Gal Cesana 318510633
*Authorization has been granted to submit the assignment individually

Natural Language Processing – Exercise 5

(a) Compare the performance of the models. Explain why this happens.

Performance Metrics:

Model	Donald Trump	Ruth Bader Ginsburg	J. K. Rowling
POS-Based	959 triplets	868 triplets	739 triplets
Dependency-Based	99 triplets	76 triplets	53 triplets
LLM-Based	36 triplets	36 triplets	83 triplets

Observations:

1. **POS-Based Extractor:**
    - Extracts the highest number of triplets.
    - Many extracted triplets are incorrect or nonsensical.
    - Relies solely on part-of-speech tagging, leading to high noise and irrelevant relations.
2. **Dependency-Based Extractor:**
    - Extracts fewer triplets but with higher accuracy.
    - Captures meaningful relationships but misses some valid triplets due to strict syntactic rules.
3. **LLM-Based Extractor:**
    - Produces fewer triplets than the POS-based approach but has higher quality.
    - Extracts contextually relevant triplets but sometimes hallucinates incorrect information.

Overall Comparison:

Model	Accuracy	Recall (Number of Triplets)	Strengths	Weaknesses
POS-Based	❌ Low	☑ Very High	Extracts many triplets	High noise, many incorrect relations
Dependency-Based	☑ Medium	⚠ Low	More accurate triplets	Misses some valid triplets
LLM-Based	☑ High	⚠ Medium	Context-aware, better relations	Can hallucinate incorrect information

(b) Two Limitations of Using an LLM as a Verifier

1. LLMs Can Introduce Hallucinations

- The LLM might **confirm false relations** if they "sound plausible" based on general knowledge.
- **Example:**
  - **Triplet:** ("Donald Trump", "born in", "Los Angeles")
  - **LLM Response:** ✅ "Yes, this is correct" → ❌ Incorrect!
- LLMs rely on **statistical patterns** and might **not fact-check** against the given document.

2. Lack of Justification & Explainability

- If the model **accepts or rejects** a triplet, it **does not explain why**.
- **Example:**
  - **Triplet:** ("J.K. Rowling", "married", "Pete Rowling")
  - LLM: ❌ "False" → But **why is it false?**
- **Problem:** Human evaluators need reasoning, and an LLM **does not provide citations** or supporting evidence.

(c) Two Limitations of Comparing Extracted Triplets to a Human-Annotated List

1. Coverage Issue – Human List is Not Exhaustive

- The human-extracted list **might not contain all valid triplets**.
- **Example:**
    - Extracted Triplet: ("Ruth Bader Ginsburg", "appointed by", "Bill Clinton")
    - Human List: **Does not contain this triplet.**
    - **Problem:** The triplet is actually **correct**, but the comparison **falsely classifies it as wrong**.

2. Exact Matching Can Be Too Strict

- Human-extracted triplets might **use different wording** than extracted ones.
- **Example:**
    - Extracted Triplet: ("Trump", "was the 45th president of", "USA")
    - Human List: ("Donald Trump", "served as", "45th president of the United States")
    - **Issue:** Even though they mean the same thing, the **exact match test** would **fail**.
- **Problem:** Semantic equivalence is not considered, leading to false negatives.

Appendix A - Python script outputs:

Page: Donald Trump

Total POS-Based Triplets: 959

Total Dependency-Based Triplets: 99

Sample POS-Based Triplets: [('Floyd', 'removed of from', 'Lafayette'), ('Trump', 'polarized approved by of', 'Republicans'), ('Americans', 'go to came from', 'House'), ('Organization', 'of inflating gain from', 'Carroll'), ('Trump', 'dictated of of', 'Trump')]

Sample Dependency-Based Triplets: [('Trump', 'posed at', 'Episcopal Church'), ('House', 'impeached', 'Trump'), ('Trump', 'withdrew from', 'Paris Agreement'), ('=', 'opened', 'Harrah'), ('Trump', 'agreed In', 'September')]

Page: Ruth Bader Ginsburg

Total POS-Based Triplets: 868

Total Dependency-Based Triplets: 76

Sample POS-Based Triplets: [('Wade', 'decided', 'Ginsburg'), ('Mentally', 'on of losing During with', 'Emily'), ('Carolina', 'of', 'Persons'), ('Project', 'on', 'International'), ('Hammer', 'as', 'Marty')]

Sample Dependency-Based Triplets: [('Ginsburg', 'dissented in', 'Wagnon'), ('Opperman Foundation', 'established', 'Award'), ('Deadpool', 'considers for', '-'), ('Ginsburg', 'went In', 'January'), ('Ginsburg', 'called on', 'Congress')]

Page: J. K. Rowling

Total POS-Based Triplets: 739

Total Dependency-Based Triplets: 53

Sample POS-Based Triplets: [('Kloves', 'wrote for with', 'Rowling'), ('£', 'allowed move to in', 'Leith'), ('Potter', 'contribute revised in resemble of', 'Harry'), ('Quirrell', 'opposes', 'Snape'), ('School', 'in', 'Winterbourne')]

Sample Dependency-Based Triplets: [('Forbes', 'named', 'Rowling'), ('Brown', 'published in', 'April'), ('Arantes', 'returned to', 'Portugal'), ('Arantes', 'arrived in', 'Scotland'), ('Emma Nicholson', 'founded with', 'MEP')]

Page: Donald Trump

Total LLM-Based Triplets: 36

Sample LLM-Based Triplets: [('Donald John Trump', 'previously served as the 45th', 'president from 2017 to 2021'), ('Republican Party', 'he is a member of the', 'political party'), ('New York City', 'Trump was born in', 'borough of Queens'), ('Jamaica Hospital', 'born at', 'Trump'), ('Fred Trump', 'is the son of', 'Trump')]

Page: Ruth Bader Ginsburg

Total LLM-Based Triplets: 36

Sample LLM-Based Triplets: [('Joan Ruth Bader', 'is', 'American lawyer and jurist'), ('Joan Ruth Bader', 'was', 'associate justice of the Supreme Court of the United States'), ('Joan Ruth Bader', 'was', 'nominated by President Bill Clinton'), ('Joan Ruth Bader', 'replaced', 'Byron White'), ('Joan Ruth Bader', 'was', 'first Jewish woman')]

Page: J. K. Rowling

Total LLM-Based Triplets: 83

Sample LLM-Based Triplets: [('Joanne Rowling', 'writes under the pen name', 'J. K. Rowling'), ('Joanne Rowling', 'also known as', 'Jo'), ('Joanne Rowling', 'married', 'Pete Rowling'), ('Anne Rowling', 'married', 'Pete Rowling'), ('David Gordon Rowling Murray', 'son of', 'Joanne Rowling')]
