

COMP 6751 Natural Language Analysis

Project Report 2 (**Demo**):

## *Developing Context Free Grammar*

Student name: Md Sakib Ullah Sourav

ID: 40264066

*Expectations of originality:*

*I, Md Sakib Ullah Sourav (student id 40264066), certify that this submission is my original work and meets the Faculty's Expectations of Originality.*

*Date: October 12, 2023*

## Contents

1.	Test Scenarios in Validation Data.....	3
1.1	Test Scenario 1.....	3
1.2.	Test Scenario 2.....	5
1.3	Test Scenario 3.....	11
1.4	Test Scenario 4.....	15
1.5	Test Scenario 5.....	22
1.6	Test Scenario 6.....	23
1.7	Test Scenario 7.....	24
2.	Outputs in CoNLL Format.....	33
3.	Limitations .....	37

## 1. Test Scenarios in Validation Data

(To access to my entire code, please visit this link-

[https://drive.google.com/file/d/1ps0xa8tx5Zhc\\_6lke1kPFBuC3Jz\\_sk8r/view?usp=sharing](https://drive.google.com/file/d/1ps0xa8tx5Zhc_6lke1kPFBuC3Jz_sk8r/view?usp=sharing)

And, let me know if the link is not working!)

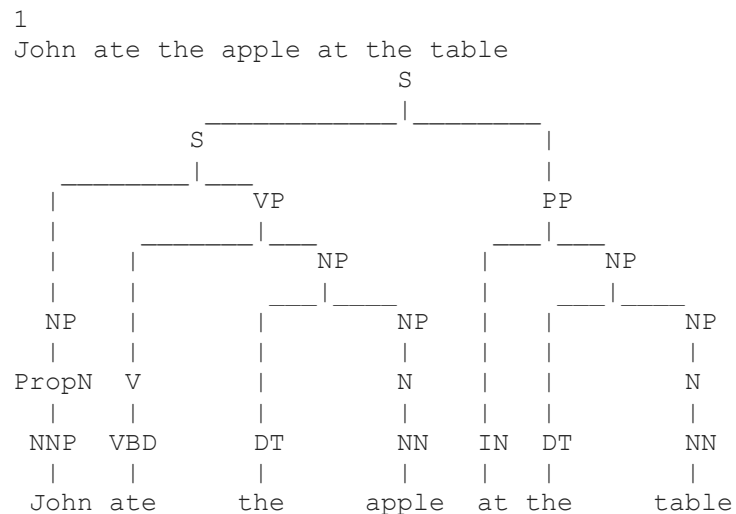
The validation data has been placed into the code as below

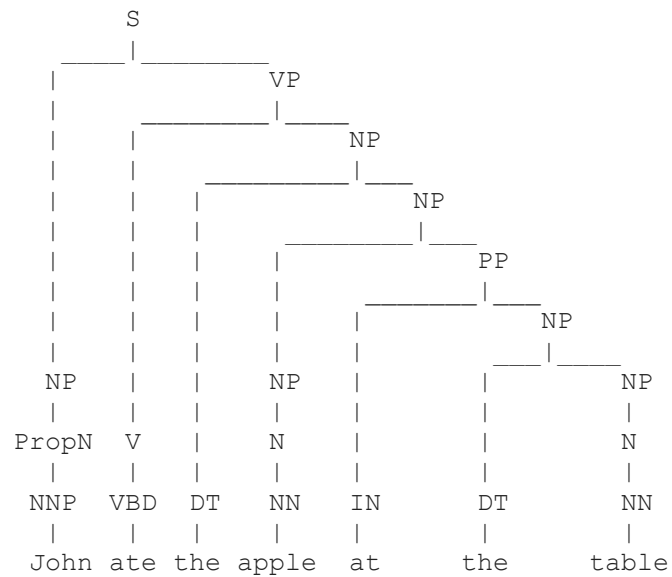
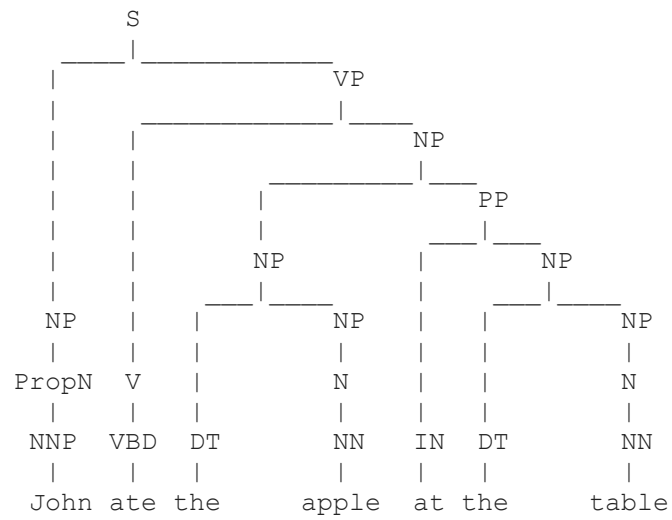
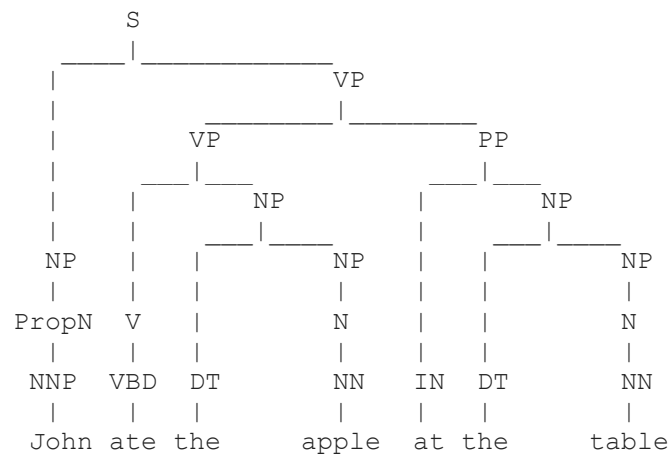
```
input = [  
    "John ate the apple at the table.",  
    "John, who ate the apple from the fridge, was sick Wednesday.",  
    "On Monday, John ate the apple in his office.",  
    "Last week, on Monday, John finally took the apple from the fridge to  
his office.",  
    "Heart rate greater than 90 beats/minute",  
    "Body temperature over 38 or under 36 degrees Celsius.",  
    "Respiratory rate greater than 20 breaths/minute or partial pressure  
of CO2 less than 32 mmHg",  
    "Leukocyte count greater than 12000 or less than 4000 /microliters or  
over 10% immature forms or bands."  
]
```

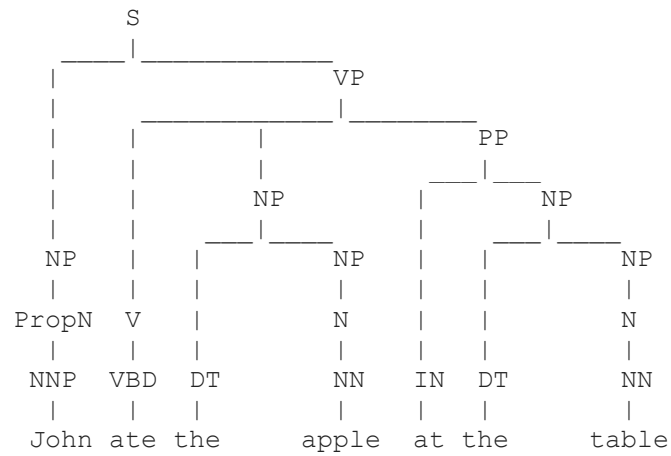
### 1.1 Test Scenario 1

Input: John ate the apple at the table.

Result: The proofreading results and text preprocess pipeline can be delivered, and the programme can operate correctly. Earley parse process results are displayed below.



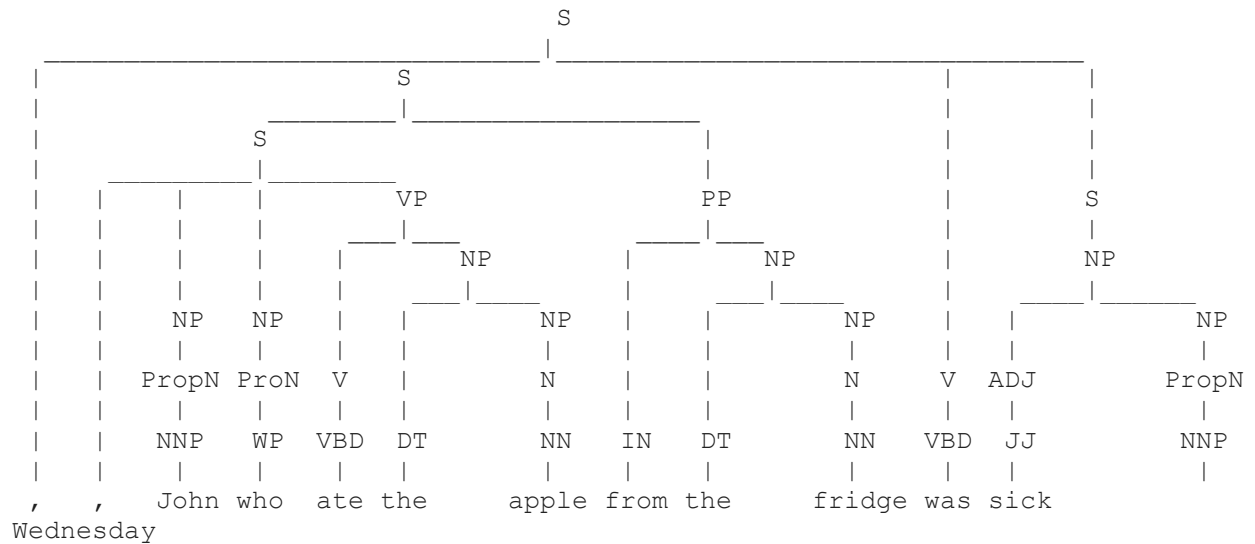


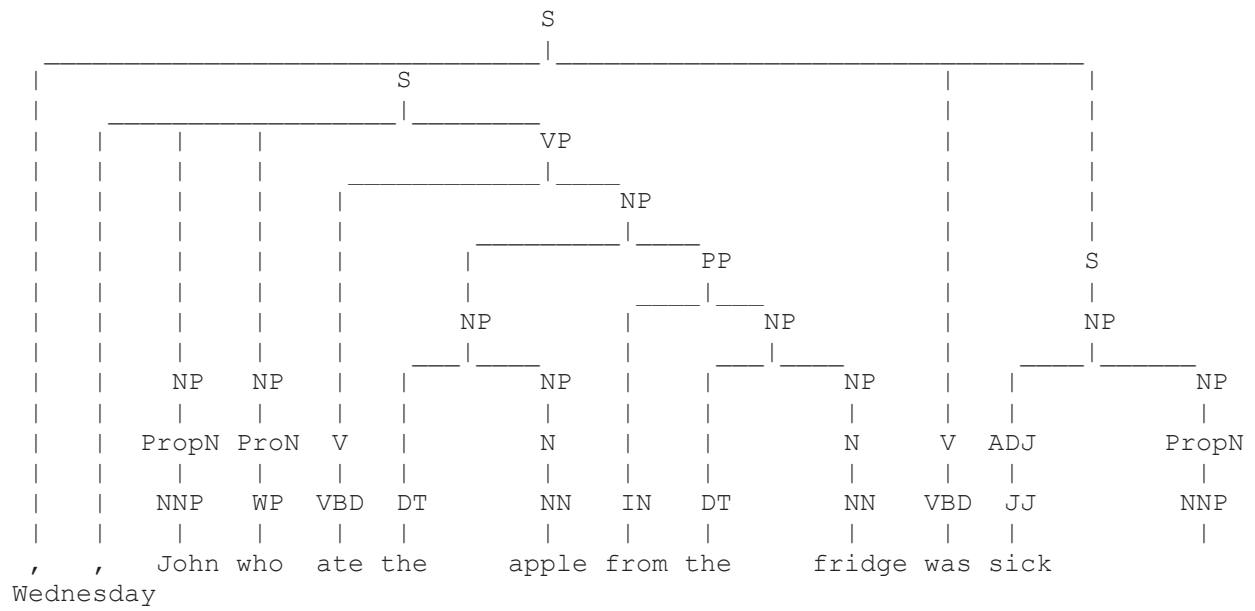
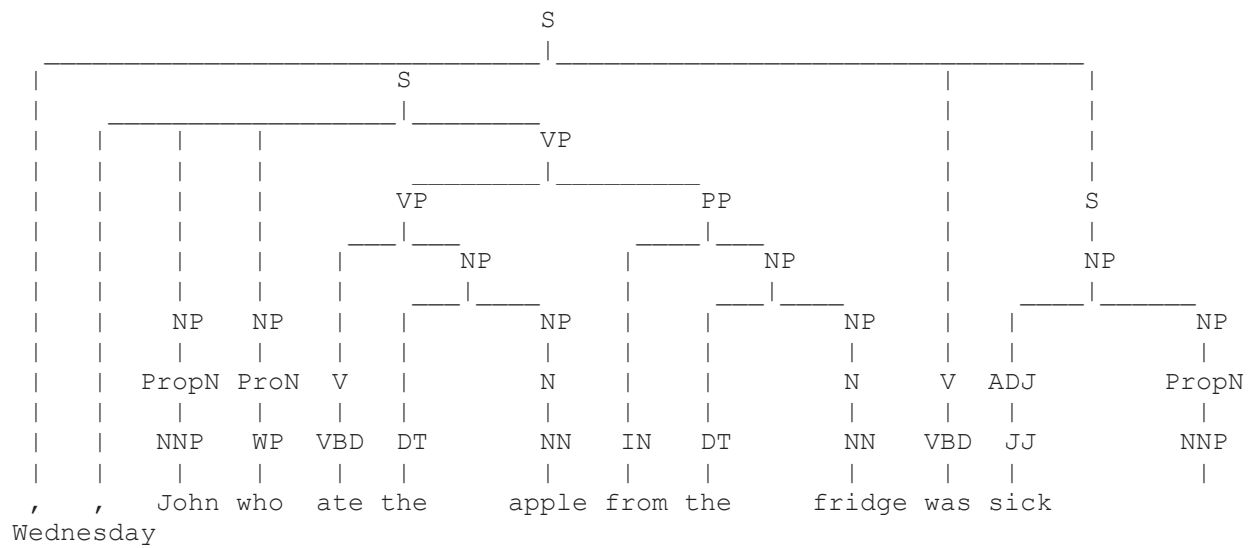


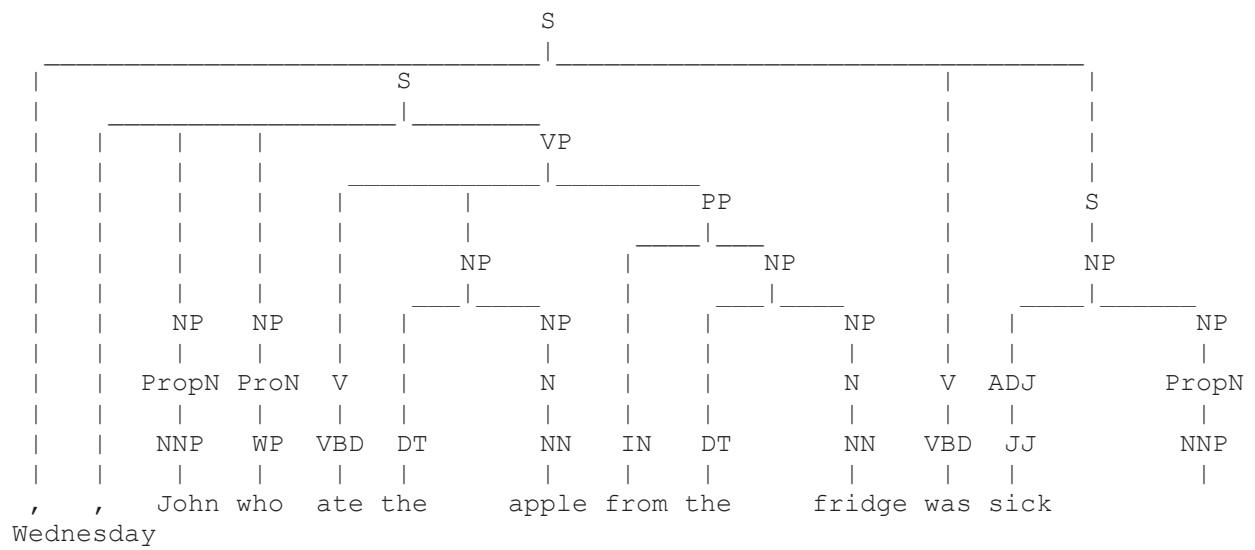
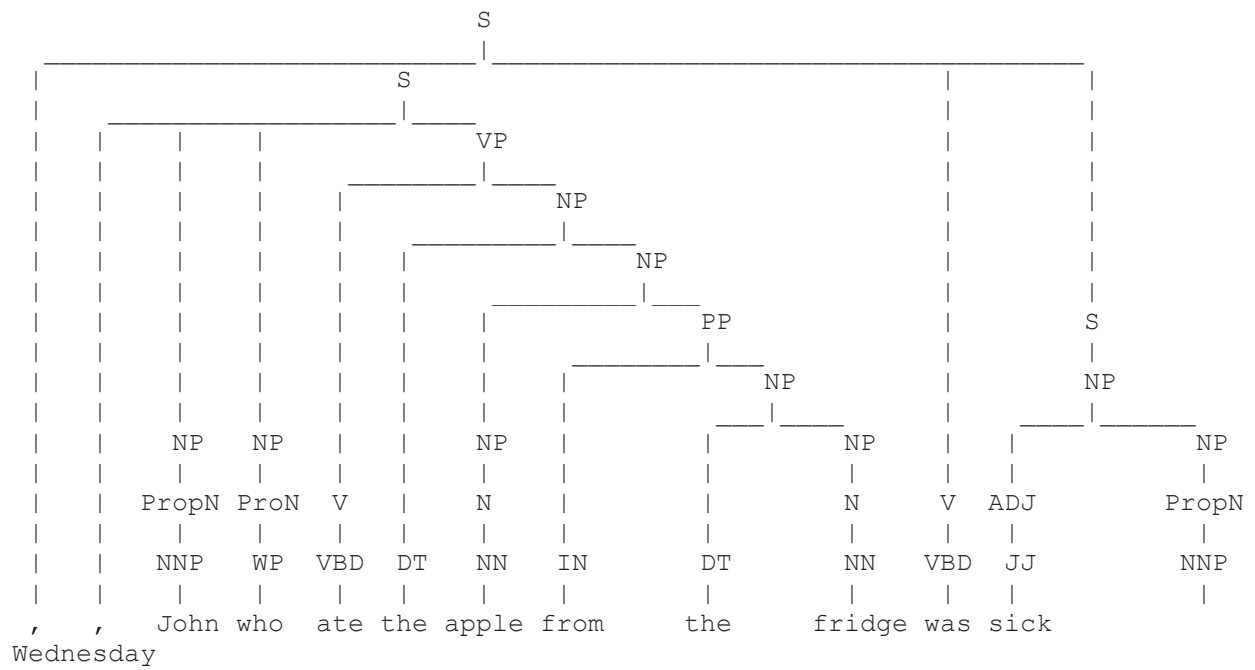
## 1.2. Test Scenario 2

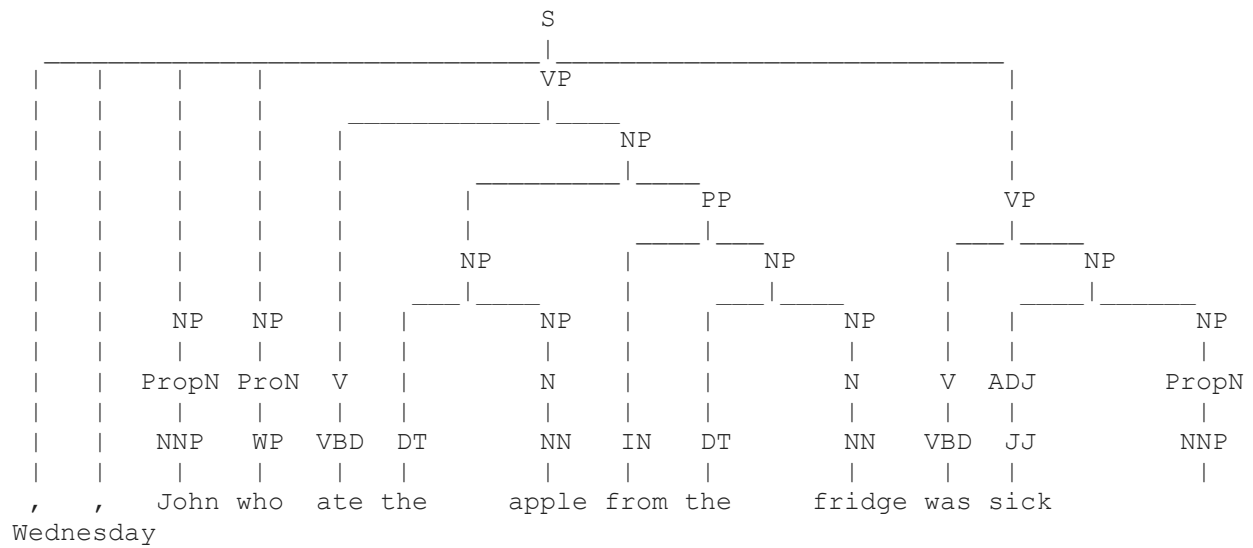
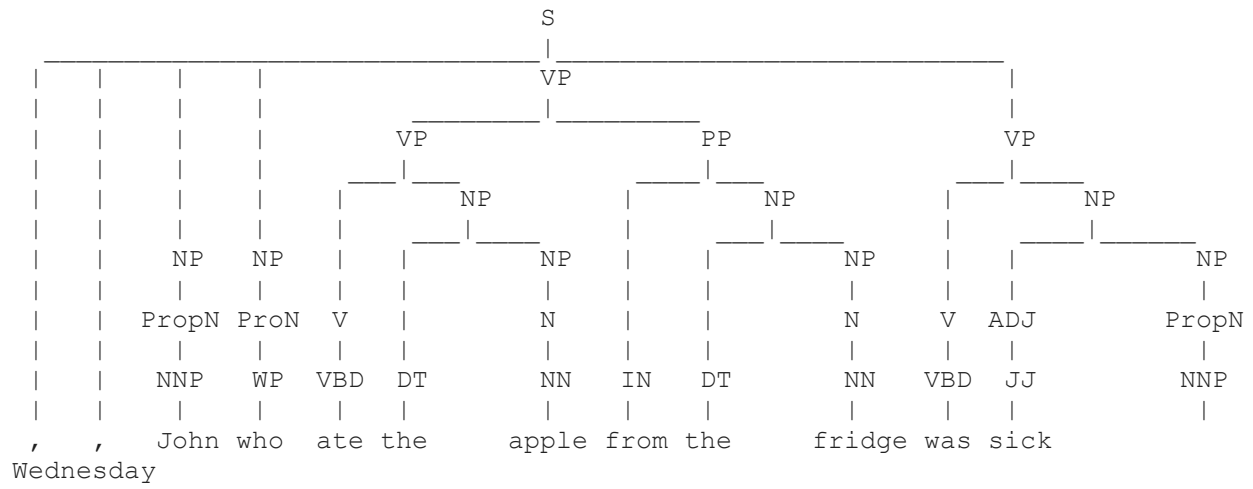
Input: John, who ate the apple from the fridge, was sick Wednesday.

Result: The proofreading results and text preprocess pipeline can be delivered, and the programme can operate correctly. Earley parse process results are displayed below.

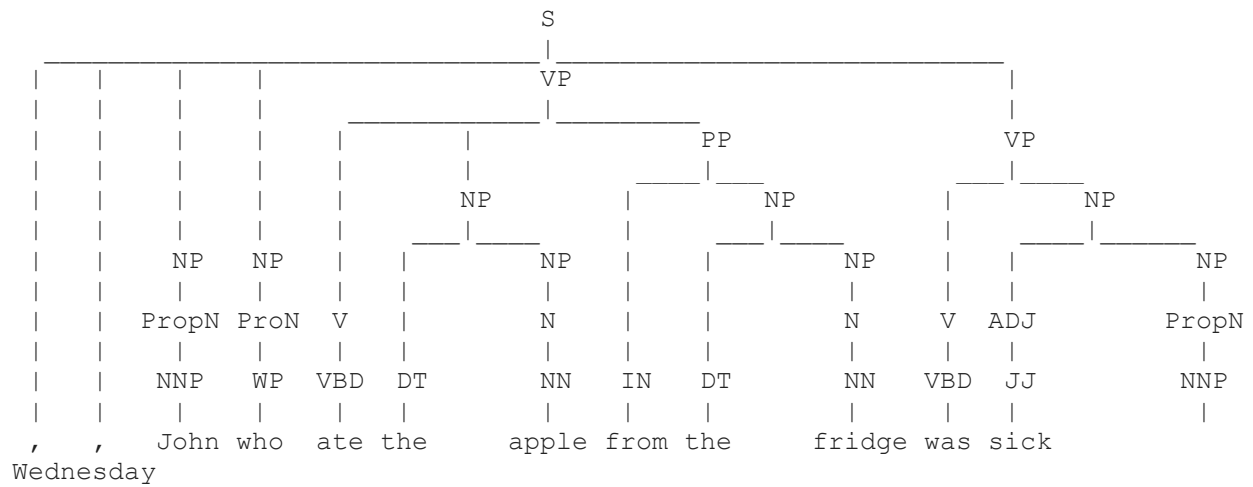
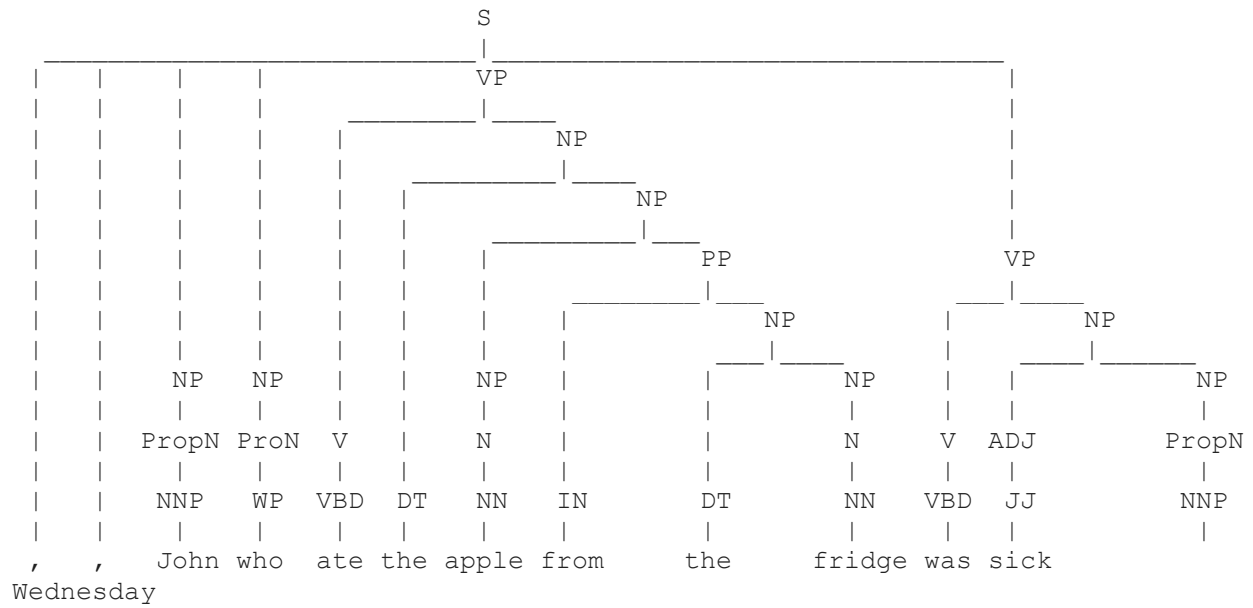


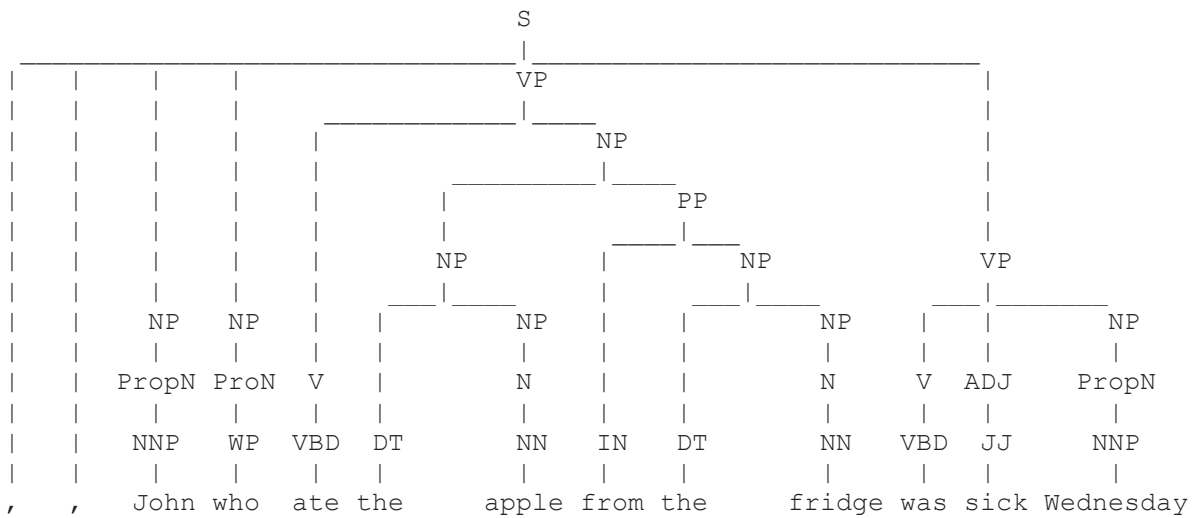
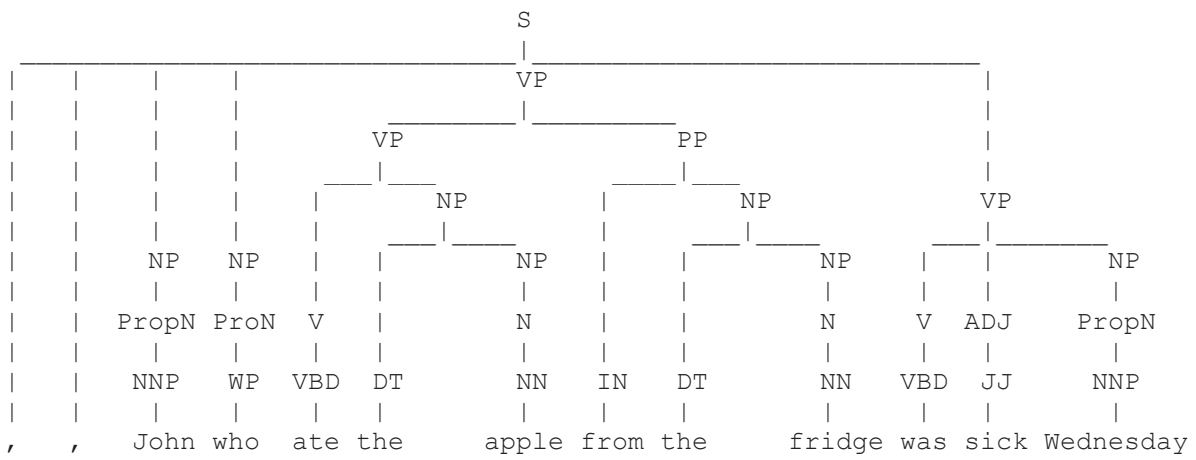


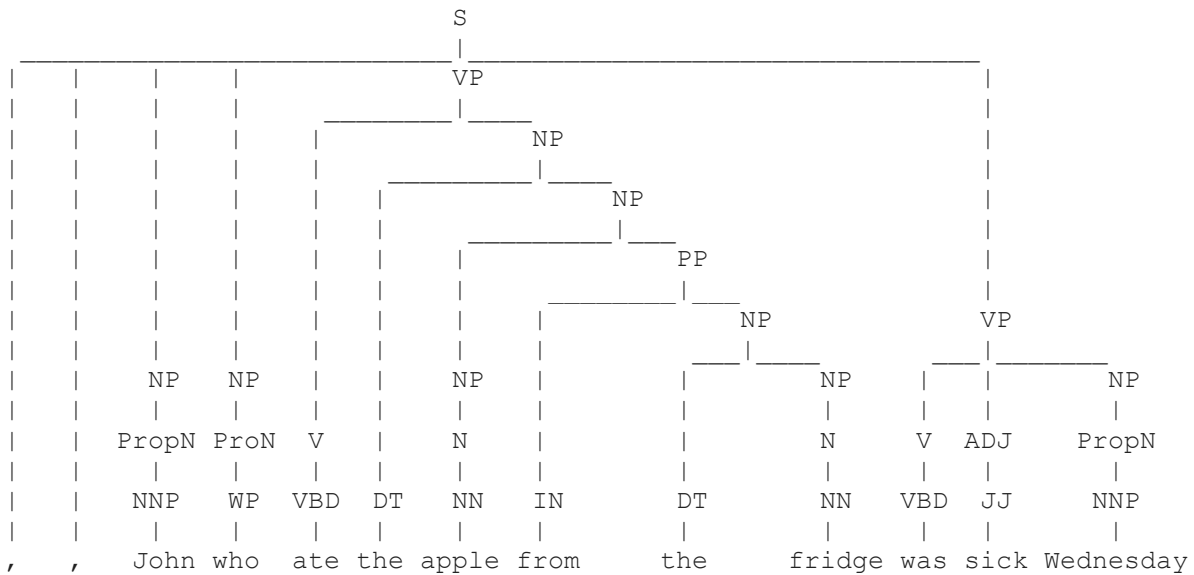










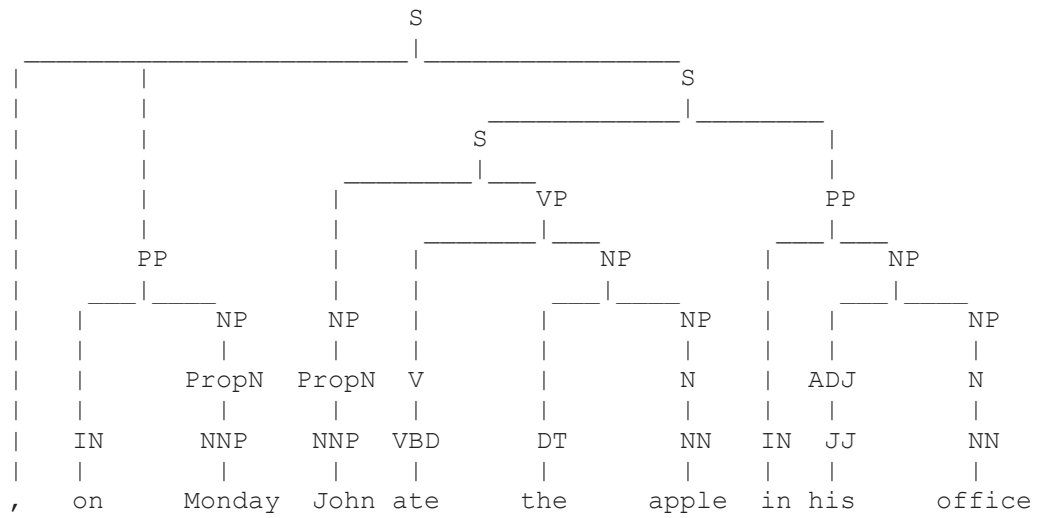
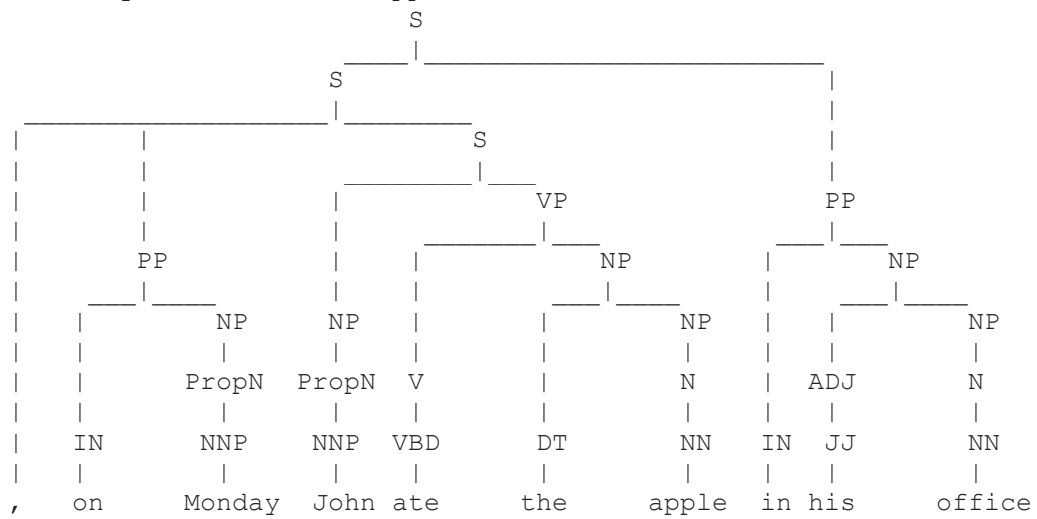


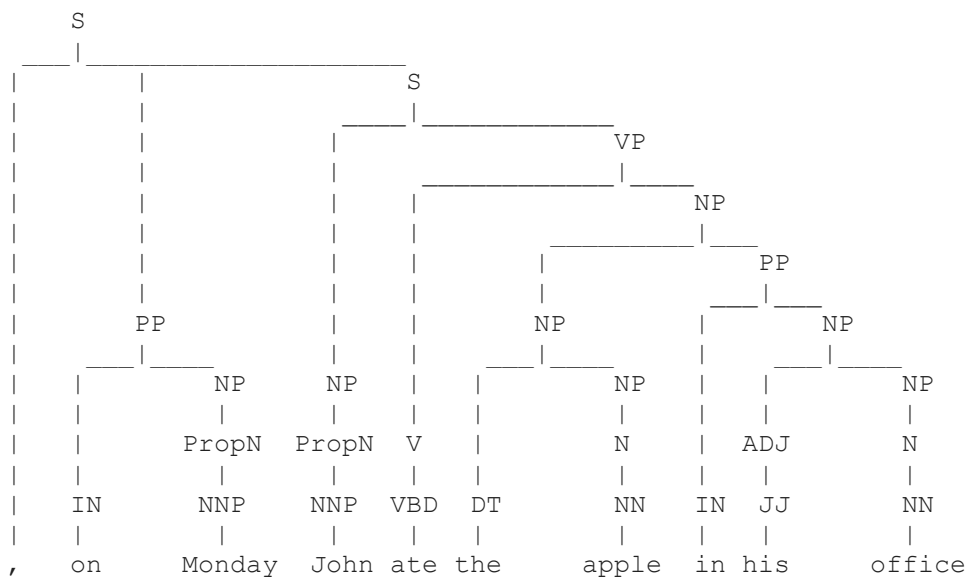
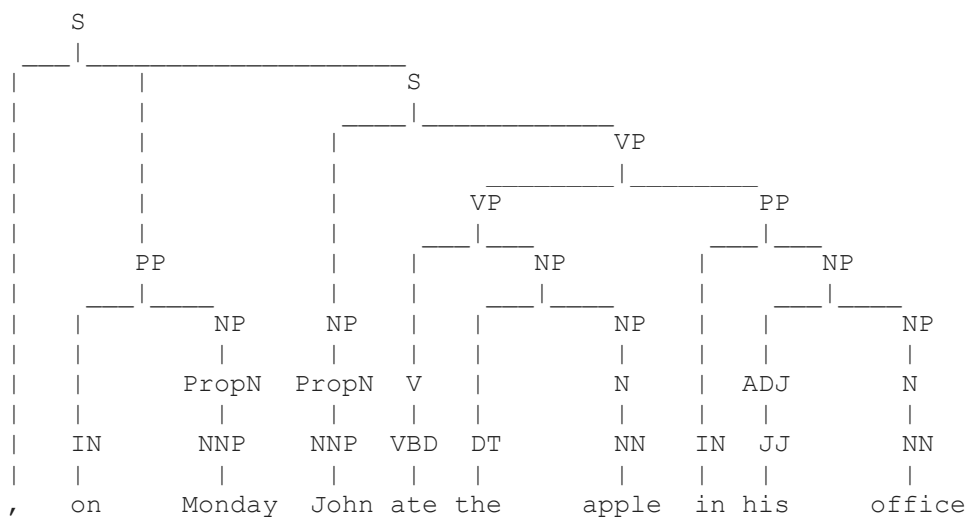
### 1.3 Test Scenario 3

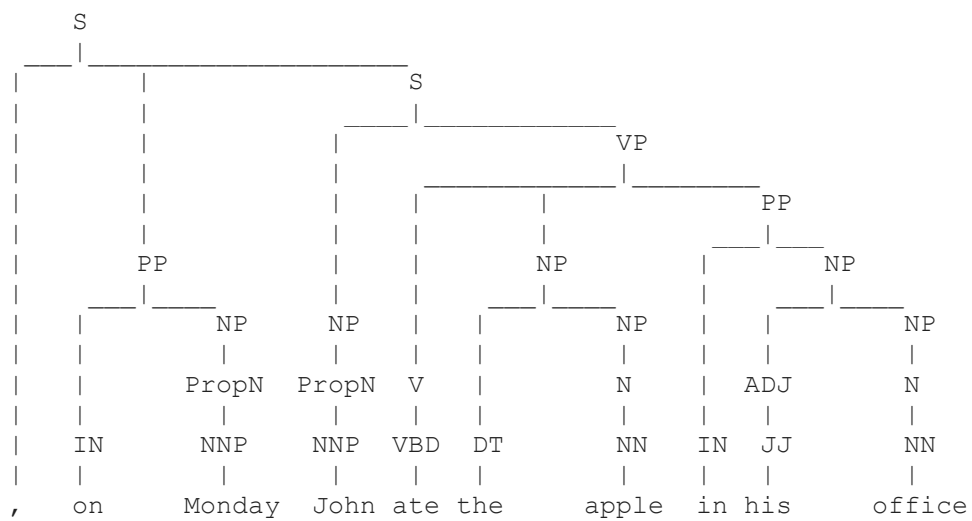
Result: The proofreading results and text preprocess pipeline can be delivered, and the programme can operate correctly. Earley parse process results are displayed below.

1

On Monday, John ate the apple in his office



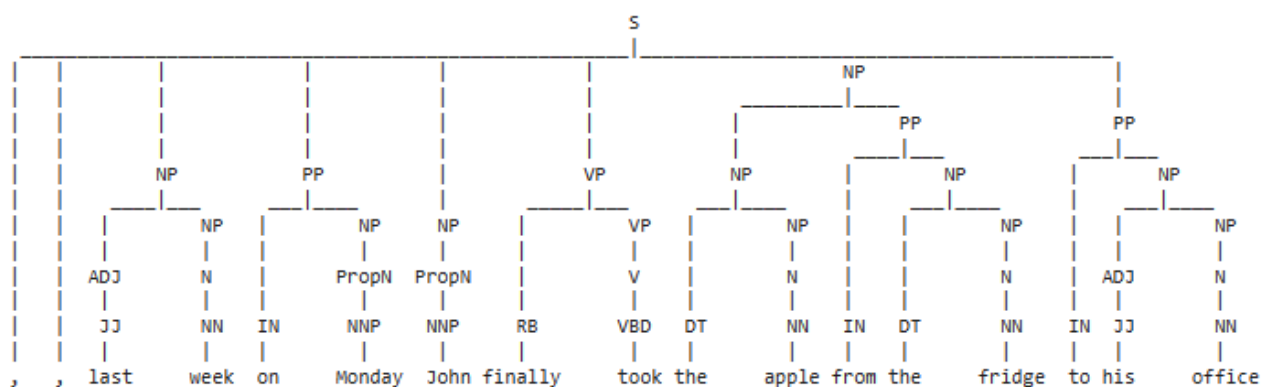
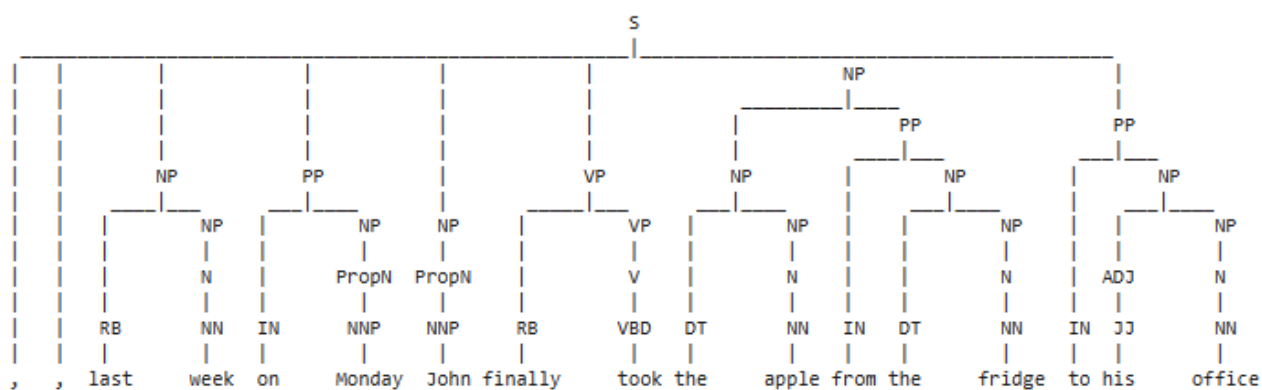


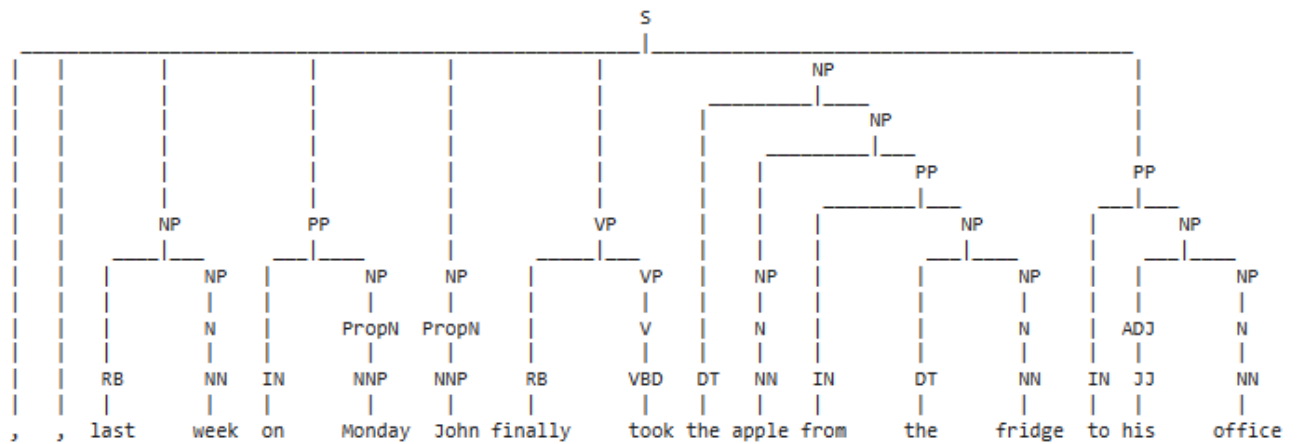
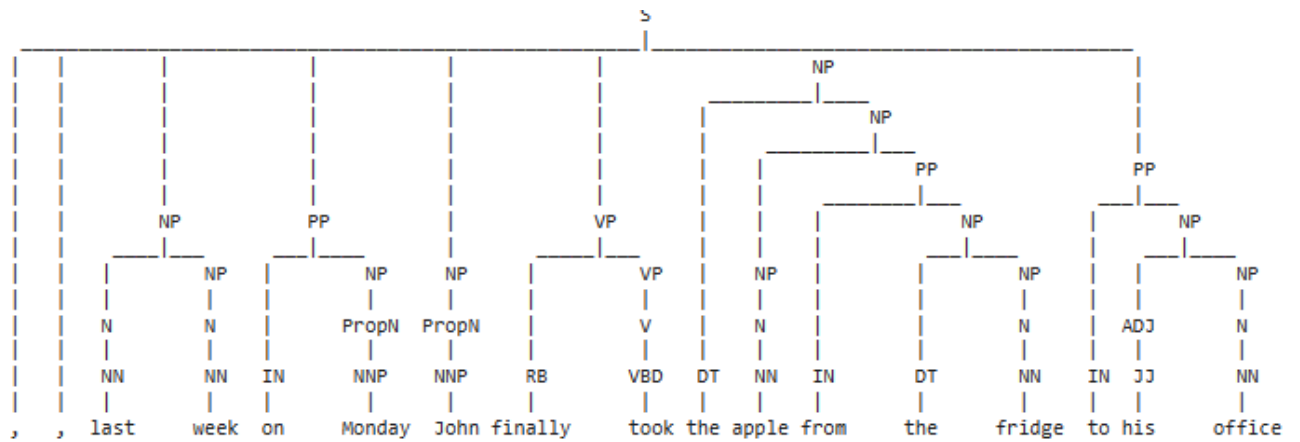


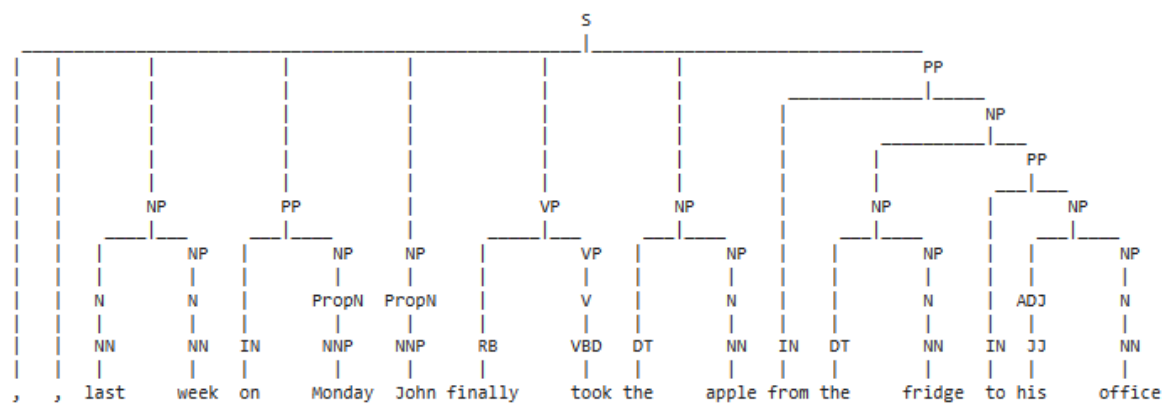
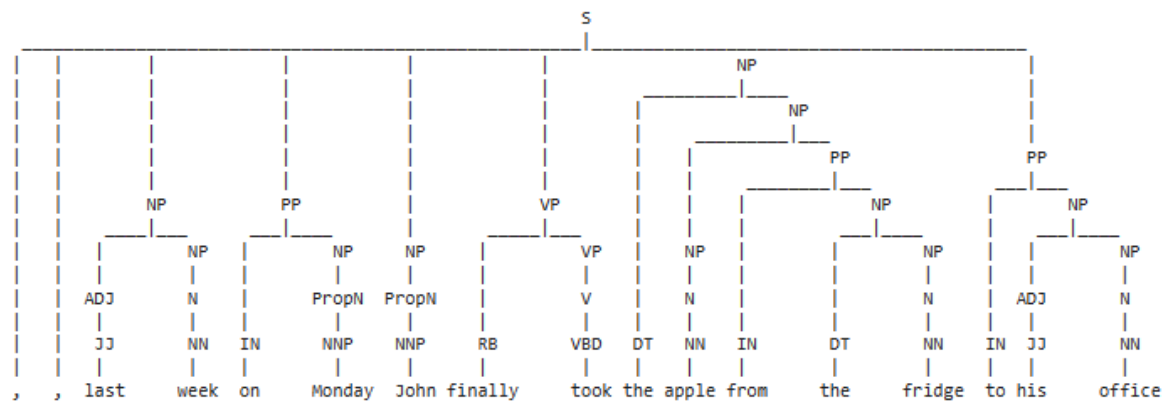
[illegible]

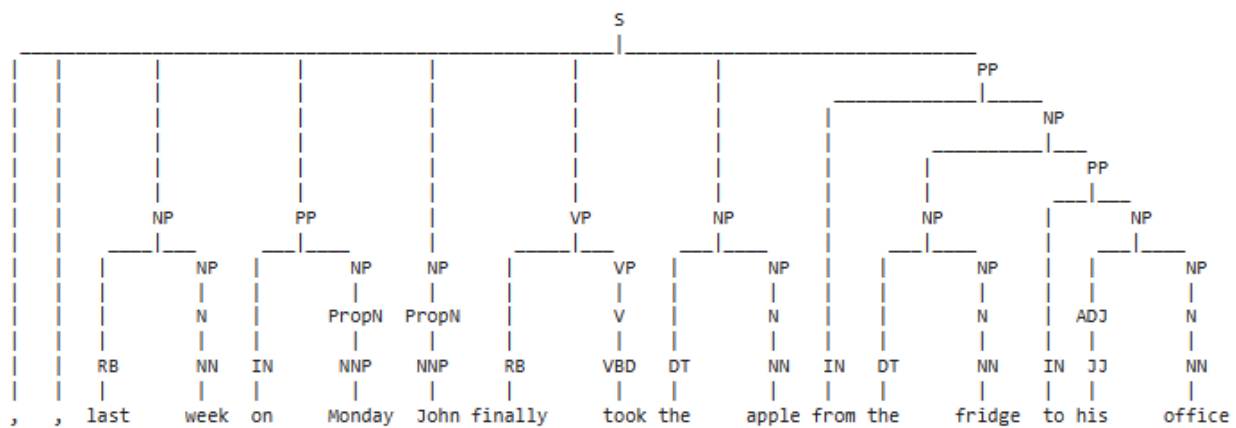
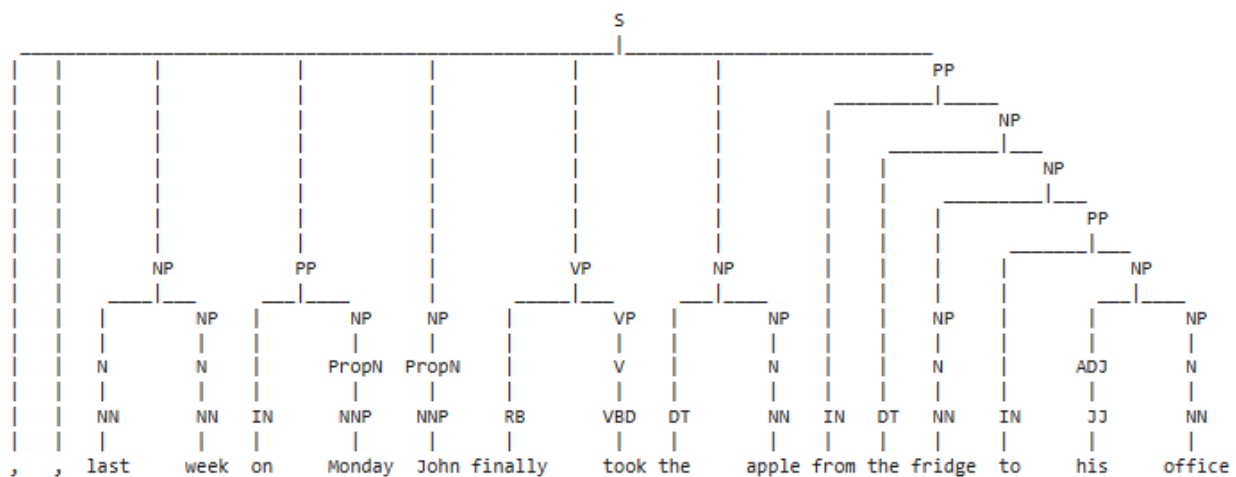


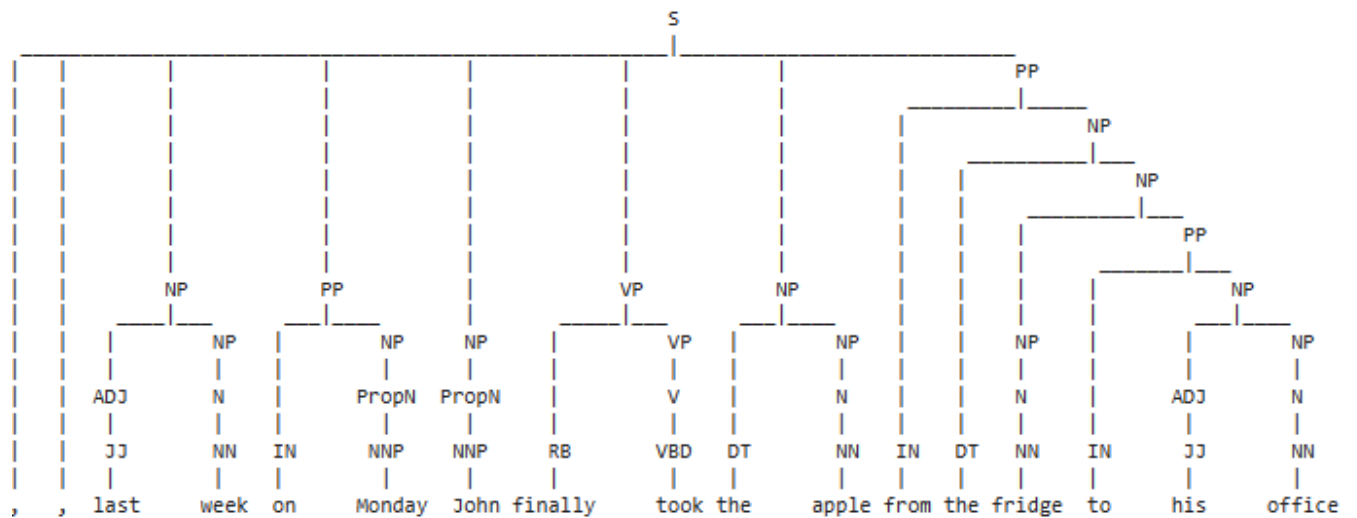
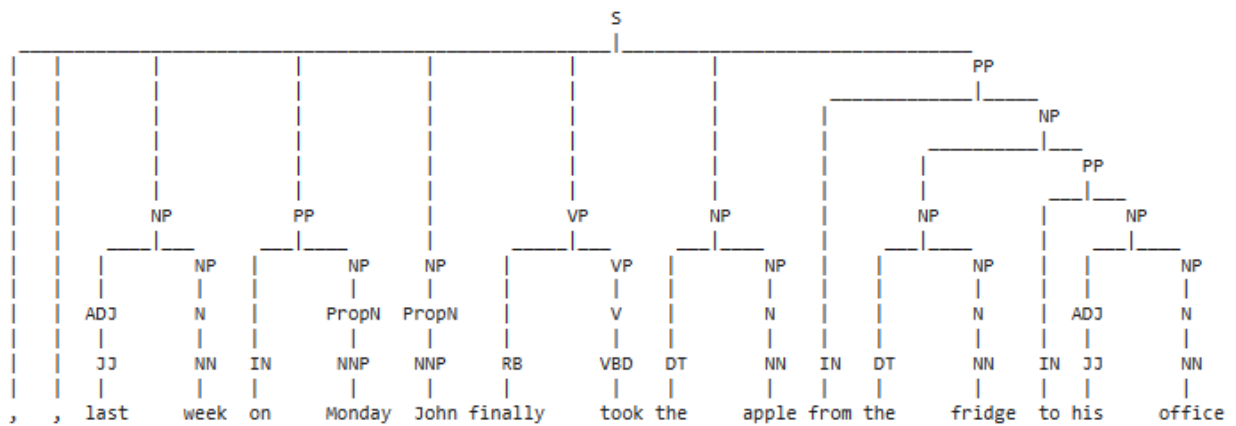
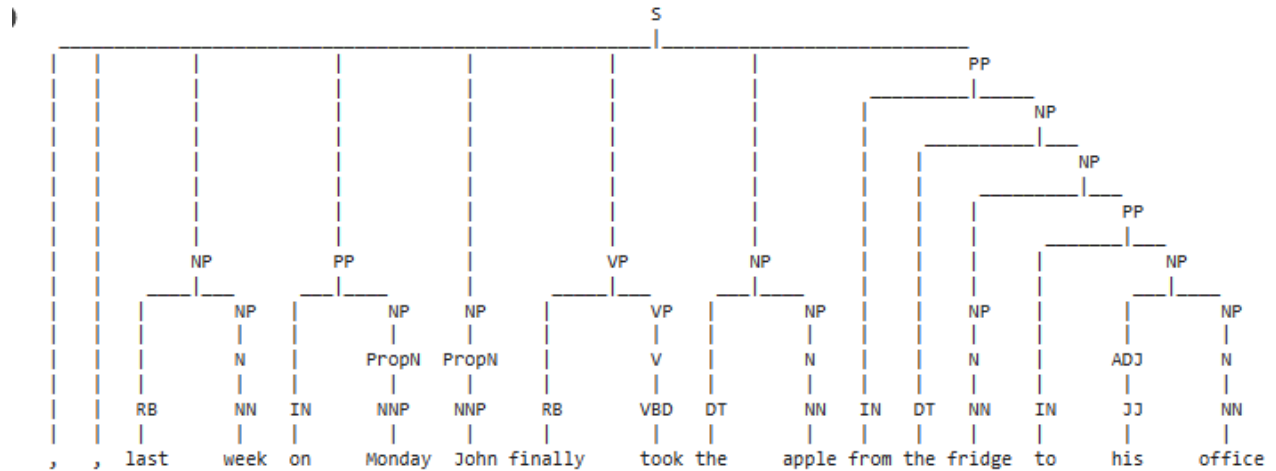








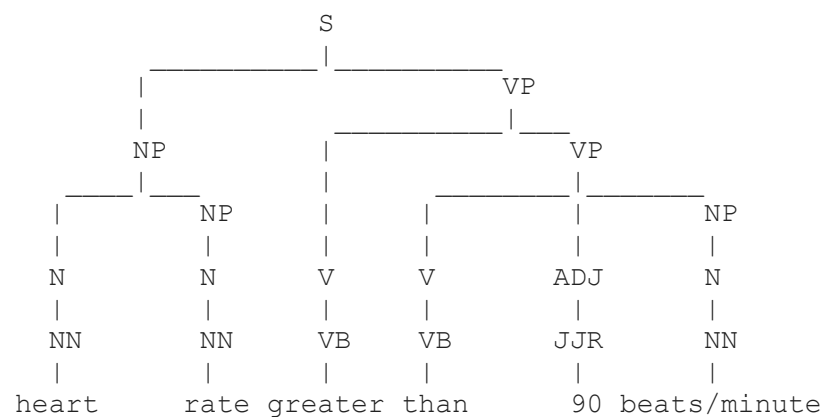
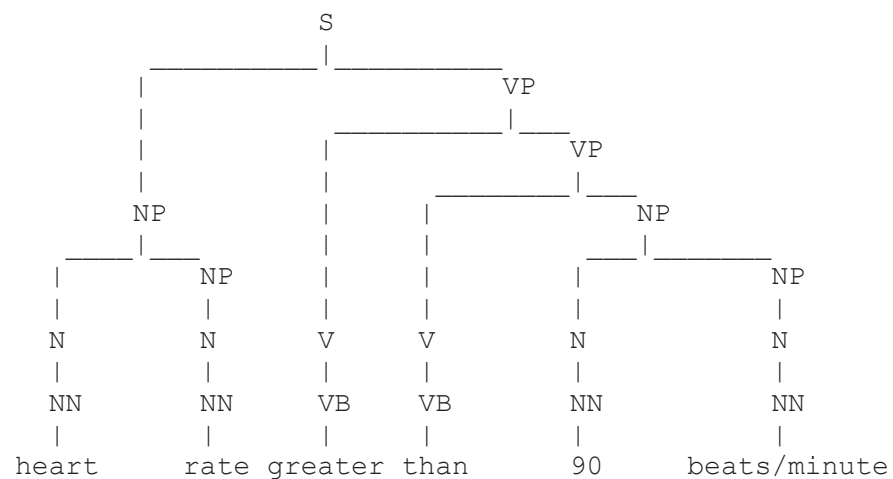
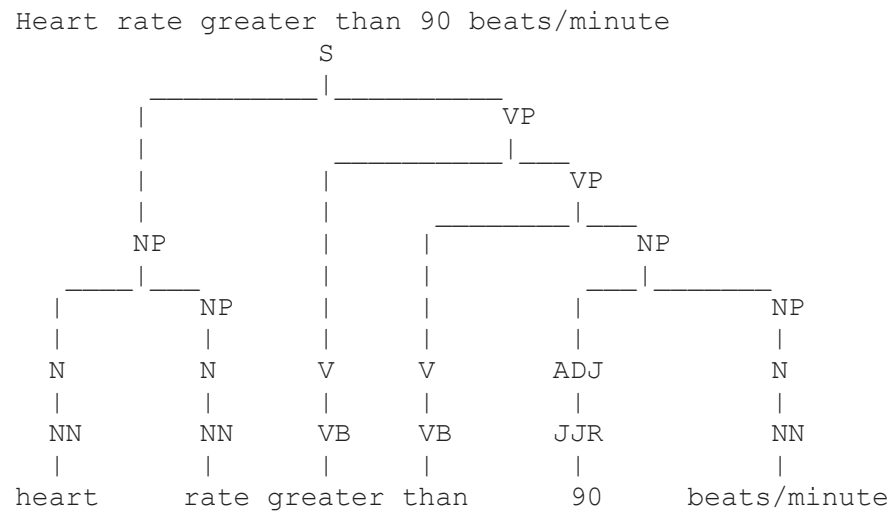




## 1.5 Test Scenario 5

Input: Heart rate greater than 90 beats/minute

Result: The proofreading results and text preprocess pipeline can be delivered, and the programme can operate correctly. Earley parse process results are displayed below.

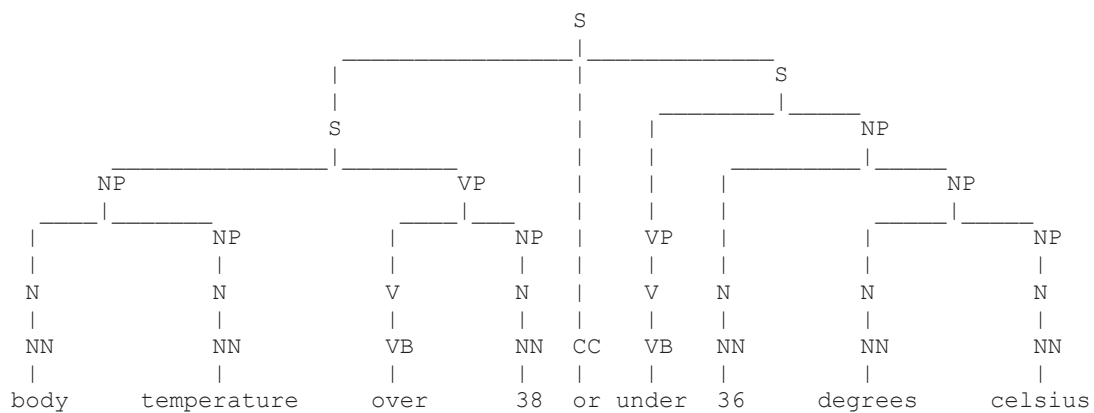
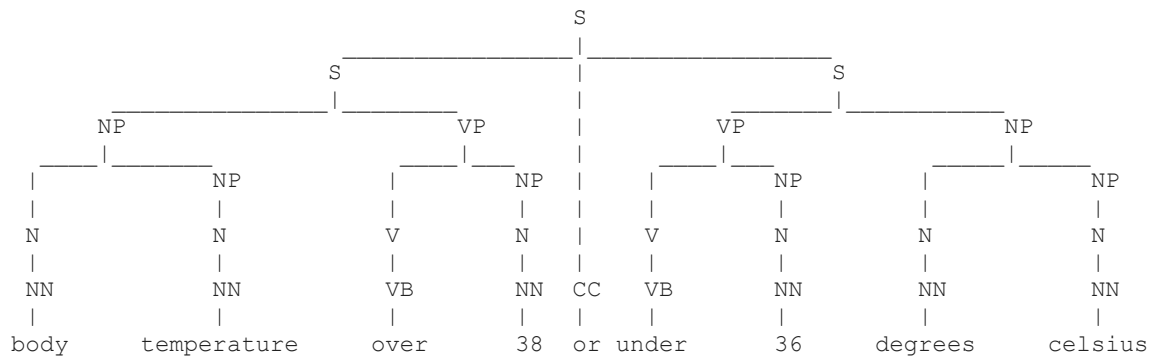
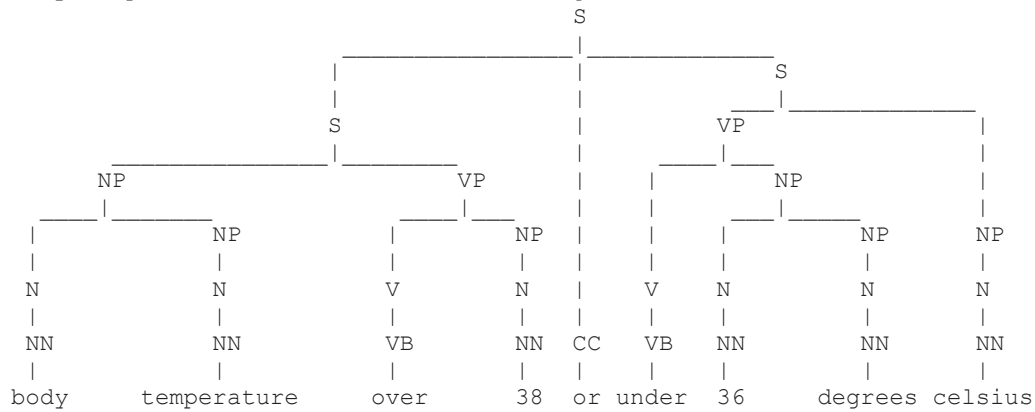


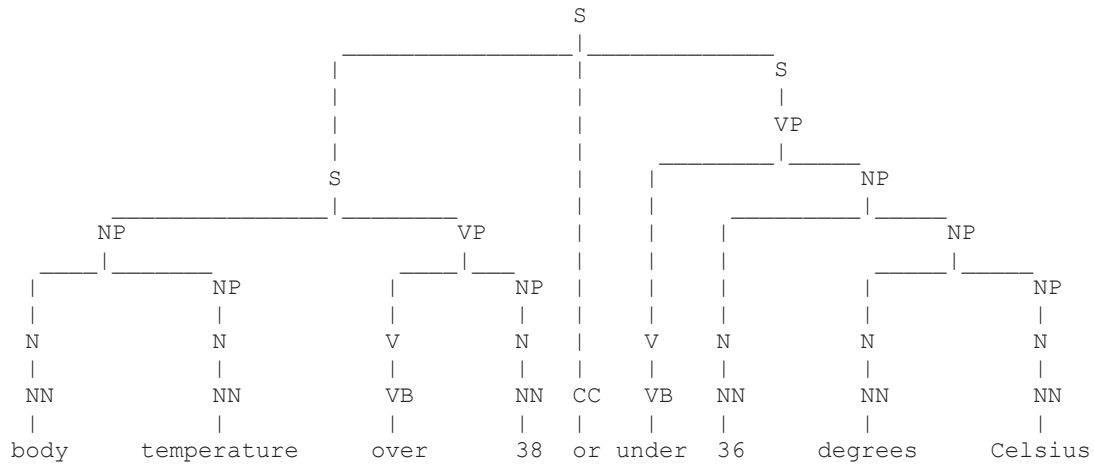
## 1.6 Test Scenario 6

Input: Body temperature over 38 or under 36 degrees Celsius.

Result: The proofreading results and text preprocess pipeline can be delivered, and the programme can operate correctly. Earley parse process results are displayed below.

Body temperature over 38 or under 36 degrees Celsius

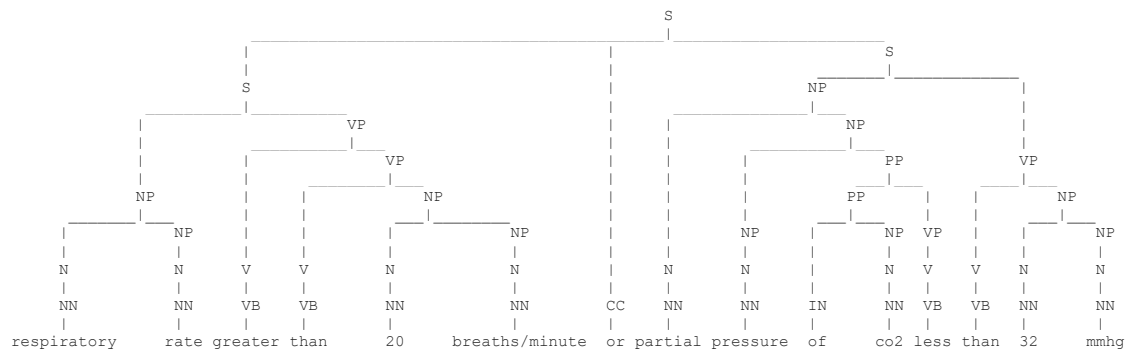
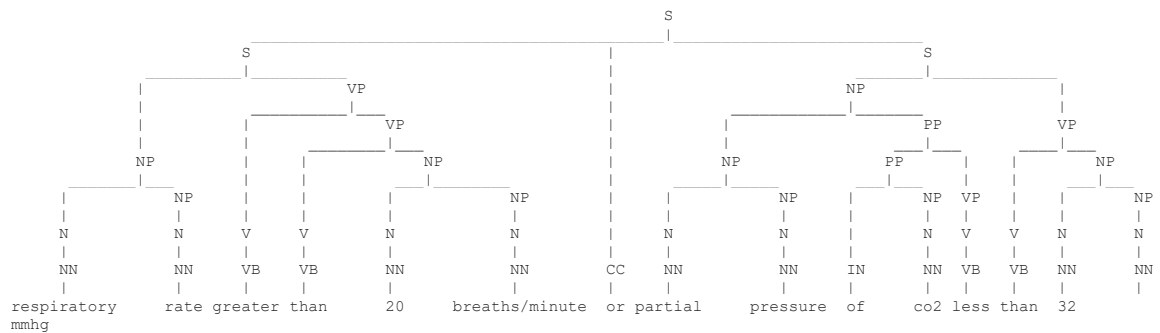




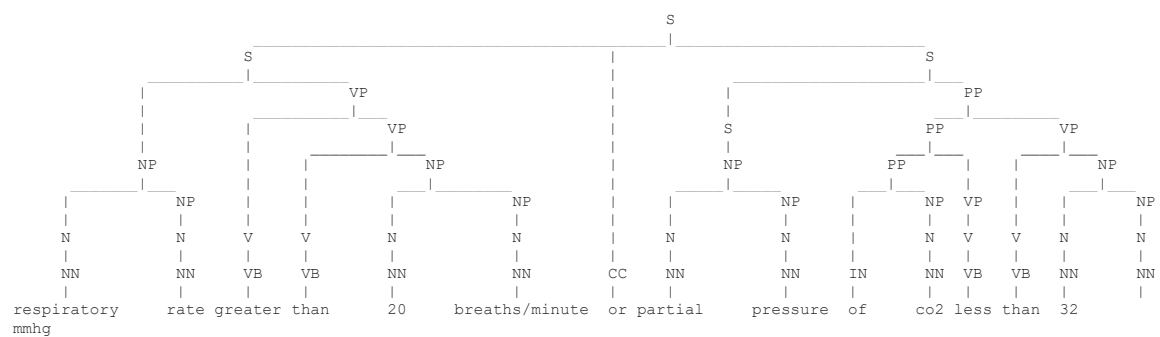
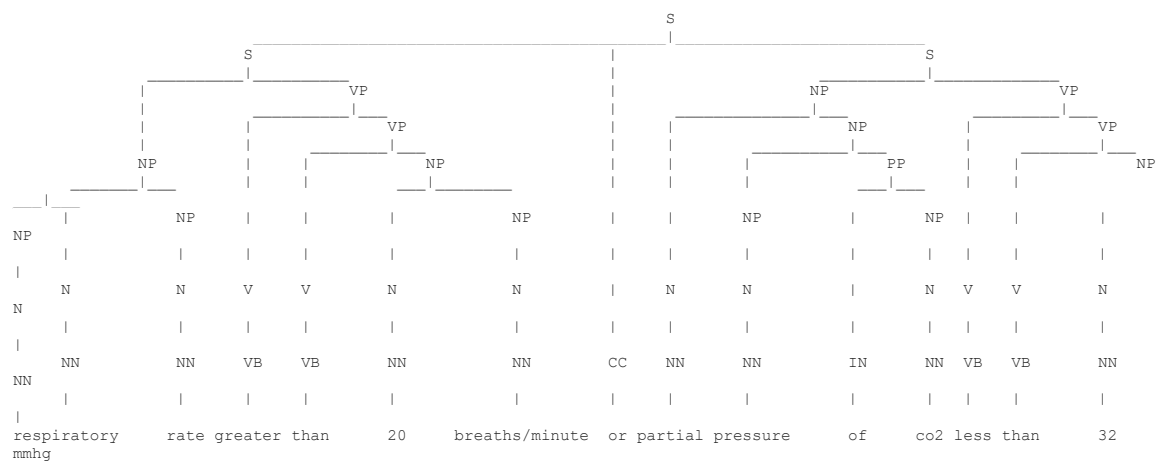
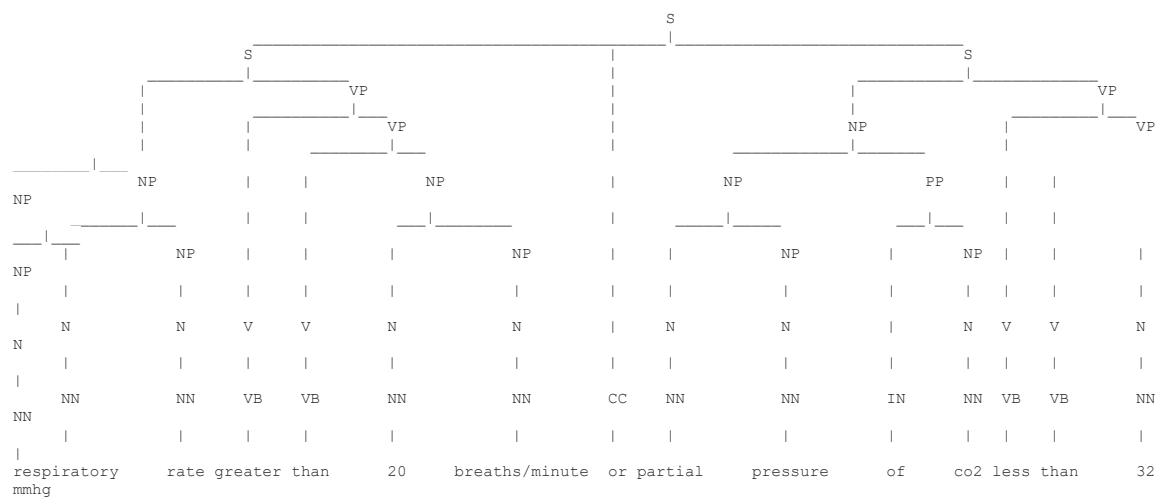
### 1.7 Test Scenario 7

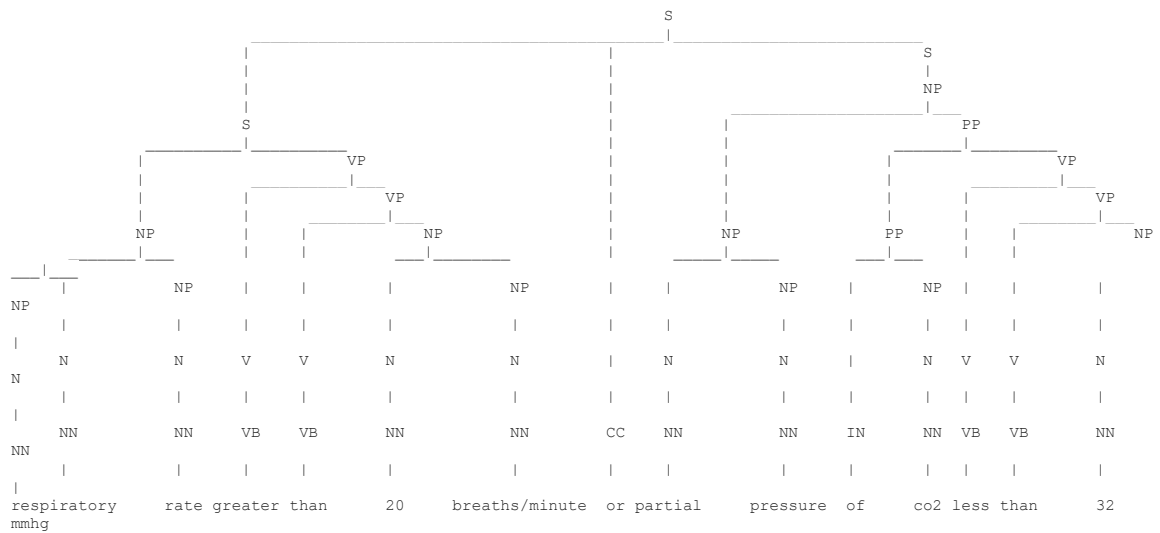
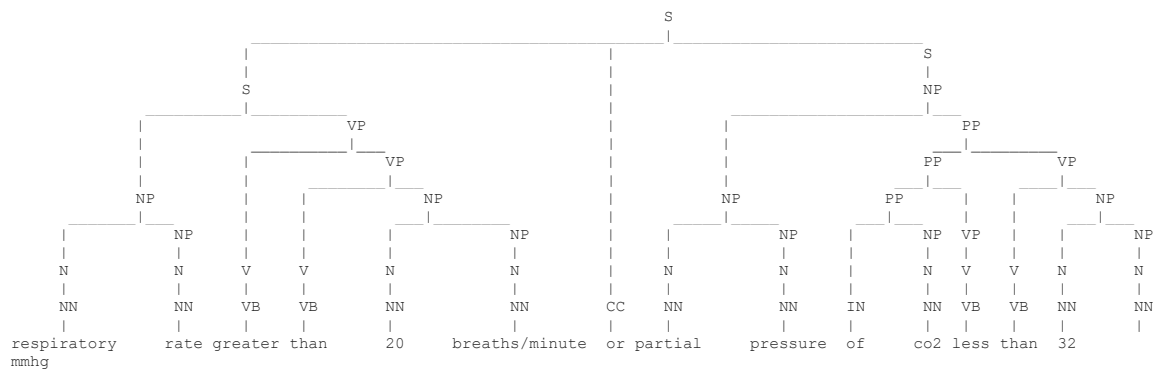
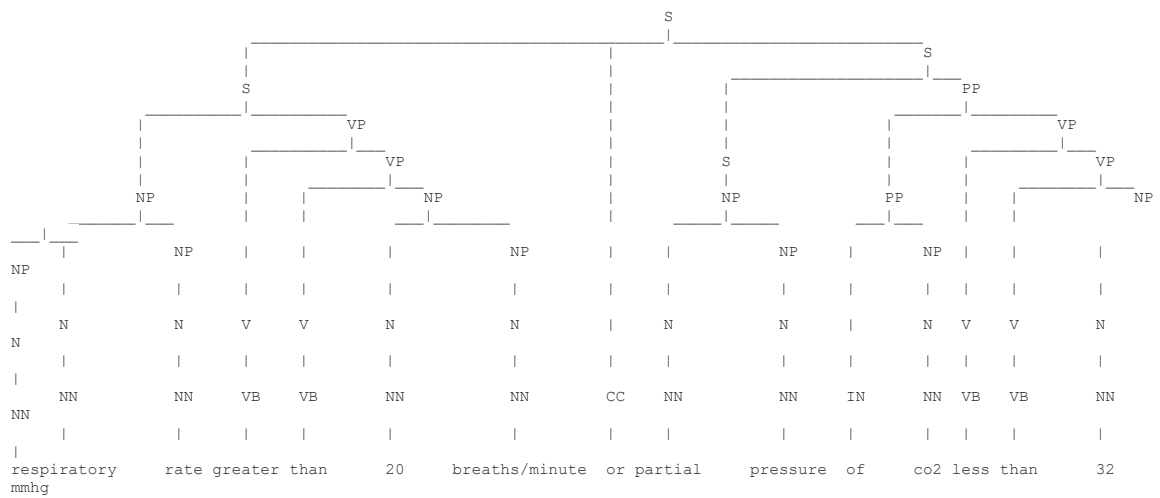
Input: Respiratory rate greater than 20 breaths/minute or partial pressure of CO2 less than 32 mmHg

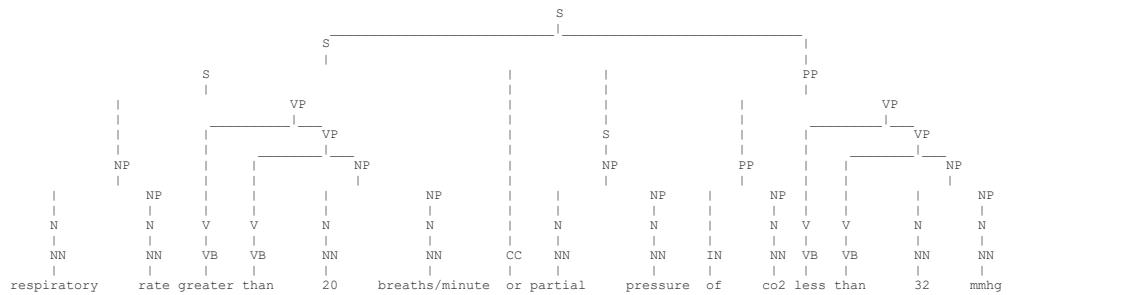
Result: The proofreading results and text preprocess pipeline can be delivered, and the programme can operate correctly. Earley parse process results are displayed below.

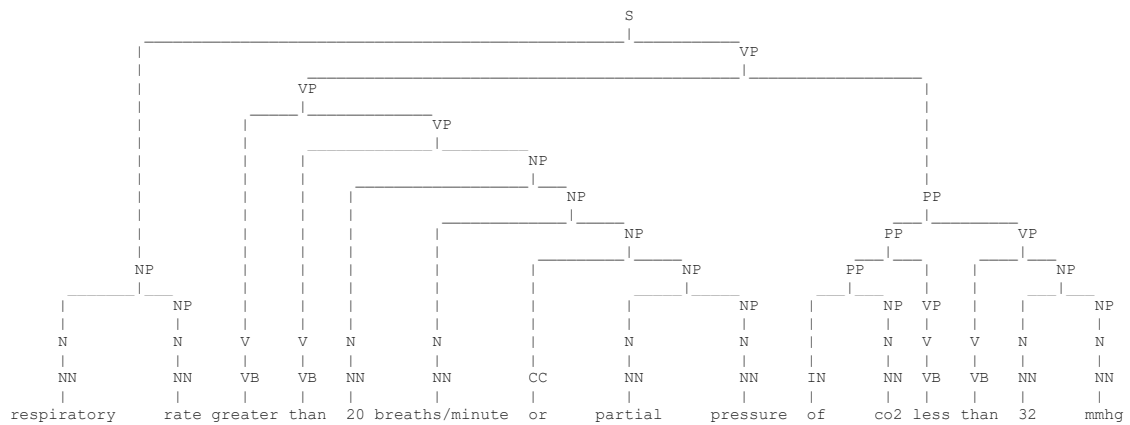
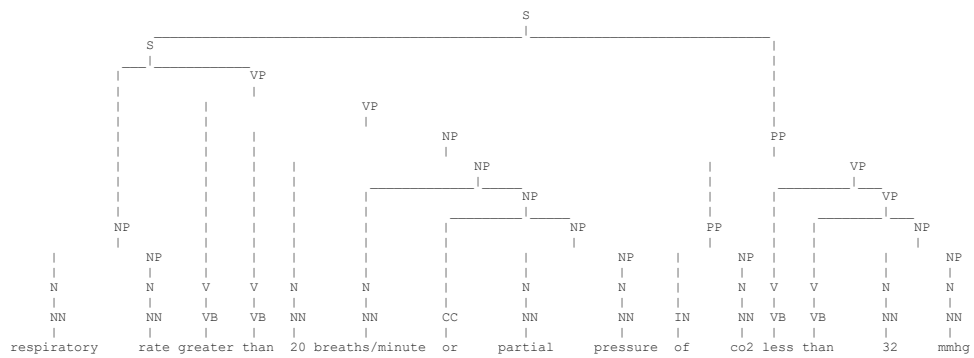
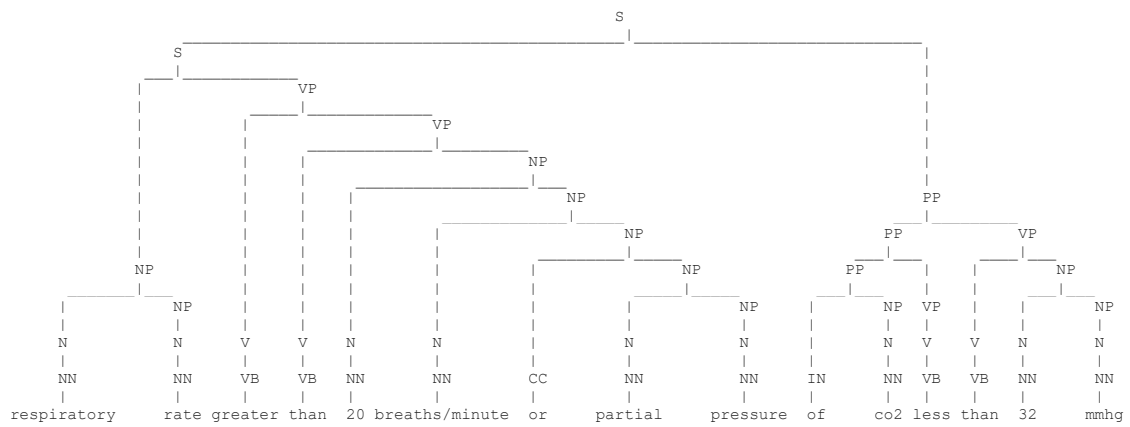


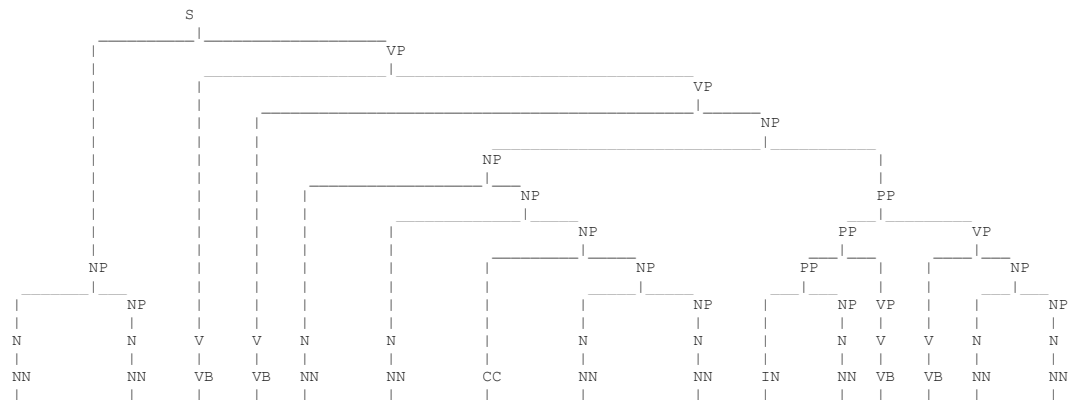
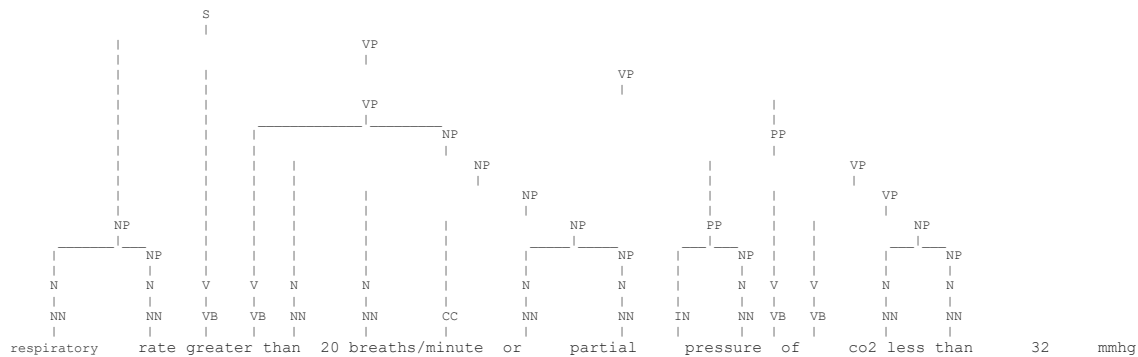
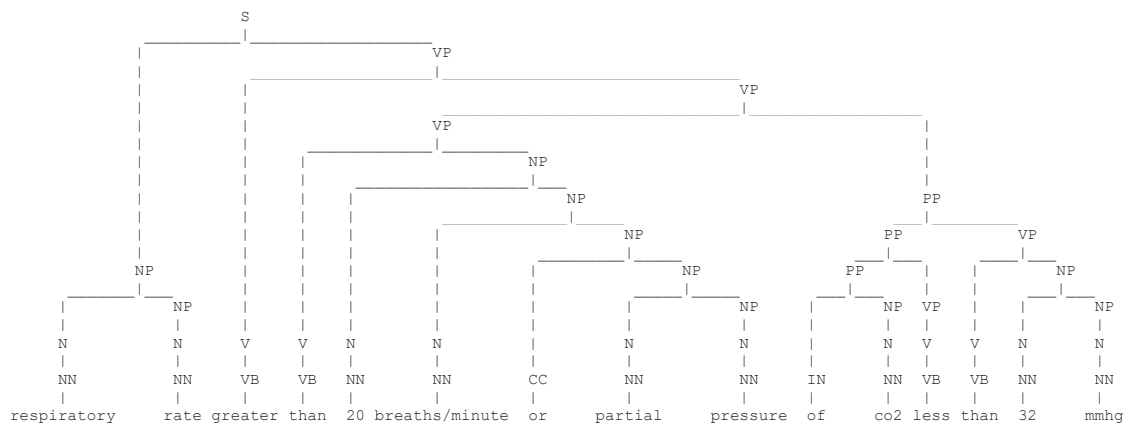
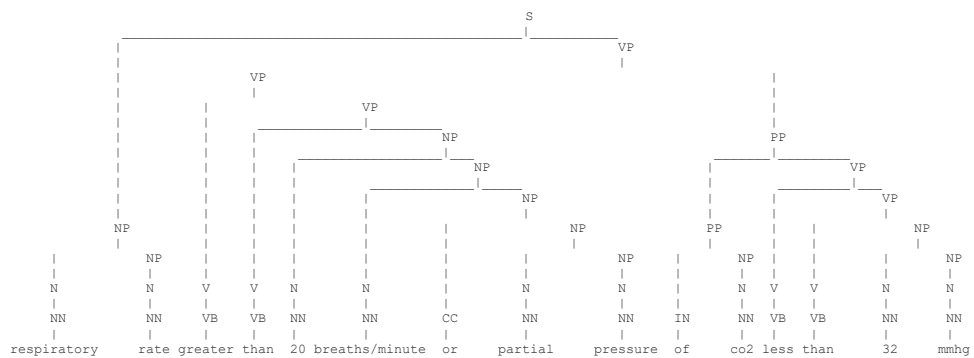




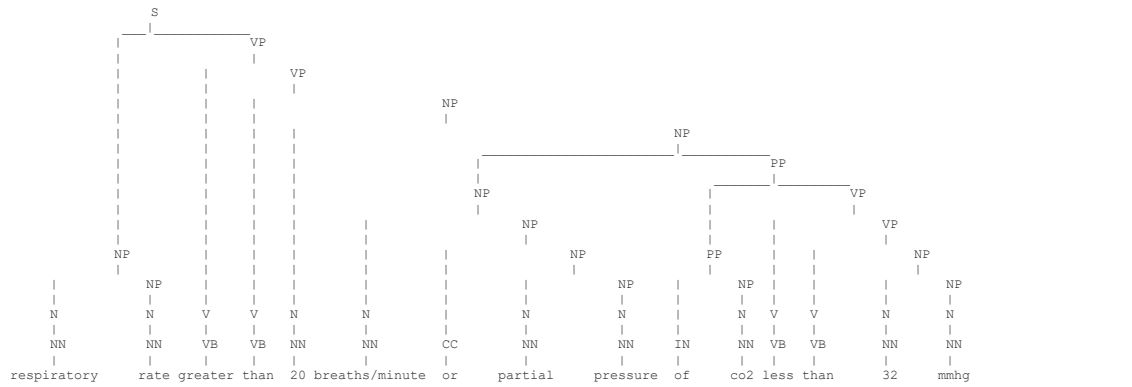
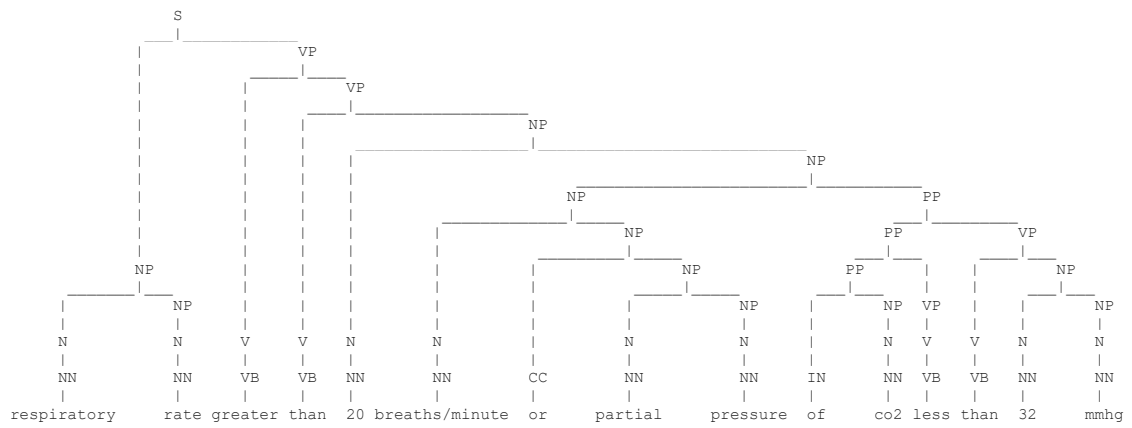
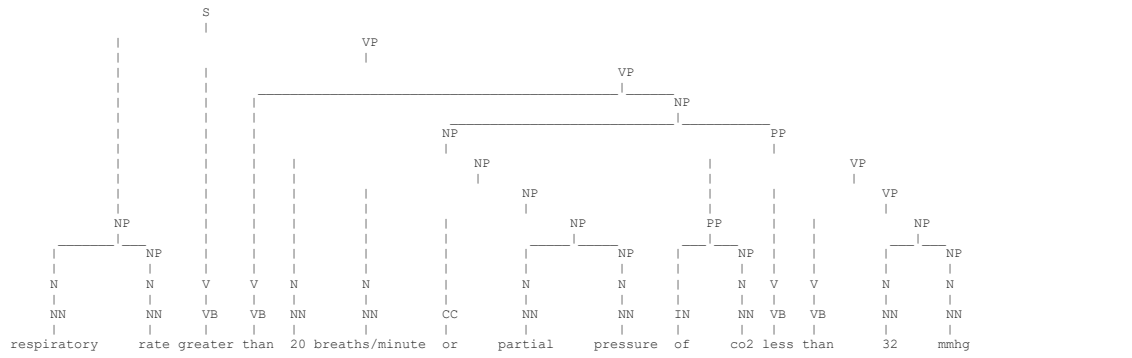


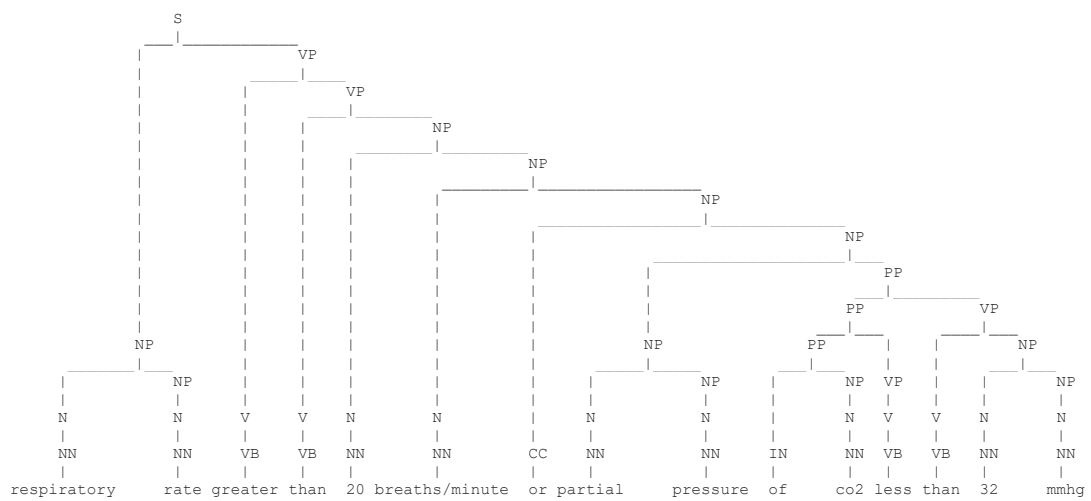
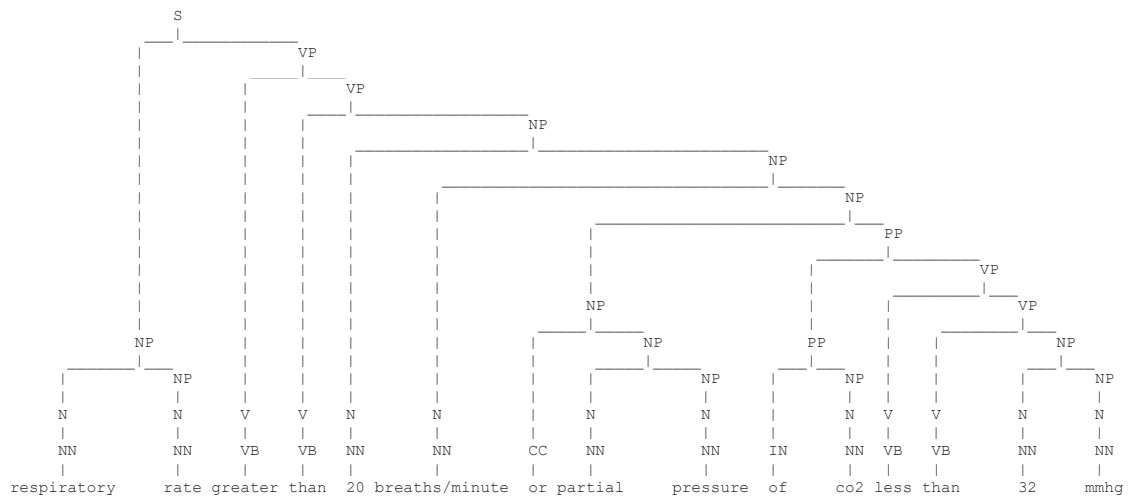
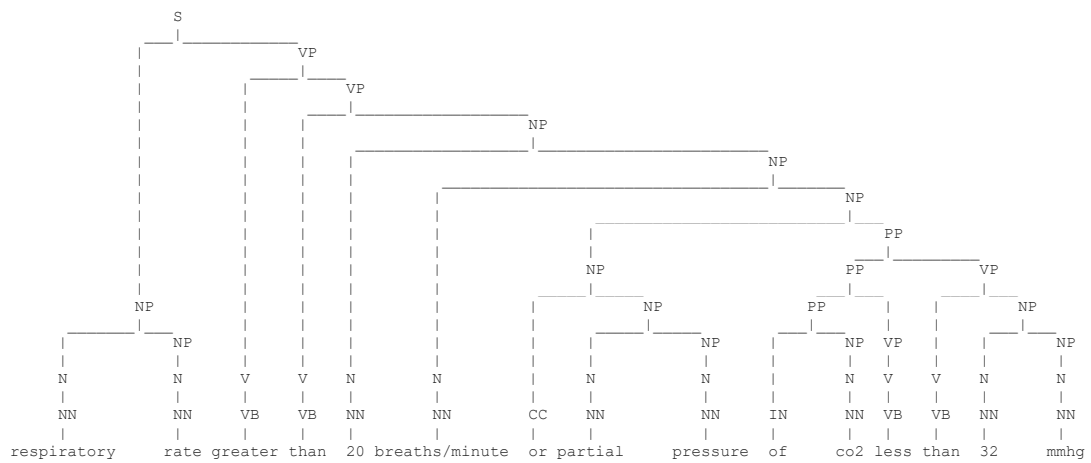


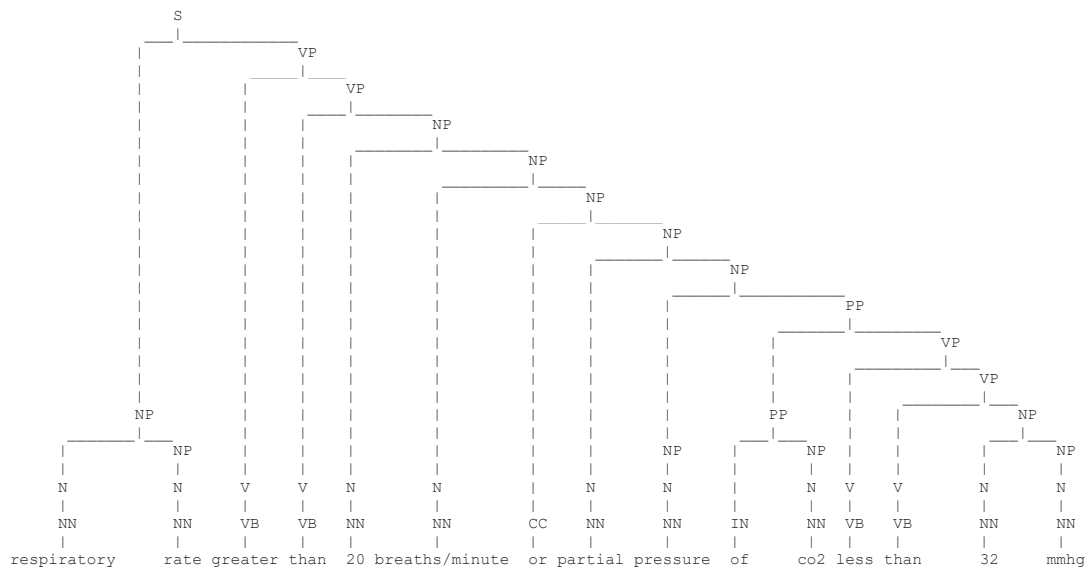
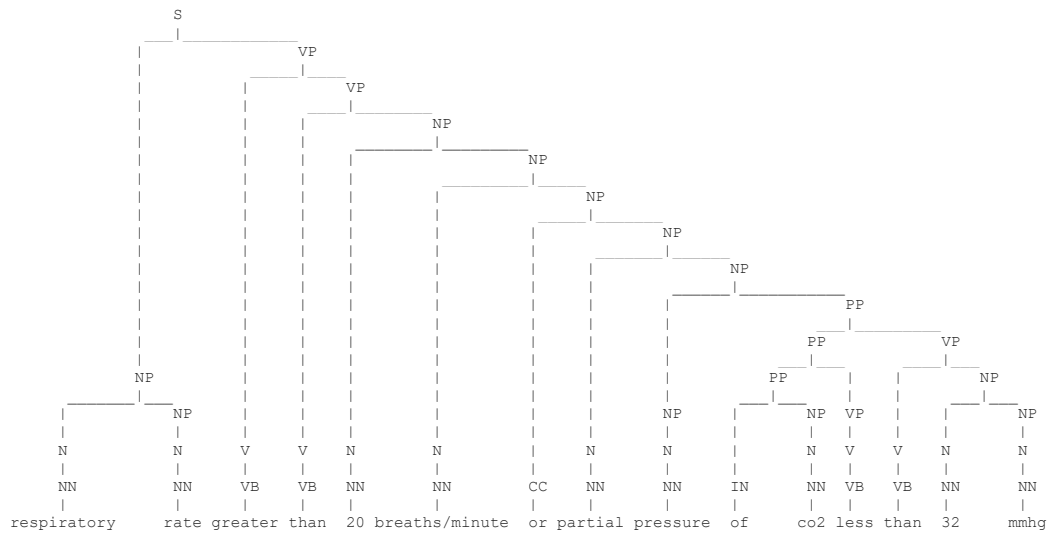
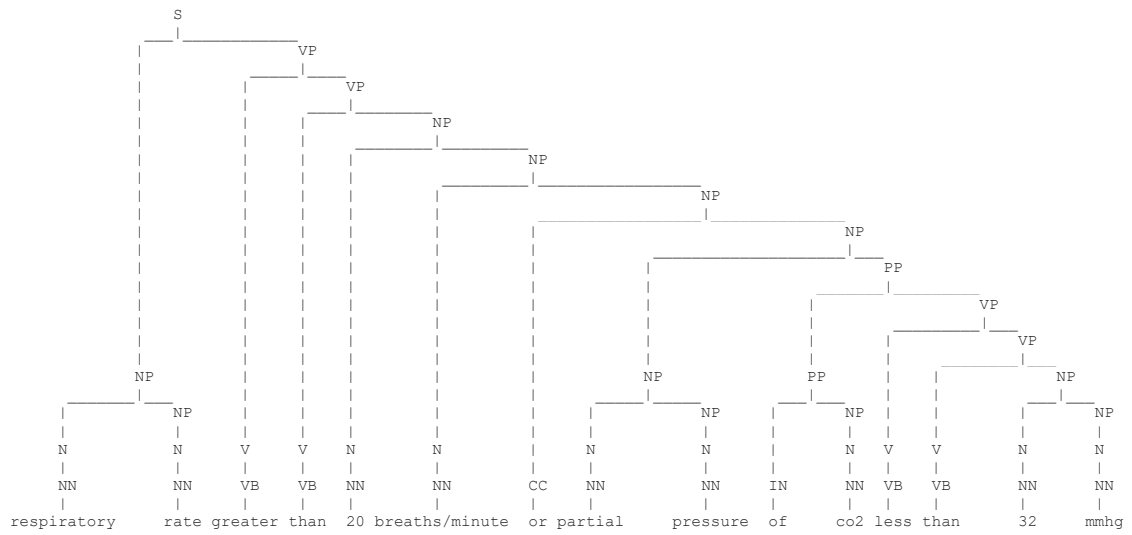




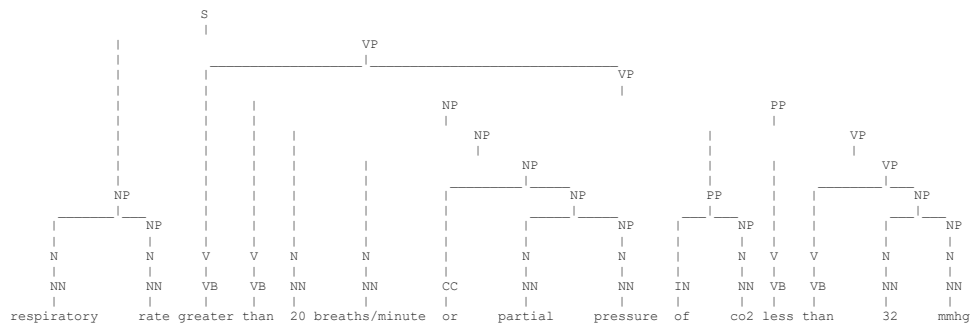
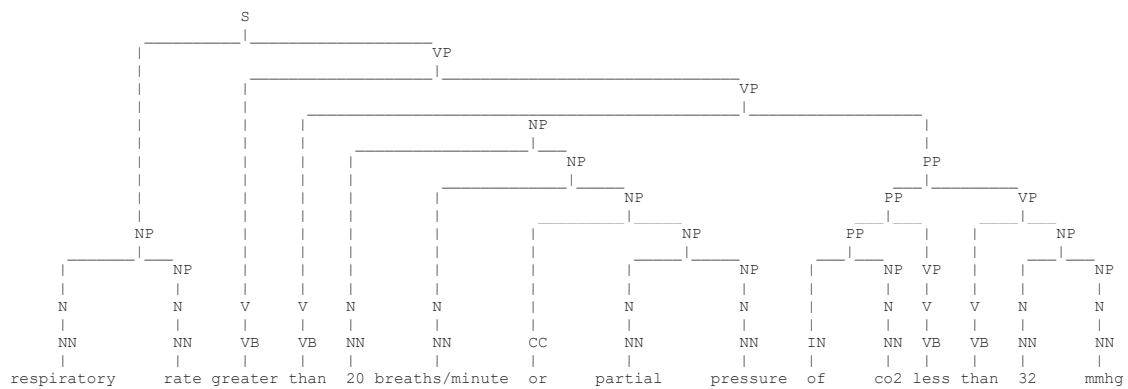
respiratory rate greater than 20 breaths/minute or partial pressure of co2 less than 32 mmhg











## 2. Outputs in CoNLL Format

For getting the output of CFG in CoNLL format, I wrote the below code,

```
# Parse and annotate sentences
for sentence in sentences:
    # Process the sentence with spaCy
    doc = nlp(sentence)

    # Extract POS and NE information and annotate with Ent and MeasEnt
    token_index = 1
    for token in doc:
        word = token.text
        pos = token.pos_
        ne = token.ent_type_ # Named Entity (NE) label
        ent = "Ent" # Entity Type (customize based on your data)
```

```

meas_ent = "MeasEnt" # Measured Entity (customize based on your
data)

# Format and print each token's information in CoNLL format
print(f"{token_index}\t{word}\t_{t}{pos}\t{ne}\t{ent}\t{meas_ent}")
token_index += 1

# Print an empty line to separate sentences
print()

```

The corresponding output in CoNLL format that I get are given below,

John ate the apple at the table.

1	John	—	PROPN PERSON	Ent	MeasEnt
2	ate	—	VERB	Ent	MeasEnt
3	the	—	DET	Ent	MeasEnt
4	apple	—	NOUN	Ent	MeasEnt
5	at	—	ADP	Ent	MeasEnt
6	the	—	DET	Ent	MeasEnt
7	table	—	NOUN	Ent	MeasEnt
8	.	—	PUNCT	Ent	MeasEnt

John, who ate the apple from the fridge, was sick Wednesday.

1	John	—	PROPN PERSON	Ent	MeasEnt
2	,	—	PUNCT	Ent	MeasEnt
3	who	—	PRON	Ent	MeasEnt
4	ate	—	VERB	Ent	MeasEnt
5	the	—	DET	Ent	MeasEnt
6	apple	—	NOUN	Ent	MeasEnt
7	from	—	ADP	Ent	MeasEnt
8	the	—	DET	Ent	MeasEnt
9	fridge	—	NOUN	Ent	MeasEnt
10	,	—	PUNCT	Ent	MeasEnt
11	was	—	AUX	Ent	MeasEnt
12	sick	—	ADJ	Ent	MeasEnt
13	Wednesday	—	PROPN DATE	Ent	MeasEnt
14	.	—	PUNCT	Ent	MeasEnt

On Monday, John ate the apple in his office.

1	On	ADP	Ent	MeasEnt
2	Monday	PROPN DATE	Ent	MeasEnt
3	,	PUNCT	Ent	MeasEnt
4	John	PROPN PERSON	Ent	MeasEnt
5	ate	VERB	Ent	MeasEnt
6	the	DET	Ent	MeasEnt
7	apple	NOUN	Ent	MeasEnt
8	in	ADP	Ent	MeasEnt
9	his	PRON	Ent	MeasEnt
10	office	NOUN	Ent	MeasEnt
11	.	PUNCT	Ent	MeasEnt

Last week, on Monday, John finally took the apple from the fridge to his office.

1	Last	ADJ DATE	Ent	MeasEnt
2	week	NOUN DATE	Ent	MeasEnt
3	,	PUNCT	Ent	MeasEnt
4	on	ADP	Ent	MeasEnt
5	Monday	PROPN DATE	Ent	MeasEnt
6	,	PUNCT	Ent	MeasEnt
7	John	PROPN PERSON	Ent	MeasEnt
8	finally	ADV	Ent	MeasEnt
9	took	VERB	Ent	MeasEnt
10	the	DET	Ent	MeasEnt
11	apple	NOUN	Ent	MeasEnt
12	from	ADP	Ent	MeasEnt
13	the	DET	Ent	MeasEnt
14	fridge	NOUN	Ent	MeasEnt
15	to	ADP	Ent	MeasEnt
16	his	PRON	Ent	MeasEnt
17	office	NOUN	Ent	MeasEnt
18	.	PUNCT	Ent	MeasEnt

Heart rate greater than 90 beats/minute

1	Heart	NOUN	Ent	MeasEnt
2	rate	NOUN	Ent	MeasEnt
3	greater	ADJ	Ent	MeasEnt
4	than	ADP	Ent	MeasEnt
5	90	NUM	Ent	MeasEnt
6	beats	NOUN	Ent	MeasEnt
7	/	SYM	Ent	MeasEnt
8	minute	NOUN	Ent	MeasEnt

Body temperature over 38 or under 36 degrees Celsius.

1	Body	_	NOUN		Ent	MeasEnt	
2	temperature	_		NOUN		Ent	MeasEnt
3	over	_	ADP		Ent	MeasEnt	
4	38	_	NUM	DATE	Ent	MeasEnt	
5	or	_	CCONJ		Ent	MeasEnt	
6	under	_	ADP		Ent	MeasEnt	
7	36	_	NUM	QUANTITY		Ent	MeasEnt
8	degrees	_	NOUN	QUANTITY		Ent	MeasEnt
9	Celsius	_	NOUN	QUANTITY		Ent	MeasEnt
10	.	_	PUNCT		Ent	MeasEnt	

Respiratory rate greater than 20 breaths/minute or partial pressure of CO2 less than 32 mmHg

1	Respiratory	_		ADJ		Ent	MeasEnt
2	rate	_	NOUN		Ent	MeasEnt	
3	greater	_	ADJ	CARDINAL		Ent	MeasEnt
4	than	_	ADP	CARDINAL		Ent	MeasEnt
5	20	_	NUM	CARDINAL		Ent	MeasEnt
6	breaths	_	NOUN		Ent	MeasEnt	
7	/	_	SYM		Ent	MeasEnt	
8	minute	_	NOUN		Ent	MeasEnt	
9	or	_	CCONJ		Ent	MeasEnt	
10	partial	_	ADJ		Ent	MeasEnt	
11	pressure	_	NOUN		Ent	MeasEnt	
12	of	_	ADP		Ent	MeasEnt	
13	CO2	_	PROPN	CARDINAL		Ent	MeasEnt
14	less	_	ADJ		Ent	MeasEnt	
15	than	_	ADP		Ent	MeasEnt	
16	32	_	NUM	CARDINAL		Ent	MeasEnt
17	mmHg	_	NOUN		Ent	MeasEnt	

Leukocyte count greater than 12000 or less than 4000 /microliters or over 10% immature forms or bands.

1	Leukocyte	_	PROPN	ORG	Ent	MeasEnt	
2	count	_	VERB		Ent	MeasEnt	
3	greater	_	ADJ		Ent	MeasEnt	
4	than	_	ADP		Ent	MeasEnt	
5	12000	_	NUM	DATE	Ent	MeasEnt	
6	or	_	CCONJ		Ent	MeasEnt	
7	less	_	ADJ	CARDINAL		Ent	MeasEnt
8	than	_	ADP	CARDINAL		Ent	MeasEnt
9	4000	_	NUM	CARDINAL		Ent	MeasEnt
10	/microliters	_		NOUN		Ent	MeasEnt
11	or	_	CCONJ		Ent	MeasEnt	
12	over	_	ADP	PERCENT	Ent	MeasEnt	
13	10	_	NUM	PERCENT	Ent	MeasEnt	
14	%	_	NOUN	PERCENT	Ent	MeasEnt	
15	immature	_	ADJ		Ent	MeasEnt	
16	forms	_	NOUN		Ent	MeasEnt	
17	or	_	CCONJ		Ent	MeasEnt	
18	bands	_	NOUN		Ent	MeasEnt	
19	.	_	PUNCT		Ent	MeasEnt	

### **3. Limitations**

The CFG grammar I developed for this project mostly work for all the validation data in terms of Earley parse output and CoNLL format output. Yet, the 8<sup>th</sup> sentence which was “Leukocyte count greater than 12000 or less than 4000 /microliters or over 10% immature forms or bands.”, it gave error while producing the Earley parse output. This is because some of the word lexicons in this particular sentence are beyond the capacity of my CFG grammar. Probably in the next project I will attempt to fix it as an extension of this current project.