

# NTCIR-7

## Almost-Unsupervised Cross-Language Opinion Analysis

NLCL group

Taras Zagibalov\*

John Carroll

T.Zagibalov@sussex.ac.uk

J.A.Carroll@sussex.ac.uk

Department of Informatics  
University of Sussex

\* supported by the Ford Foundation International Fellowships Program.

# Overview

- Introduction
- Tasks
- Our Approach
- Lexical Item Extraction
- Relevance Classification
- Subjectivity Classification
- Results
- Error Analysis and Conclusion

# Introduction

- Our main focus is **portability** of natural language processing systems **across languages**
- Our basic approach is an **almost unsupervised** approach

# Tasks

- Japanese
- English
- Simplified Chinese
- Traditional Chinese

# Tasks

- Relevance Classification
- Subjectivity Classification
- Opinion Classification
- Target Detection
- Opinion Holder Detection

# Our Approach

- Lexical Item Extraction
- Relevance Classification
- Subjectivity Classification

# Lexical Item Extraction

Lexical Item (LI) extraction problems:

- A problem of the word boundary detection in Chinese and Japanese.
- A problem of idioms / collocations

# Lexical Item Extraction

LI extraction technique used:

- Any sequence of characters that occurs at least three times is a candidate to be a LI
- If the frequency of a LI is the same as that of a shorter sub-unit then the latter is deleted.



# Lexical Item Extraction

LI extraction technique used:

- Any sequence of characters that occurs at least three times is a candidate to be a LI
- If the frequency of a LI is the same as that of a shorter sub-unit then the latter is deleted.

LI candidate	Frequency	Length	
美国司法	31	4	√
美国司	31	3	X
司	519	1	√

# Relevance Classification

- All LI are ranked according to their frequency in each document
- LI frequency ranks are compared across all the documents
- LI with the biggest rank differences are selected as relevance indicators

# Relevance Classification

- All LI are ranked according to their frequency in each document
- LI frequency ranks are compared across all the documents
- LI with the biggest rank differences are selected as relevance indicators

LI	Topic 1 rank	Topic 2 rank	Difference	
the	2	3	1	X
netscape	0	10	10	√
law	24	6	18	√

# Relevance Classification

- Example:
  - Topic:  
*'What is the relationship between AOL and Netscape?' (N11)*
  - Relevance indicators:  
*america online, appliances, designed, dominant, link, maker, netscape, online, services, start-ups, sun, technological change, they have, windows operating*

# Subjectivity Classification

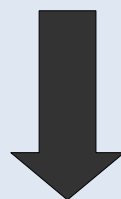
- For each LI we found immediate neighbours:

第五次缔约方大会的中国代表团

# Subjectivity Classification

- For each LI we found immediate neighbours:

第五次缔约方大会的中国代表团



中国： 的 \_0, 大会的 \_0, 代表团 \_1

# Subjectivity Classification

- For each neighbour word we calculated chi-square ( $\chi^2$ ) score
- LI with  $\chi^2 > 3.84$  were included into the list
- All such words were ranked according to their score
- Lists of every two headwords were compared to find how much of context words they shared

# Subjectivity Classification

- Syntactic and Semantic relations separated:

跟 中国 经济 的 快速  
对 美国 经济 的 信心

Syntactic relations	Semantic relations
跟 + 中国 中国 + 经济 美国 + 经济 经济 + 的	中国 + 美国



# Subjectivity Classification

Headwords	中国	美国	经济	的
Context words	经济	经济	中国	经济
Context words	跟	对	的	快速

- Good pairs: 中国 + 美国
- Bad pairs: 中国 + 经济 ; 美国 + 经济 ;  
经济 + 的

# Subjectivity Classification

- Syntactic and Semantic relations separated:  
*there are good years and bad years*  
*stable and good conditions*

Syntactic relations	Semantic relations
are + good good + years and + bad and + good	good + bad

# Subjectivity Classification

Headwords	good	bad	and	years
Context words	and	and	bad	bad
Context words	years	years	good	and

- Good pairs: good + bad
- Bad pairs: and + bad; and + good;  
and + years; years + bad;  
good + years

# Subjectivity Classification

- Filtering the paired headwords:
  - Filter 1:  
Excluded all pairs with a too small association score (the score value less than  $\bar{x} - 1.96\sigma$ )
  - Filter 2:  
Deleted all words that occurred in too many pairs ( LI that occurred in more than  $\bar{x} + 1.96\sigma$  pairs);

# Subjectivity Classification

- RunID1:
  - Use manually filtered words:

*important, difficult, effective, popular,  
successful, easily, troubled, striking,  
best, bad, painful, strong, good*

Result: low recall

# Subjectivity Classification

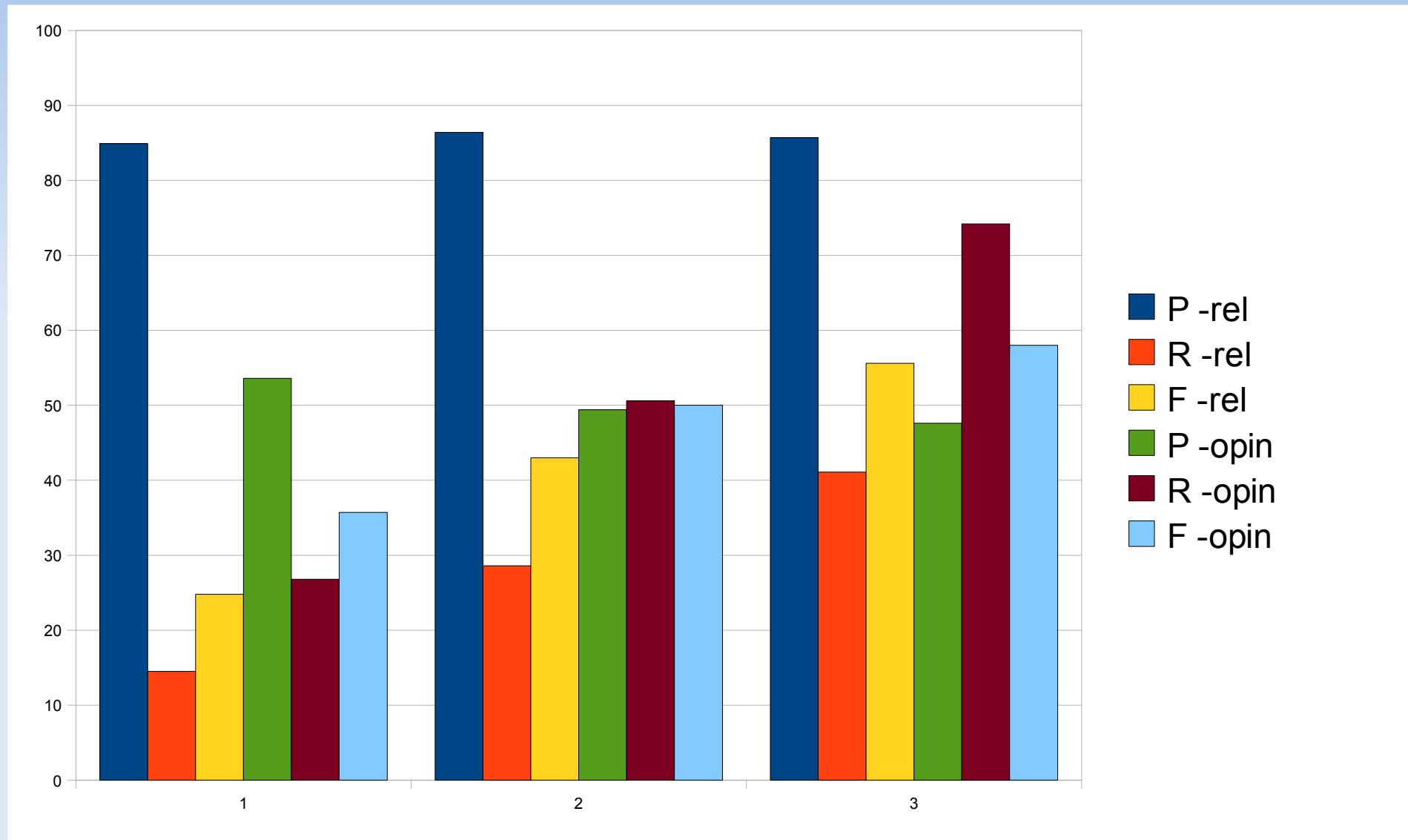
- RunID1:
  - Use manually filtered words
- RunID2:
  - RunID1 + ( $\chi^2 > \text{average}$ )
- RunID3:
  - RunID1 + ( $\chi^2 > 3.84$ )

# Subjectivity Classification

Classification algorithm:

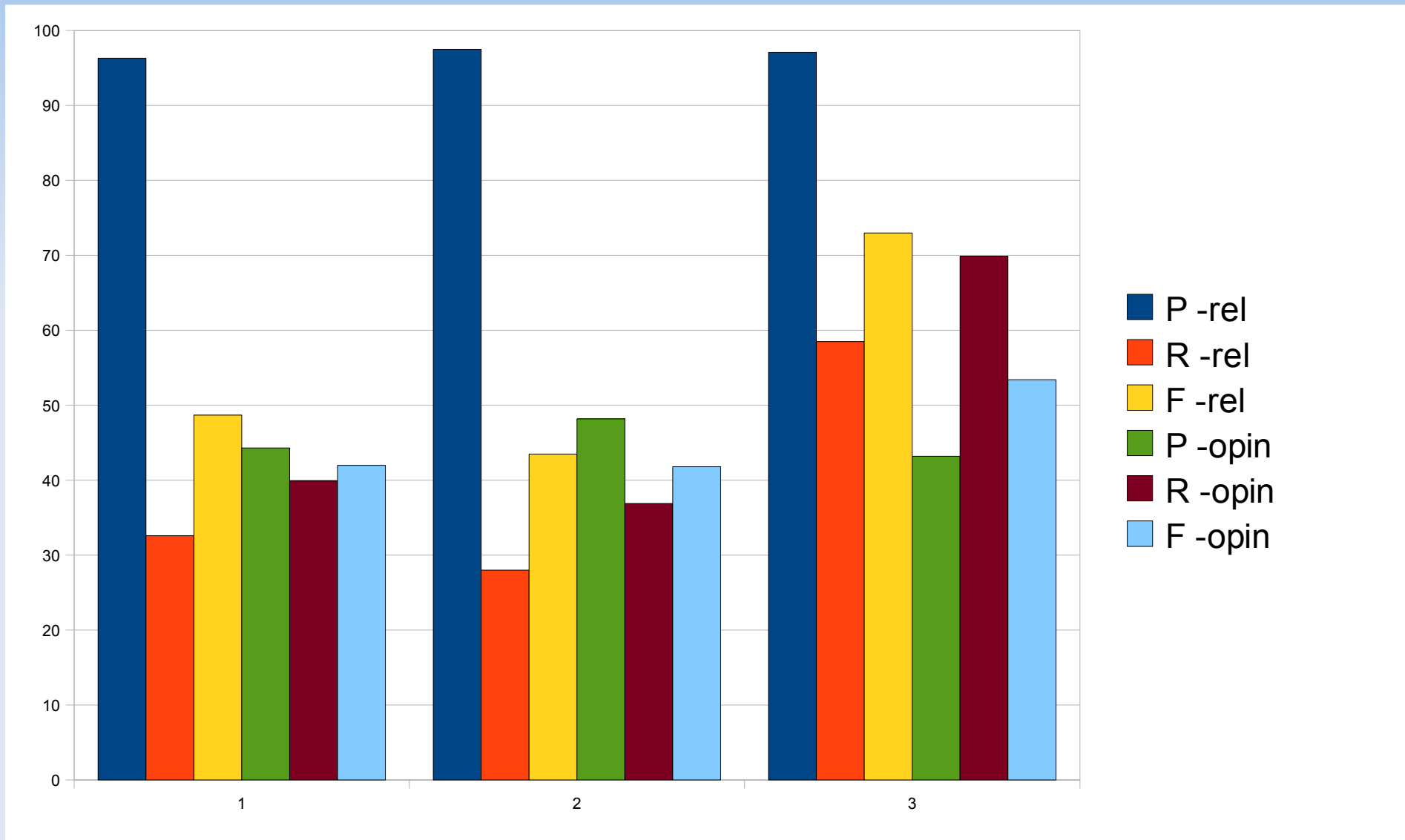
1. If a sentence contains a **relevance marker** >  
RELEVANT
2. If a sentence is RELEVANT and contains a  
**subjectivity marker** >  
OPINIONATED
3. Otherwise >  
NA

# Results: Trad. Chinese (lenient)

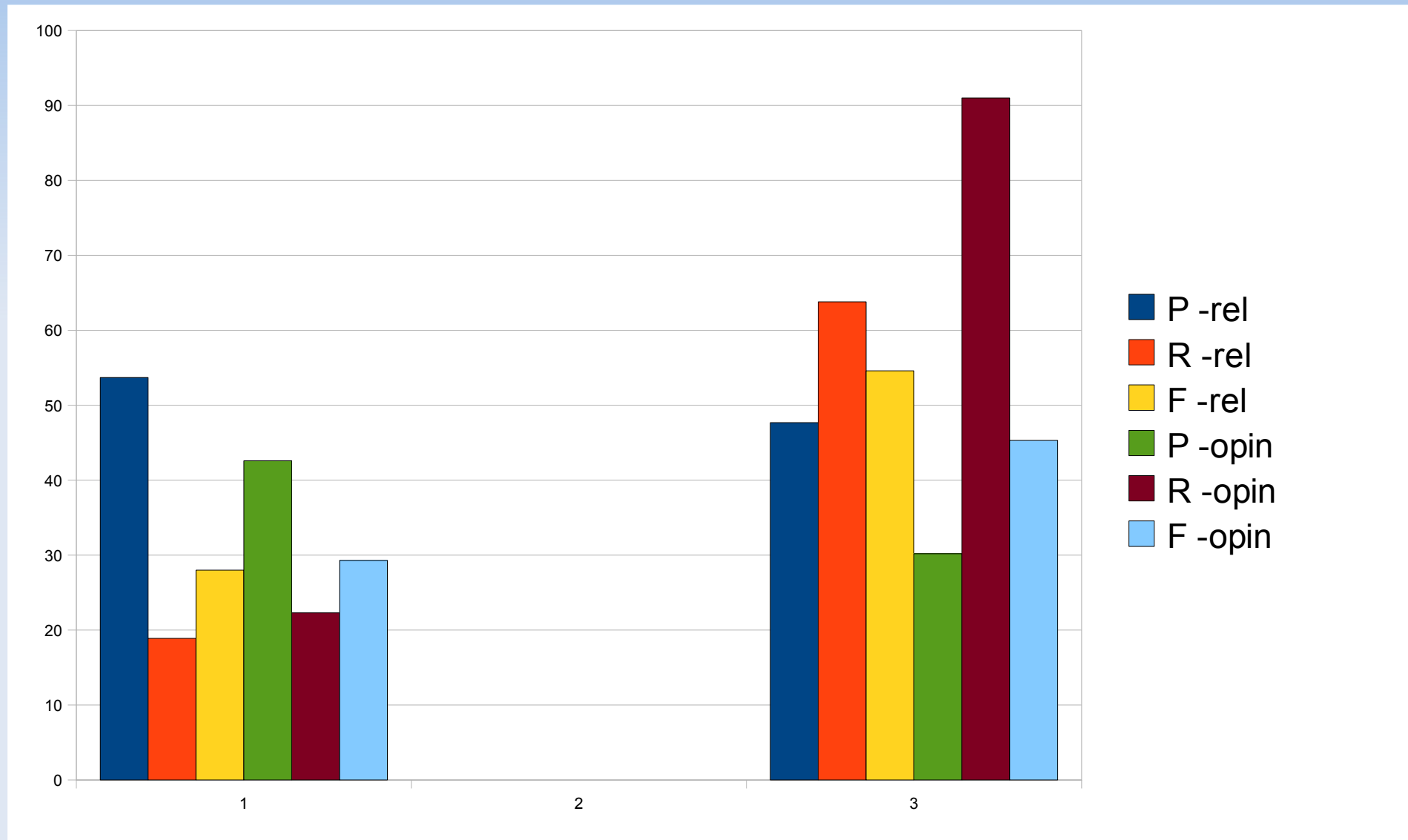




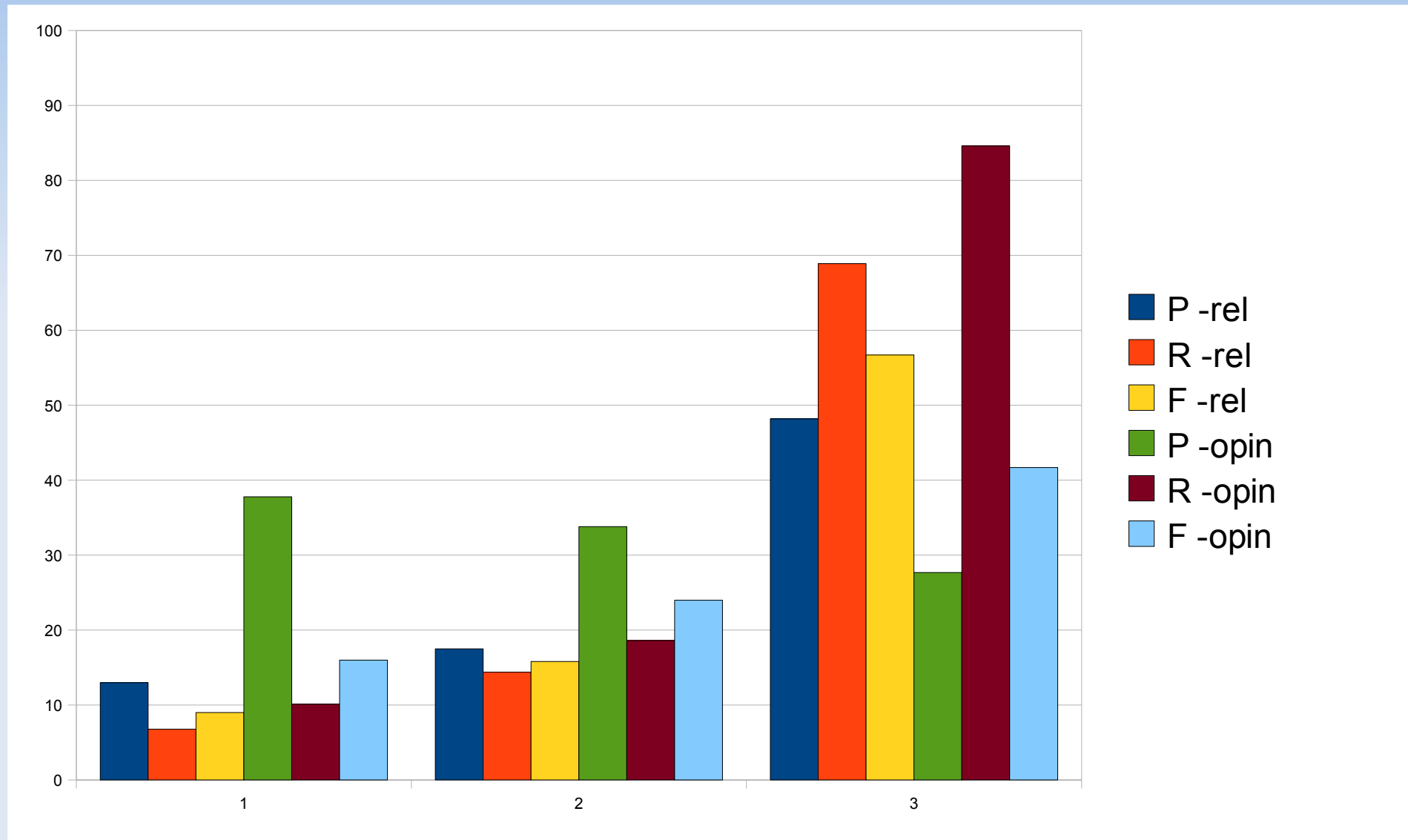
# Results: Simp. Chinese (lenient)



# Results: Japanese (lenient)



# Results: English (lenient)



# Results

## Best results (lenient)

Language	Sub-task (RunID)	Precision	Recall	F-value
T. Chinese	Relevance (3)	48.2	68.9	56.7
	Opinion (3)	27.7	84.6	41.7
S. Chinese	Relevance (3)	97.1	58.5	73.0
	Opinion (3)	43.2	69.9	53.4
Japanese	Relevance (3)*	47.7	63.8	54.6
	Opinion (3)*	30.2	91.0	45.3
English	Relevance (3)	87.5	41.1	55.6
	Opinion (3)	47.6	74.2	58.0

\*Note that the RunID3 results were obtained after the official submission.

# Error Analysis

- Small amount of data
- More noise with higher recall
- Word segmentation for the Asian languages
  - 发展中国家：发展中 + 国家 / 发展 + 中国 + 家
- POS tagging

# Conclusion

- Simple almost unsupervised cross-lingual system
- Satisfactory results for the Japanese and English tasks
- Rather poor performance for the Chinese (both)

# Future Work

- Reduce noise
- Automate subjectivity marker selection
- Develop unsupervised language independent (quasi-)POS tagging technique

ありがとうございます  
謝謝  
谢谢  
Thank you



# Results

## Traditional Chinese (lenient)

Sub-task (RunID)	Precision	Recall	F-value
Relevance (1)	84.9	14.5	24.8
Opinion (1)	53.6	26.8	35.7
Relevance (2)	86.4	28.6	43.0
Opinion (2)	49.4	50.6	50.0
Relevance (3)	85.7	41.1	55.6
Opinion (3)	47.6	74.2	58.0

# Results

## Simplified Chinese (lenient)

Sub-task (RunID)	Precision	Recall	F-value
Relevance (1)	96.3	32.6	48.7
Opinion (1)	44.3	39.9	42.0
Relevance (2)	97.5	28.0	43.5
Opinion (2)	48.2	36.9	41.8
Relevance (3)	97.1	58.5	73.0
Opinion (3)	43.2	69.9	53.4

# Results

## Japanese (lenient)

Sub-task (RunID)	Precision	Recall	F-value
Relevance (1)	53.7	18.9	28.0
Opinion (1)	42.6	22.3	29.3
Relevance (2)	-	-	-
Opinion (2)	-	-	-
Relevance (3)*	47.7	63.8	54.6
Opinion (3)*	30.2	91.0	45.3

\*Note that the RunID3 results were obtained after the official submission.

# Results

## English (lenient)

Sub-task (RunID)	Precision	Recall	F-value
Relevance (1)	13.0	6.8	9.0
Opinion (1)	37.8	10.1	16.0
Relevance (2)	17.5	14.4	15.8
Opinion (2)	33.8	18.6	24.0
Relevance (3)	48.2	68.9	56.7
Opinion (3)	27.7	84.6	41.7