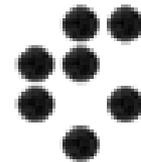


# Triplet Extraction from Sentences

Lorand Dali  
Blaž Fortuna

lorand.dali@ijs.si  
blaz.fortuna@ijs.si

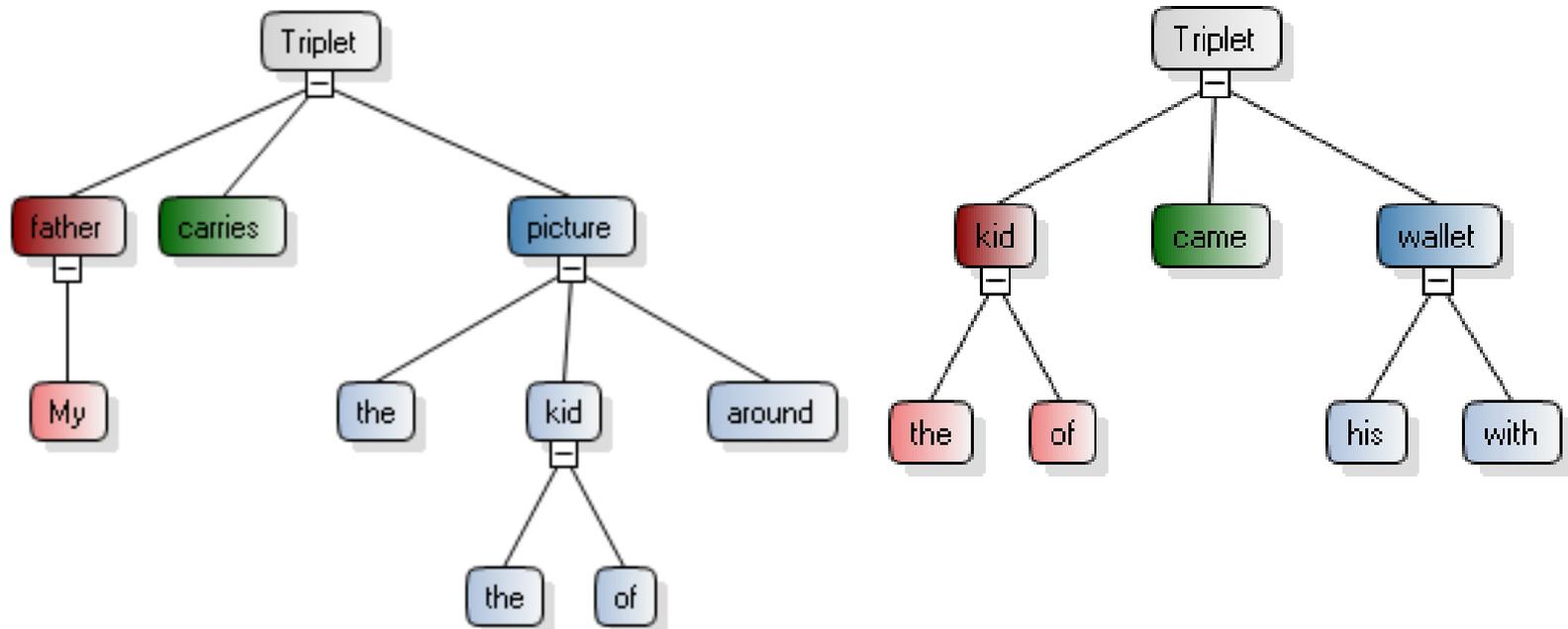
“Jožef Stefan” Institute, Ljubljana



**“Jožef Stefan”  
Institute**

17<sup>th</sup> of October 2008

My father carries around the picture of the kid who came with his wallet.



# Motivation of Triplet Extraction

- Advantages

- compact and simple representation of the information contained in a sentence
- avoids the complexity of a full parse
- contains semantic information

- Applications

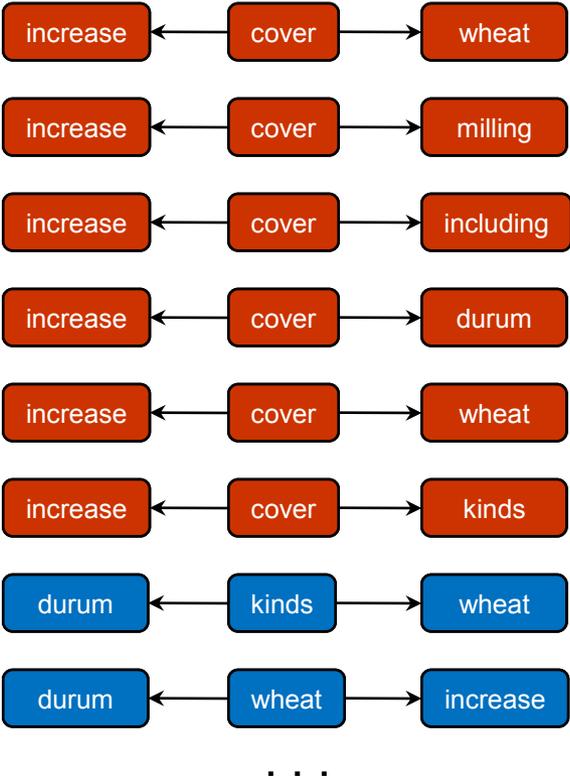
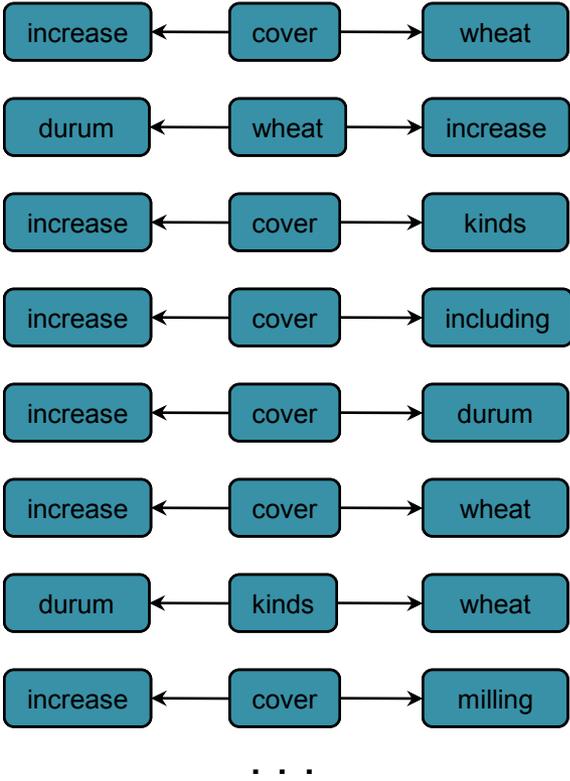
- building the semantic graph of a document
- summarization
- question answering

# Triplet Extraction – 2 Approaches

- Extraction from the parse tree of the sentence using heuristic rules
  - OpenNLP – Treebank Parsetree
  - Link Parser – Link Grammar (a type of dependency grammar)
- Extraction using Machine Learning
  - Support Vector Machines (SVM) are used
  - The SVM model is trained on human annotated data

The increase will cover all kinds of wheat including durum and milling wheat.

The increase will cover all kinds of wheat including durum and milling wheat .



# Merging Triplets

increase ← cover → wheat

increase ← cover → milling

increase ← cover → including

increase ← cover → durum

increase ← cover → wheat

increase ← cover → kinds

Merge triplets with same or adjacent words

increase ← cover → kinds wheat including durum milling wheat

Add stopwords

increase ← cover → all kinds of wheat including durum and milling wheat

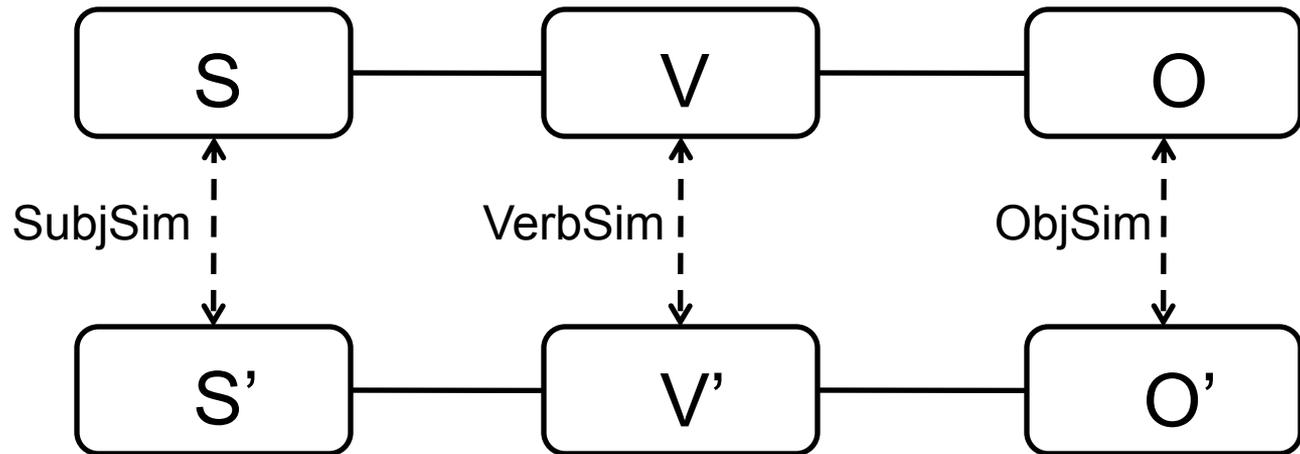
# Features of the triplet candidates

- Over 300 features depending on:
  - **Sentence**
    - length of sentence, number of words, etc
  - **Candidate**
    - context of Subj, Verb and Obj;
    - distance between Subj, Verb, Obj
  - **Linkage**
    - number of links, of link types, nr of links from S, V, O
  - **Minipar**
    - depth, diameter, siblings, uncles, cousins, categories, relations
  - **Treebank**
    - depth, diameter, siblings, uncles, cousins, path to root, POS

# Evaluation and Testing

- Training set = 700 annotated sentences  
Test set = 100 annotated sentences
- **Compare** the **extracted** triplets from a sentence to the **annotated** triplets from that same sentence
- Comparison is done according to a **similarity measure**  $\in [0, 1]$  between two triplets
- extracted to annotated => **precision**
- annotated to extracted => **recall**

# Triplet Similarity Measure



$$\text{TrSim} = (\text{SubjSim} + \text{VerbSim} + \text{ObjSim}) / 3$$

$\text{TrSim}, \text{SubjSim}, \text{VerbSim}, \text{ObjSim} \in [0, 1]$

# String Similarity Measure

The road to success is always under construction

road

success

under

construction

$$Sim = nMatch / maxLen = 3 / 5 = 0.6$$

way

success

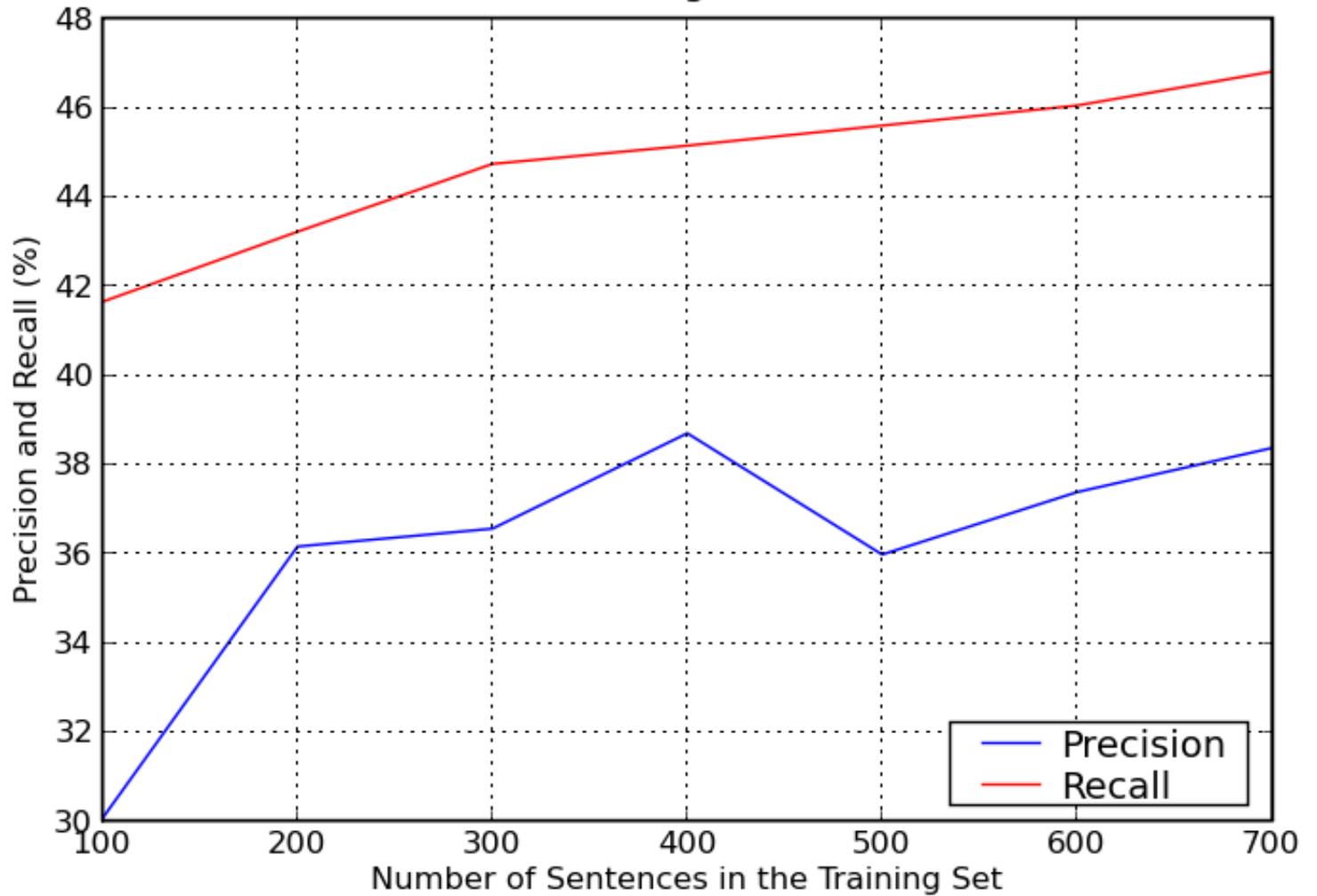
under

heavy

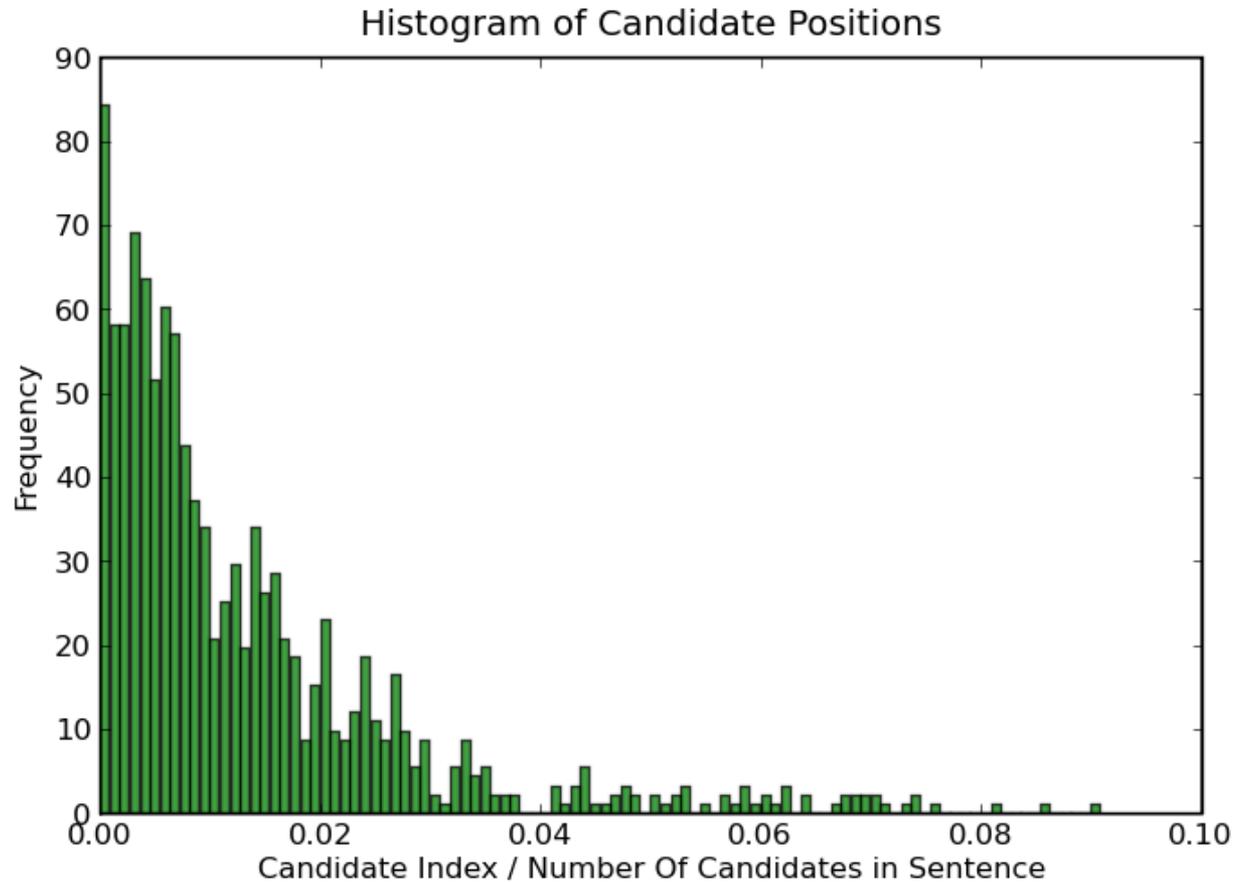
construction

The way to success is under heavy construction

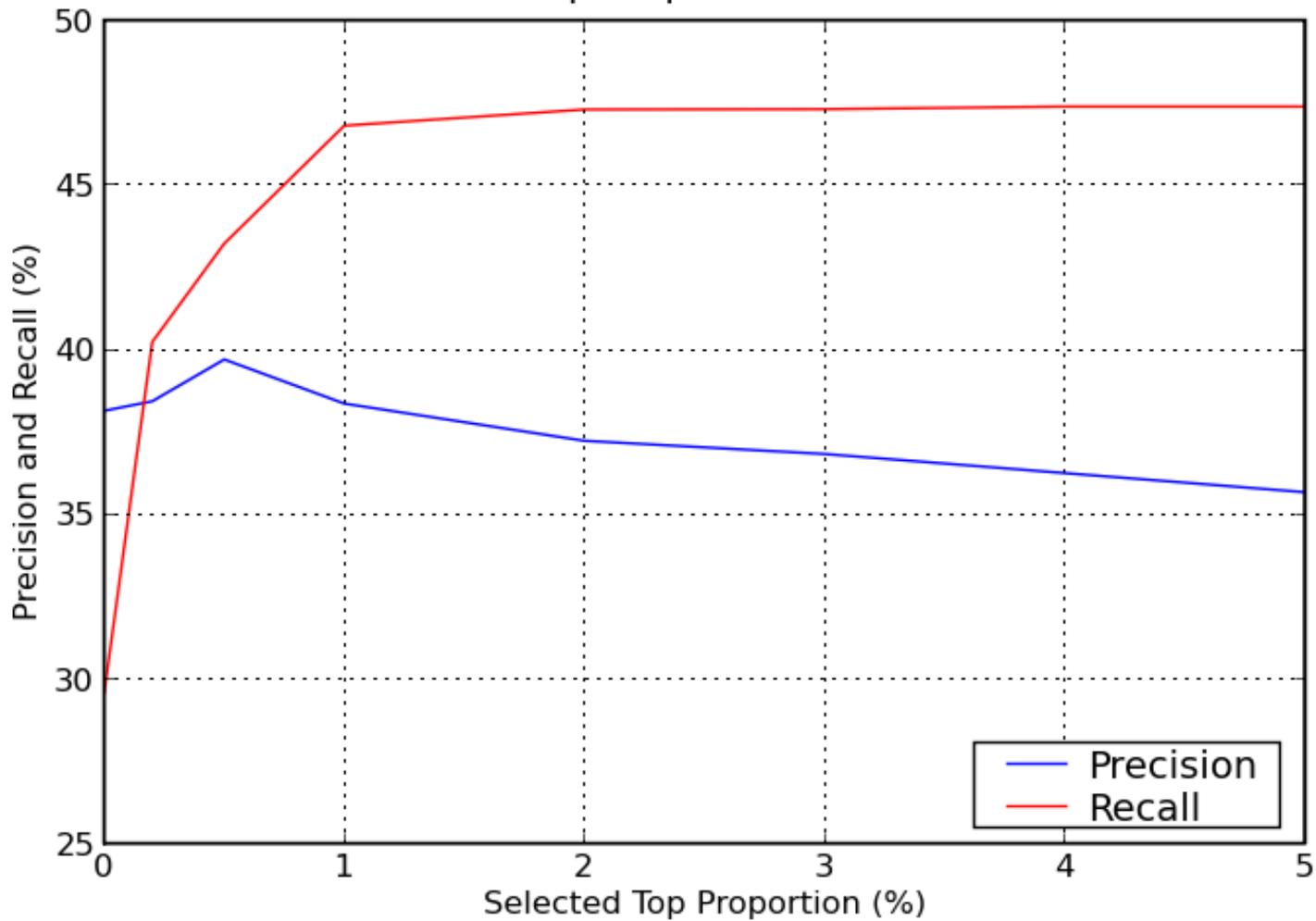
Effects of Increase of Training Set on Precision and Recall

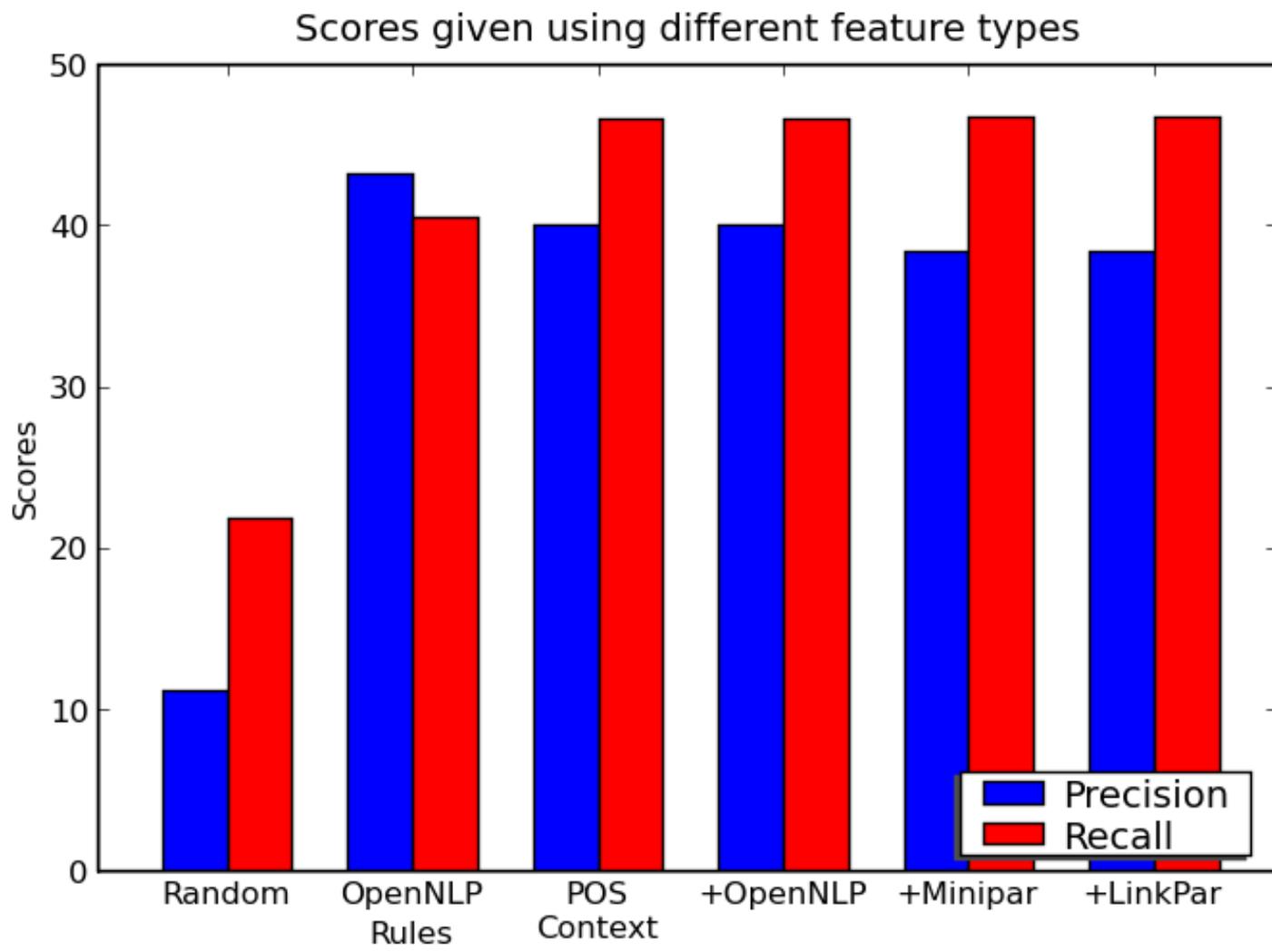


# Candidate Position Histogram

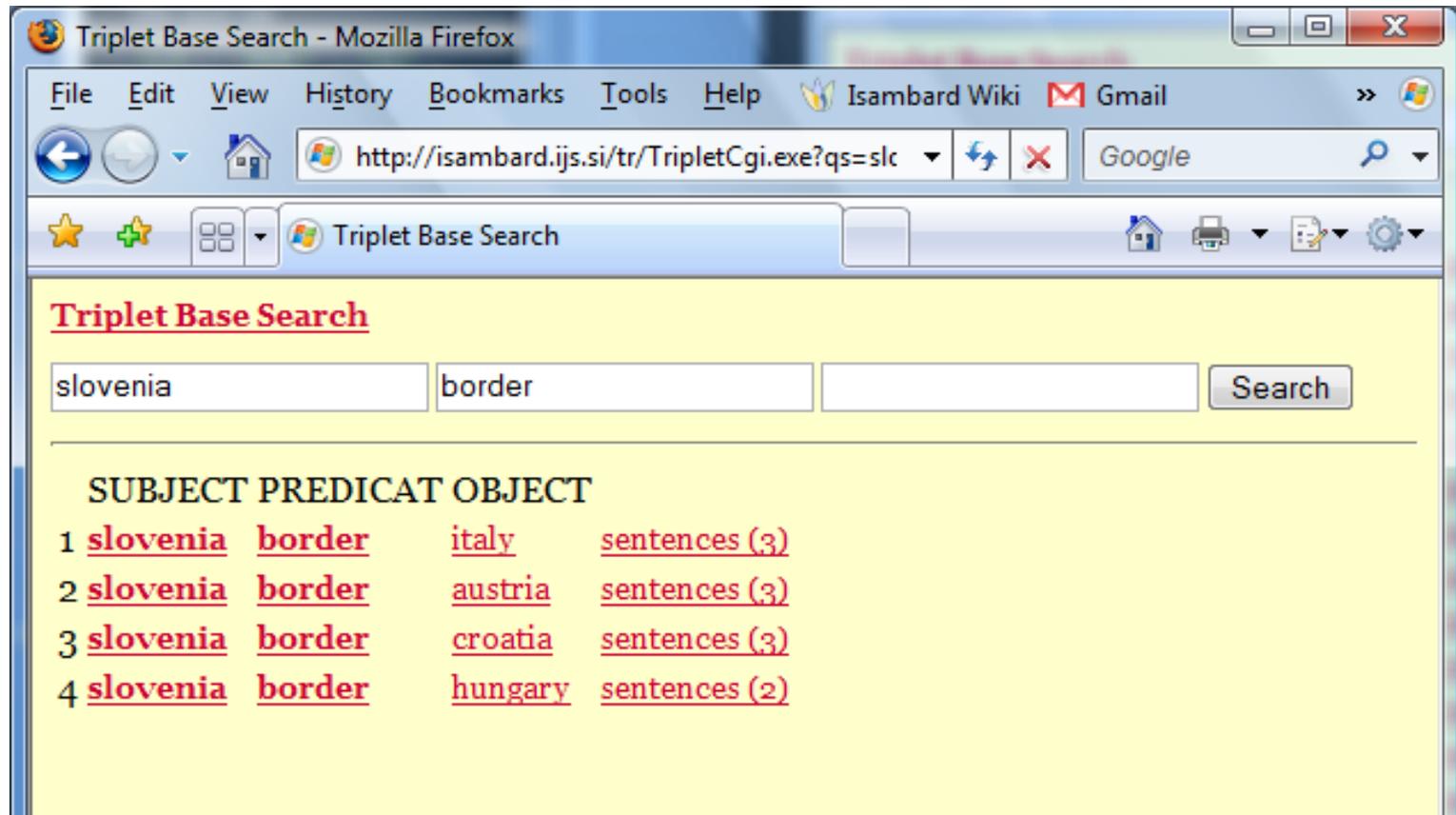


Effects of Selected Top Proportion on Precision and Recall





# Search the extracted triplets



The screenshot shows a Mozilla Firefox browser window titled "Triplet Base Search - Mozilla Firefox". The address bar contains the URL "http://isambard.ijs.si/tr/TripletCgi.exe?q=slc". The search bar contains the text "slovenia" and "border". The search results are displayed in a table format with columns for SUBJECT, PREDICAT, and OBJECT.

**Triplet Base Search**

slovenia border Search

---

**SUBJECT PREDICAT OBJECT**

1	<a href="#">slovenia</a>	<a href="#">border</a>	<a href="#">italy</a>	<a href="#">sentences (3)</a>
2	<a href="#">slovenia</a>	<a href="#">border</a>	<a href="#">austria</a>	<a href="#">sentences (3)</a>
3	<a href="#">slovenia</a>	<a href="#">border</a>	<a href="#">croatia</a>	<a href="#">sentences (3)</a>
4	<a href="#">slovenia</a>	<a href="#">border</a>	<a href="#">hungary</a>	<a href="#">sentences (2)</a>

### Question Answering

Who declared war? Ask

#### Bavaria

Bavaria declares war on canned beer.

Germany's beer capital Bavaria has declared war against an unlikely invader -- canned beer.

#### Australia

Australia declares war on pesky rabbits.

Australia has declared biological war on an estimated 200 million rabbits.

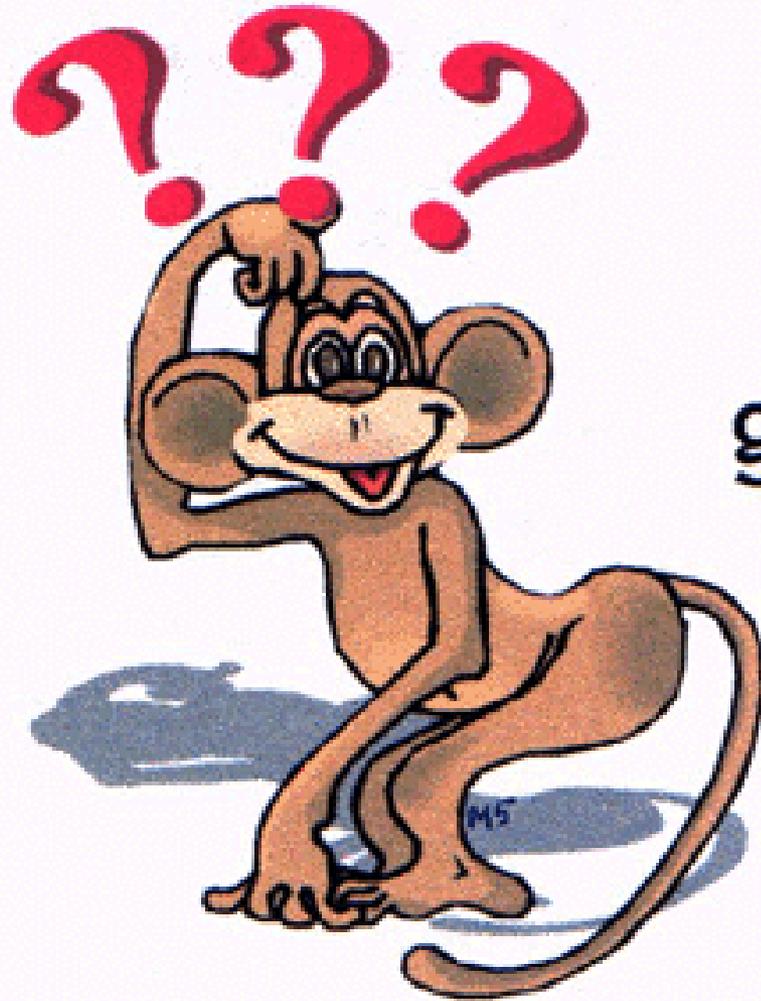
- 1911 - Italy declared war on Turkey for possession of Tripolitania, a province of the kingdom of Libya.
- 1915 - Italy and Russia declare war on Bulgaria.
- 1941 - Italy and Germany declared war on the United States; the U.S. Congress in turn declared war on both.
- 1915 : Having been neutral, Italy declared war on
- 1940 - Italy declared itself at war with France and Britain, effective June 11.

#### France

- 1939 - Great Britain and France declared war on Germany after its invasion of Poland.
- 1914 - France declared war on Turkey.
- 1793 - France declared war on England and Holland.
- 1793 - France declared war on Spain during the French Revolution.
- 1792 - France declared war on Austria, Prussia and Sardinia in the War of the First Coalition.

#### Yeltsin

Done



Questions  
are  
guaranteed in  
life;  
Answers  
aren't.