

# COMP 472: Artificial Intelligence

## Natural Language Processing

### *Solutions*

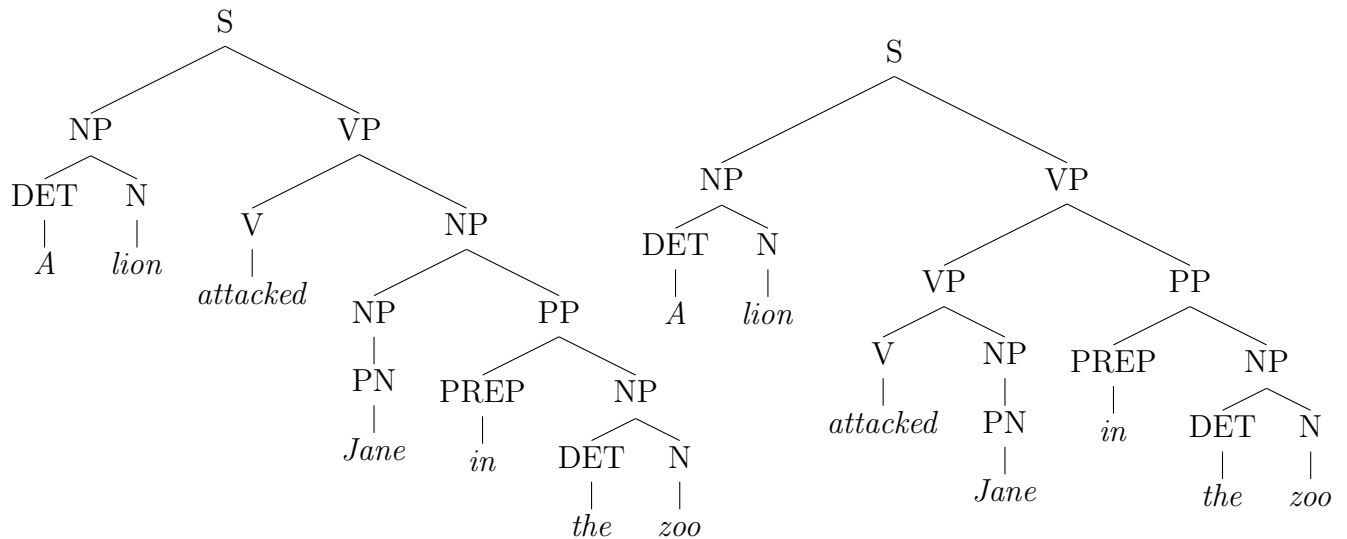
**Question 1** Consider the following context-free grammar:

$S \rightarrow NP VP$ $VP \rightarrow V$ $VP \rightarrow V NP$ $VP \rightarrow VP PP$ $NP \rightarrow DET N$ $NP \rightarrow PN$ $NP \rightarrow NP PP$ $PP \rightarrow PREP NP$	$N \rightarrow \text{lion}$ $N \rightarrow \text{knife}$ $N \rightarrow \text{zoo}$ $V \rightarrow \text{attacked}$ $PN \rightarrow \text{Jane}$ $DET \rightarrow \text{the}$ $DET \rightarrow \text{a}$ $PREP \rightarrow \text{with}$ $PREP \rightarrow \text{in}$
---	--

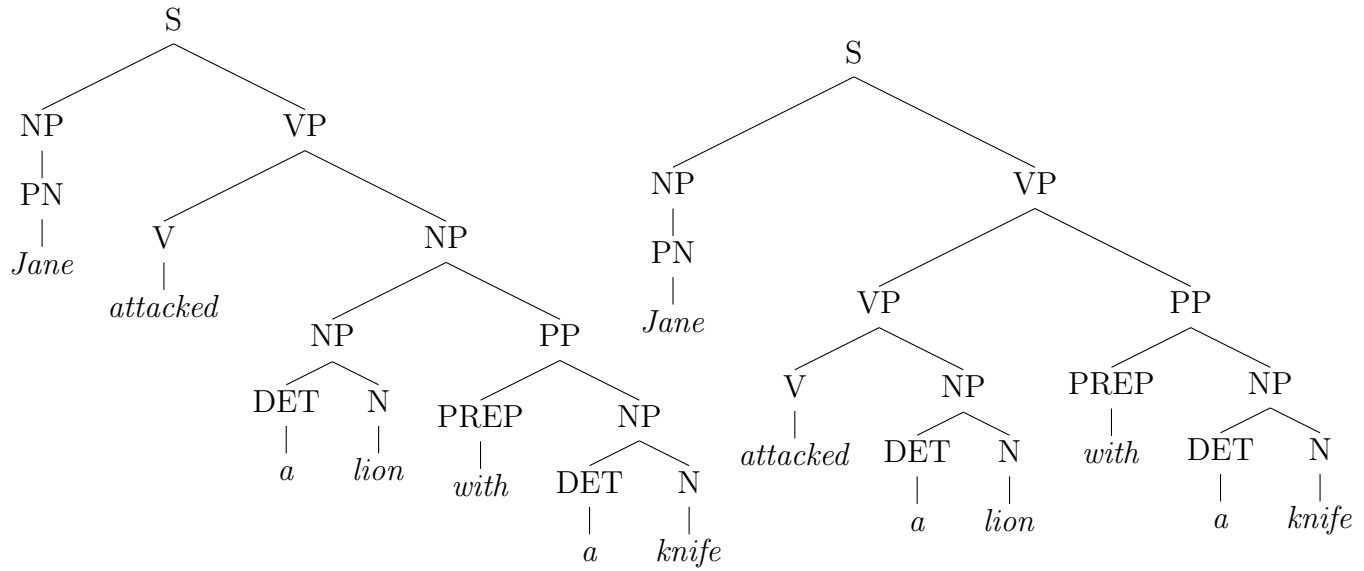
(a) Generate all possible parse trees for the following sentences:

i *A lion attacked Jane in the zoo*

There are be two possible parse trees:

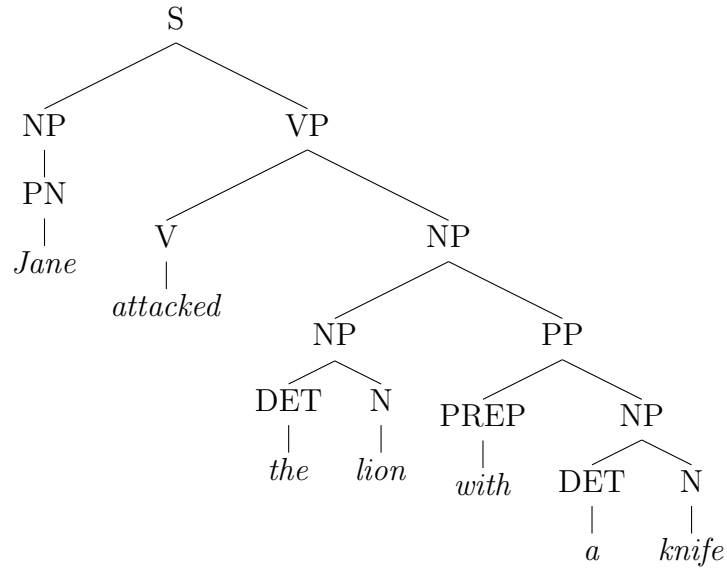


- ii *Jane attacked a lion with a knife*  
 There are two possible parse trees:

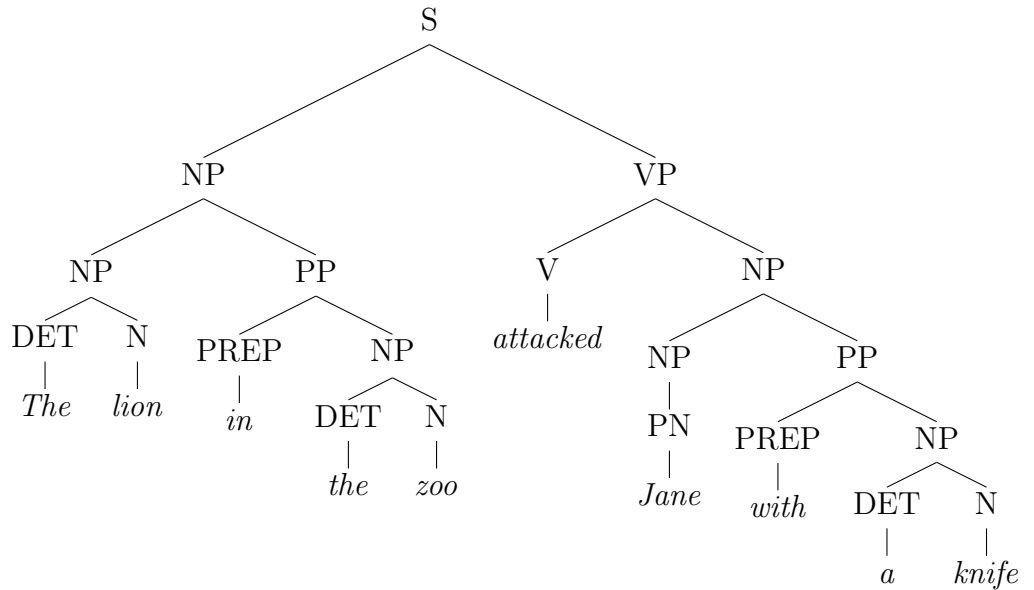


- (b) Using the following parse trees as training set, generate a probabilistic context-free grammar.

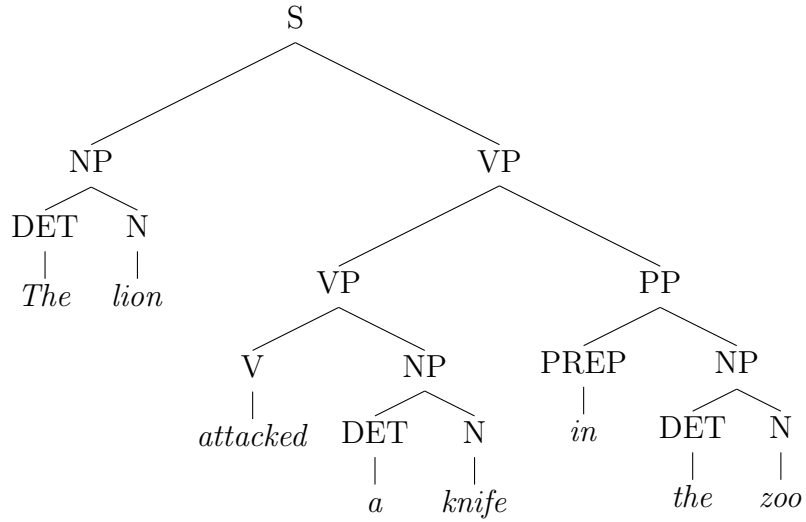
i. Jane attacked the lion with a knife



ii. The lion in the zoo attacked Jane with a knife



iii. The lion attacked a knife in the zoo



$S \rightarrow NP VP$       *Probability* = 1

$VP \rightarrow V$       *Probability* = 0

$VP \rightarrow V NP$       *Probability* =  $3/4 = 0.75$

$VP \rightarrow VP PP$       *Probability* =  $1/4 = 0.25$

$NP \rightarrow DET N$       *Probability* =  $8/13 = 0.62$

$NP \rightarrow PN$       *Probability* =  $2/13 = 0.15$

$NP \rightarrow NP PP$       *Probability* =  $3/13 = 0.23$

$PP \rightarrow PREP NP$       *Probability* = 1

$N \rightarrow lion$       *Probability* =  $3/8 = 0.375$

$N \rightarrow knife$       *Probability* =  $3/8 = 0.375$

$N \rightarrow zoo$       *Probability* =  $2/8 = 0.25$

$V \rightarrow attacked$       *Probability* = 1

$PN \rightarrow Jane$       *Probability* = 1

$DET \rightarrow the$       *Probability* =  $5/8 = 0.625$

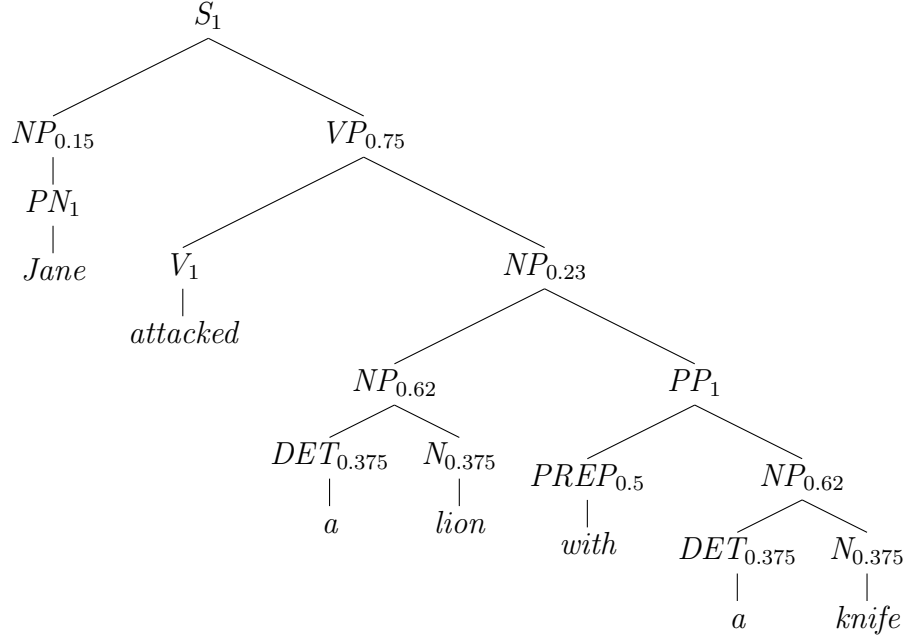
$DET \rightarrow a$       *Probability* =  $3/8 = 0.375$

$PREP \rightarrow with$       *Probability* =  $2/4 = 0.50$

$PREP \rightarrow in$       *Probability* =  $2/4 = 0.50$

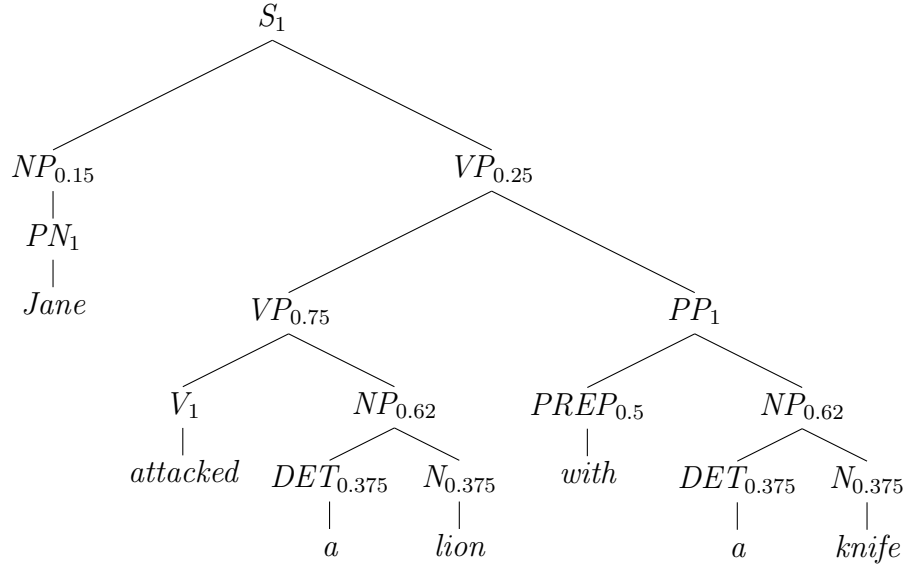
- (c) Determine the most probable parse of sentence (a)ii above (*Jane attacked a lion with a knife*) using the PCFG from (b).

**Tree 1**



$$P(\text{tree}_1) = 1 \times 0.15 \times 1 \times 0.75 \times 1 \times 0.23 \times 0.62 \times 0.375 \times 0.375 \times 1 \times 0.5 \times 0.62 \times 0.375 \times 0.375 = 9.83 \times 10^{-5}$$

**Tree 2**



$$P(\text{tree}_2) = 1 \times 0.15 \times 1 \times 0.25 \times 0.75 \times 1 \times 0.62 \times 0.375 \times 0.375 \times 1 \times 0.5 \times 0.62 \times 0.375 \times 0.375 = 1.07 \times 10^{-4}$$

Thus Tree 2 is more probable.

**Question 2** Consider the following sentences as training set for disambiguating the sense of the word *light*.

Sentence	Sense
There is a ray of light even if it is dim.	Sense1
This faint ray of light is barely enough to see.	Sense1
It needs to be light in weight to be portable.	Sense2
I can see dim red light coming from that room.	Sense1
This laptop is light enough to carry around.	Sense2
I am not hurt since I was hit by something as light as a candle.	Sense2
There is a faint hint of light at the end of the street.	Sense1
But the room was bright and light from the candle was warm as it fell on the surroundings.	Sense1
My bright red overcoat is light but fairly warm.	Sense2
This red candle stand is light enough to carry it to the room.	Sense2

Further consider this list of stop words:

*a, and, are, as, at, be, for, from, in, is, it, if, of, that, this, the, to, was*

Using a Naive Bayes approach with:

- a context window of  $\pm 3$  words,
- a vocabulary with the size of 31,
- smoothing with the value of 0.2, and
- stop-word removal

Calculate the scores of each possible sense and find the most probable sense of the word *light* in the following sentence:

*In her dark red room, that dim light from the candle was warm enough.*

The sentence with the word to disambiguate-

~~In~~ her dark red room, ~~that~~ dim **light** ~~from the~~ candle ~~was~~ warm enough.

Note : The stop words have been depicted as strike outs, since they will not be considered in predicting the sense. The words in the box are the words that are a part of the sentence and lie within the context window. The probabilities of these words for each sense in the training set will be considered.

Each of the words in the box above is also highlighted (boxed) in each of the training sentences below, when it lies within the context window of the word we are trying to disambiguate:

There is a ray of **light** even if it is dim.

This faint ray of **light** is barely enough to see.

It needs to be **light** in weight to be portable.

I can see dim red **light** coming from that room.

This laptop is **light** enough to carry around.

I am not hurt since I was hit by something as **light** as a candle.

There is a faint hint of **light** at the end of the street.

But the room was bright and **light** from the candle was warm as it fell on  
the surroundings.

My bright red overcoat is **light** but fairly warm.

This red candle stand is **light** enough to carry it to the room.

$$P(\text{Sense1}) = 5/10 = 1/2 = 0.5$$

$$P(\text{Sense2}) = 5/10 = 1/2 = 0.5$$

$$P(\text{red}|\text{Sense1}) = \frac{1 + 0.2}{25 + (31 \times 0.2)} = \frac{1.2}{31.2} = 0.038461538$$

$$P(\text{room}|\text{Sense1}) = \frac{2 + 0.2}{25 + (31 \times 0.2)} = \frac{2.2}{31.2} = 0.070512821$$

$$P(\text{dim}|\text{Sense1}) = \frac{2 + 0.2}{25 + (31 \times 0.2)} = \frac{2.2}{31.2} = 0.070512821$$

$$P(\text{candle}|\text{Sense1}) = \frac{1 + 0.2}{25 + (31 \times 0.2)} = \frac{1.2}{31.2} = 0.038461538$$

$$P(\text{warm}|\text{Sense1}) = \frac{1 + 0.2}{25 + (31 \times 0.2)} = \frac{1.2}{31.2} = 0.038461538$$

$$P(\text{enough}|\text{Sense1}) = \frac{1 + 0.2}{25 + (31 \times 0.2)} = \frac{1.2}{31.2} = 0.038461538$$

$$P(\text{red}|\text{Sense2}) = \frac{2 + 0.2}{23 + (31 \times 0.2)} = \frac{2.2}{29.2} = 0.075342466$$

$$P(\text{room}|\text{Sense2}) = \frac{1 + 0.2}{23 + (31 \times 0.2)} = \frac{1.2}{29.2} = 0.04109589$$

$$P(\text{dim}|\text{Sense2}) = \frac{0 + 0.2}{23 + (31 \times 0.2)} = \frac{0.2}{29.2} = 0.006849315$$

$$P(\text{candle}|\text{Sense2}) = \frac{2 + 0.2}{23 + (31 \times 0.2)} = \frac{2.2}{29.2} = 0.075342466$$

$$P(\text{warm}|\text{Sense2}) = \frac{1 + 0.2}{23 + (31 \times 0.2)} = \frac{1.2}{29.2} = 0.04109589$$

$$P(\text{enough}|\text{Sense2}) = \frac{2 + 0.2}{23 + (31 \times 0.2)} = \frac{2.2}{29.2} = 0.075342466$$

$$\text{score}(\text{Sense1})$$

$$= \log(\text{Sense1}) + \log(P(\text{red}|\text{Sense1})) + \log(P(\text{room}|\text{Sense1})) + \log(P(\text{dim}|\text{Sense1})) +$$

$$\log(P(\text{candle}|\text{Sense1})) + \log(P(\text{warm}|\text{Sense1})) + \log(P(\text{enough}|\text{Sense1}))$$

$$= \log(0.5) + \log(0.038461538) + \log(0.070512821) + \log(0.070512821) + \log(0.038461538) +$$

$$\log(0.038461538) + \log(0.038461538)$$

$$= -8.264387214$$

$$\begin{aligned}
& \text{score}(\text{Sense2}) \\
&= \log(\text{Sense2}) + \log(P(\text{red}|\text{Sense2})) + \log(P(\text{room}|\text{Sense2})) + \log(P(\text{dim}|\text{Sense2})) + \\
&\log(P(\text{candle}|\text{Sense2})) + \log(P(\text{warm}|\text{Sense2})) + \log(P(\text{enough}|\text{Sense2})) \\
&= \log(0.5) + \log(0.075342466) + \log(0.006849315) + \log(0.04109589) + \log(0.075342466) + \\
&\log(0.04109589) + \log(0.075342466) \\
&= -8.606666574
\end{aligned}$$

Since  $\text{score}(\text{Sense1}) > \text{score}(\text{Sense2})$ , **Sense1** is more probable.