG provided. Draw the two syntactic trees. (5 points)
2. Calculate the probability of each of the trees. (4 points)
3. Has the PCFG assigned a higher probability to the correct parsed tree? Why do you think this is the correct parsed tree? (1 point)
4. When this PCFG is used to analyze the sentence *"He enjoyed pizzas with friends"*, would the same syntactic structure be preferred (i.e. be assigned higher probability)? Why? (2 points)

3 Semantics (4 points)

1. The following table shows the counts of the words occurring in the context of **breakfast**, **car** and **lunch**.

	counts of words in context								
	cook	menu	eat	food	tasty	drive	fast	driver	park
breakfast	23	11	21	35	11	13	1	1	0
car	0	0	0	0	0	25	28	12	16
lunch	11	14	29	17	16	0	0	0	2

(a) Calculate the cosine similarity between *"breakfast"* and *"car"*, and *"breakfast"* and *"lunch"*. Explain in terms of distributional semantics why *"breakfast"* is more similar to *"lunch"* than *"car"*. (3 points)

(b) Why shouldn't we include function words such as prepositions and determiners when counting the contextual words? (1 point)

Submission Details:

Upload your submission to our CMS in groups of two to three students until *December 22, 2024 at 17:59.* Late submissions will not be graded! The submission should be uploaded by exactly **one** team member. Make sure that your submission contains the name and matriculation number of each team member. Submit your solution as a ***pdf* file** with your answers.

1.

1. The probability of a sentence based on a PCFG (2 points)

Skip ☐ can consider more than one part of speech category that could be assigned to the words of the sentence.

Wrong ☐ depends on the context surrounding the sentence.

Correct ☐ is the sum of the probabilities of all possible parsed trees based on the PCFG.

Correct ☐ could never be zero as long as the PCFG includes the part of speech categories of all the words in the sentence.

2. Which statement(s) about neural language models is/are true? (2 points)

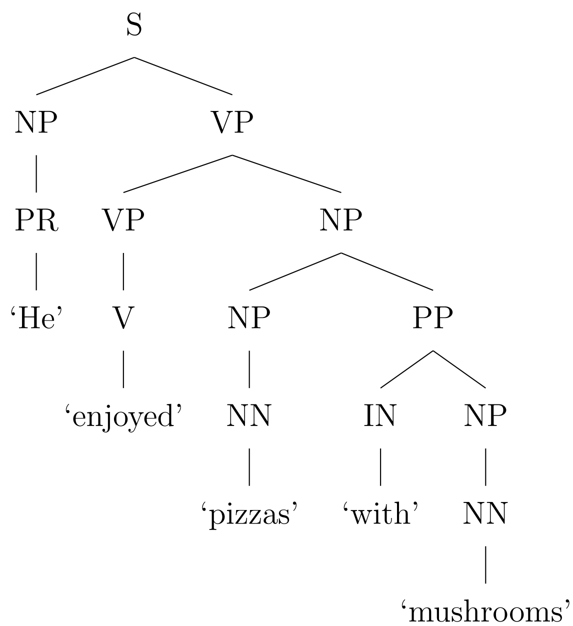
Correct ☐ Autoregressive language models predict the next token based on previous tokens.

Correct ☐ Masked language models predict a token x based on tokens to the left of x and tokens to the right of x.

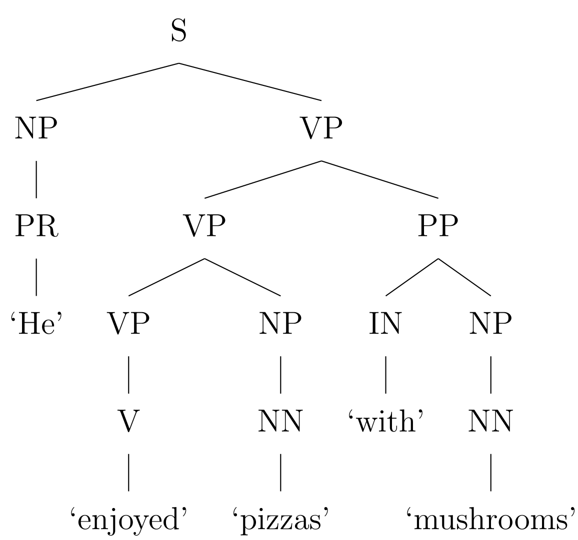
Wrong ☐ Autoregressive language models use self-attention while masked language models don't.

Wrong ☐ Neural language models assign non-zero probabilities only to sequences of tokens that appear in the training data.

2



$P(t1) = 0.45 \times 0.38 \times 0.46 \times 0.33 \times 0.07 \times 0.04 \times 0.71 \times 0.38 \times 0.013 \times 0.56 \times 0.075 \times 0.38 \times 0.019 = 7.73 \times 10^{-11}$



$P(t2) = 0.45 \times 0.38 \times 0.46 \times 0.24 \times 0.33 \times 0.07 \times 0.04 \times 0.13 \times 0.013 \times 0.56 \times 0.075 \times 0.38 \times 0.019 = 8.94 \times 10^{-12}$

The PCFG assigned the higher probability to the correct parsed tree. The first tree is correct because *"with mushrooms"* is a modifier of the noun *"pizzas"*. In the sentence *"He enjoyed pizzas with friends"* we should prefer another syntactic structure because *"with friends"* has a different logical meaning.

3.

a)

	counts of words in context								
	cook [1]	menu [2]	eat [3]	food [4]	tasty [5]	drive [6]	fast [7]	driver [8]	park [9]
A ← **breakfast**	23	11	21	35	11	13	1	1	0
B ← **car**	0	0	0	0	0	25	28	12	16
C ← **lunch**	11	14	29	17	16	0	0	0	2

Cosine Similarity Between "breakfast" and "car":

$$\frac{A \cdot B}{\|A\| \cdot \|B\|} = \frac{\sum_{i=1}^{n=9} A_i B_i}{\sqrt{\sum_{i=1}^{n=9} A_i^2} \cdot \sqrt{\sum_{i=1}^{n=9} B_i^2}} = \frac{23 \cdot 0 + 11 \cdot 0 + 21 \cdot 0 + 35 \cdot 0 + 11 \cdot 0 + 13 \cdot 25 + 1 \cdot 28 + 1 \cdot 12 + 0 \cdot 16}{\sqrt{23^2 + 11^2 + 21^2 + 35^2 + 11^2 + 13^2 + 1^2 + 1^2 + 0^2} \cdot \sqrt{0^2 + 0^2 + 0^2 + 0^2 + 0^2 + 25^2 + 28^2 + 12^2 + 16^2}} = \frac{365}{\sqrt{2608} \cdot \sqrt{1809}} = 0.168$$

Cosine Similarity Between "breakfast" and "lunch":

$$\frac{A \cdot C}{\|A\| \cdot \|C\|} = \frac{\sum_{i=1}^{n=9} A_i C_i}{\sqrt{\sum_{i=1}^{n=9} A_i^2} \cdot \sqrt{\sum_{i=1}^{n=9} C_i^2}} = \frac{23 \cdot 11 + 11 \cdot 14 + 21 \cdot 29 + 35 \cdot 17 + 11 \cdot 16 + 13 \cdot 0 + 1 \cdot 0 + 1 \cdot 0 + 0 \cdot 2}{\sqrt{23^2 + 11^2 + 21^2 + 35^2 + 11^2 + 13^2 + 1^2 + 1^2 + 0^2} \cdot \sqrt{11^2 + 14^2 + 29^2 + 17^2 + 16^2 + 0^2 + 0^2 + 0^2 + 2^2}} = \frac{1787}{\sqrt{2608} \cdot \sqrt{1707}} = 0.847$$

The more the result of cosine similarity is closer to 1, the more two contexts are similar. Hence, "breakfast" is more similar to "lunch" than "car". This means that they share more similar contexts.

b)

Because when they come together with other words, they may form collocations, which usually drastically changes the original meaning of the word.