

Introduction

- Large corpus of regulatory texts from European Commission are Difficult to read and interpret.
- Especially, Reporting Obligations, are one type of regulations that are crucial for many organization.
- Failure to comply the result in subsequent fines.
- In this work, we present a framework for transforming regulatory texts in reporting obligations into into structured format which enable to find out who was a reporter, to report what, to whom and by when.
- We proposed 2 approaches to extract addresser, addressee, action result, reported verb.

Dataset

- Eur-lex: <https://eur-lex.europa.eu/homepage.html>
- Language: English
- Size: 5000 texts



Figure 1. Dataset Source: Eur-Lex

Learning Reporting Verbs

- Unsupervised using Word2Vec language model.
- Cosine similarity for word vector distances

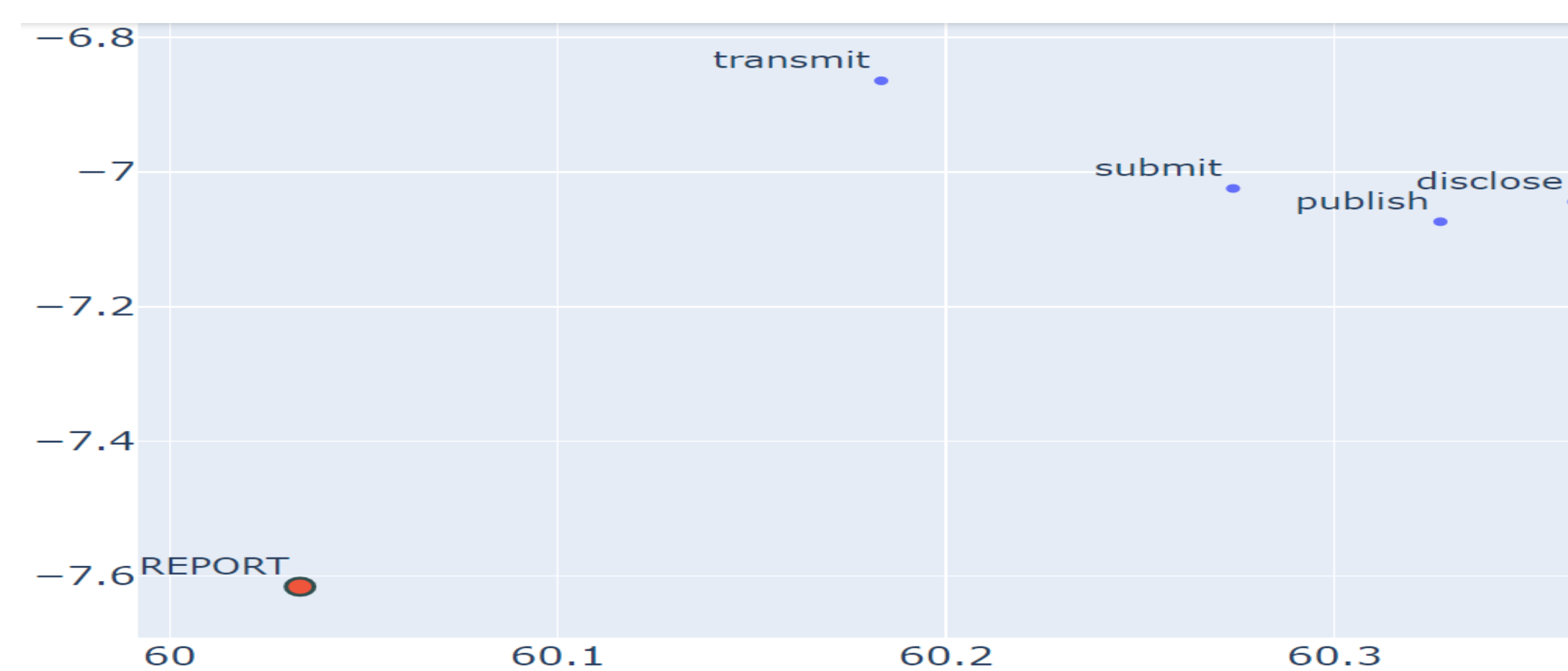


Figure 2. Word Vectors

Extracting Concepts based on Syntactic Structure

- Generate syntactic tree using dependency parser
- Parse tree to locate concepts based on syntactic roles
 - nsubj: addresser
 - dobj: addressee
 - pobj: action result
 - pobj: date
 - Handle conjunction, passive forms
- Extract concepts

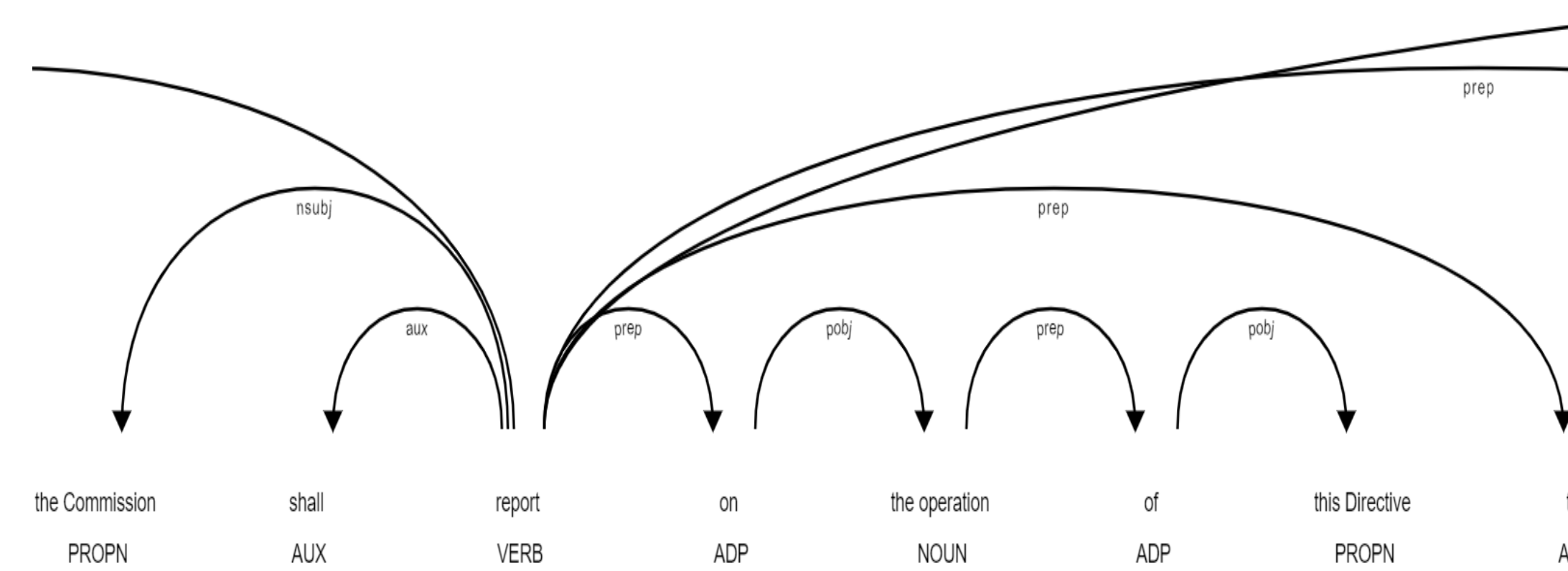


Figure 3. Syntactic Dependencies

By 5 September 2010, the Commission shall report on the operation of this Directive to the European Parliament and the Council, including on its impact on small and medium-sized issuers and on the application of sanctions, in particular whether ...

Figure 4. Extracted Concepts

Extracting Concepts using LLM

- Model: Llama3-8B-Instruct
- Design Specific Prompting
- Use few shot prompting



Figure 5. Llama LLM

LLM Prompts and Results

Sample prompts:

- LLM prompted to act as a Compliance Office
- Provided with a handful of examples for learning (few-shot)

```
You are a virtual annotation. For each sentence, you annotate the addresser, addressee,
actionResult, date. Here are some samples:
sentence: EU commission will report the result to the each member.
addresser: eu commission
Action: report
ActionResult: the result
addressee: the_each_member
Date: None
sentence: By 31 December 2010 and, thereafter, at least every three years, the Commission
shall review the provisions concerning its implementing powers and present a report to the
European Parliament and to the Council ...
addresser: the commission
Action: present
ActionResult: a report on the functioning of those powers
addressee: the_european_parliament, the_council
Date: By 31 December 2010
```

Figure 6. Example Prompts

Sample results:

- Concepts automatically detected by LLMs

```
assistant
Here are the annotations:
sentence: The Commission shall, no later than 5 September 2010, submit a report to the European
Parliament and to the Council on the ...
addresser: the Commission
Action: submit
ActionResult: a report
addressee: the_european_parliament, the_council
Date: no later than 5 September 2010
sentence: The result will be reported by EU Commission to the each member.
addresser: EU Commission
Action: report
ActionResult: the result
addressee: the_each_member
Date: None
```

Figure 7. Results