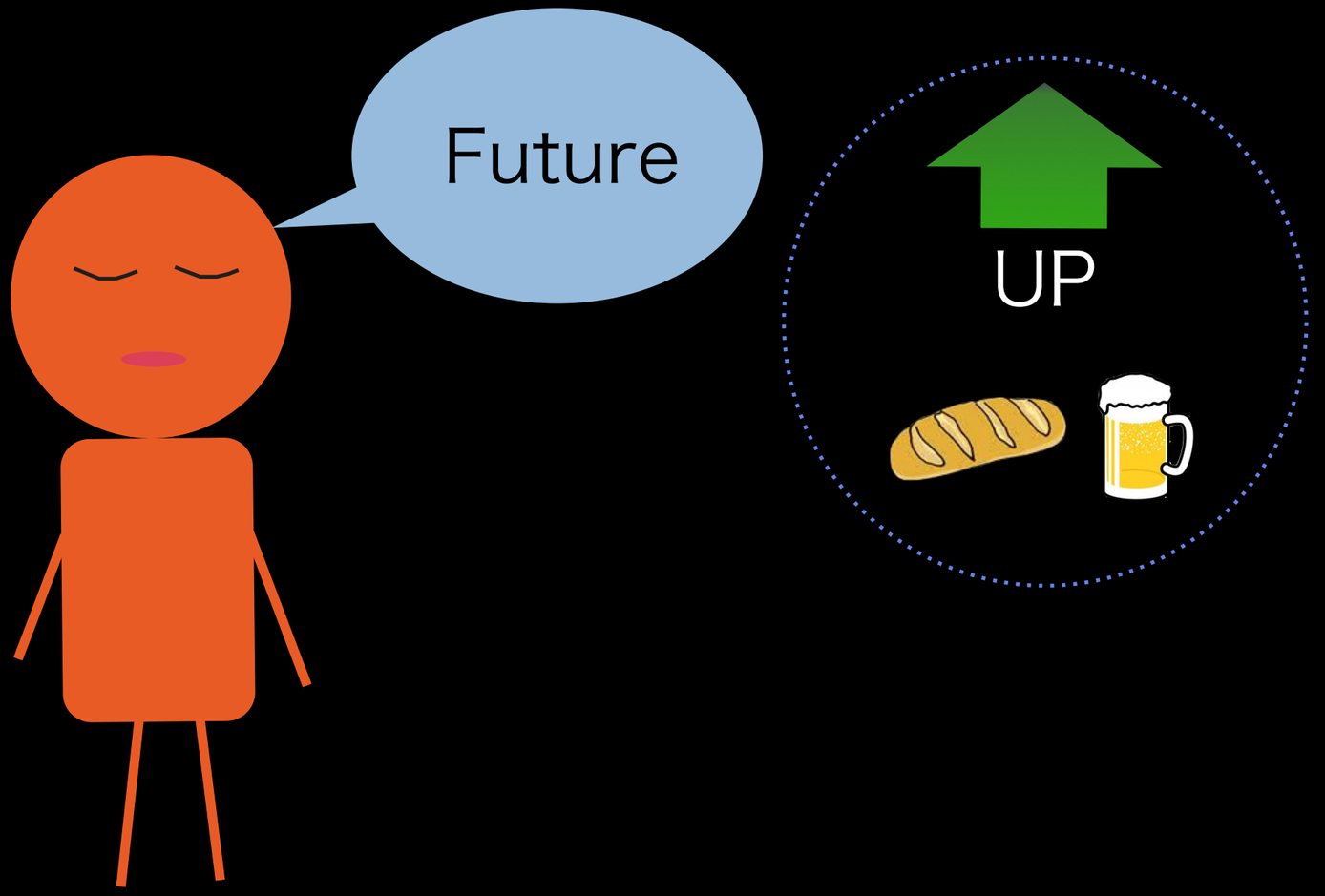
Investigation of Future Reference Expressions in Trend Information

Yoko Nakajima
Michal Ptaszynski
Hirotoshi Honma
Fumito Masui

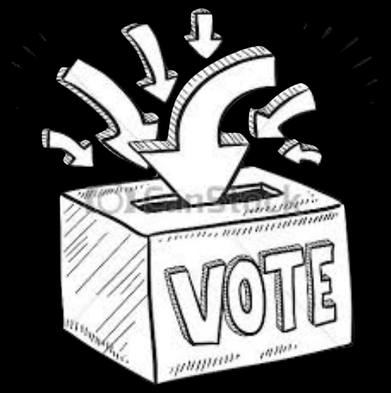
“The price of wheat is expected to rise from October. ”

If this report is trustworthy, ...



People want to know the future

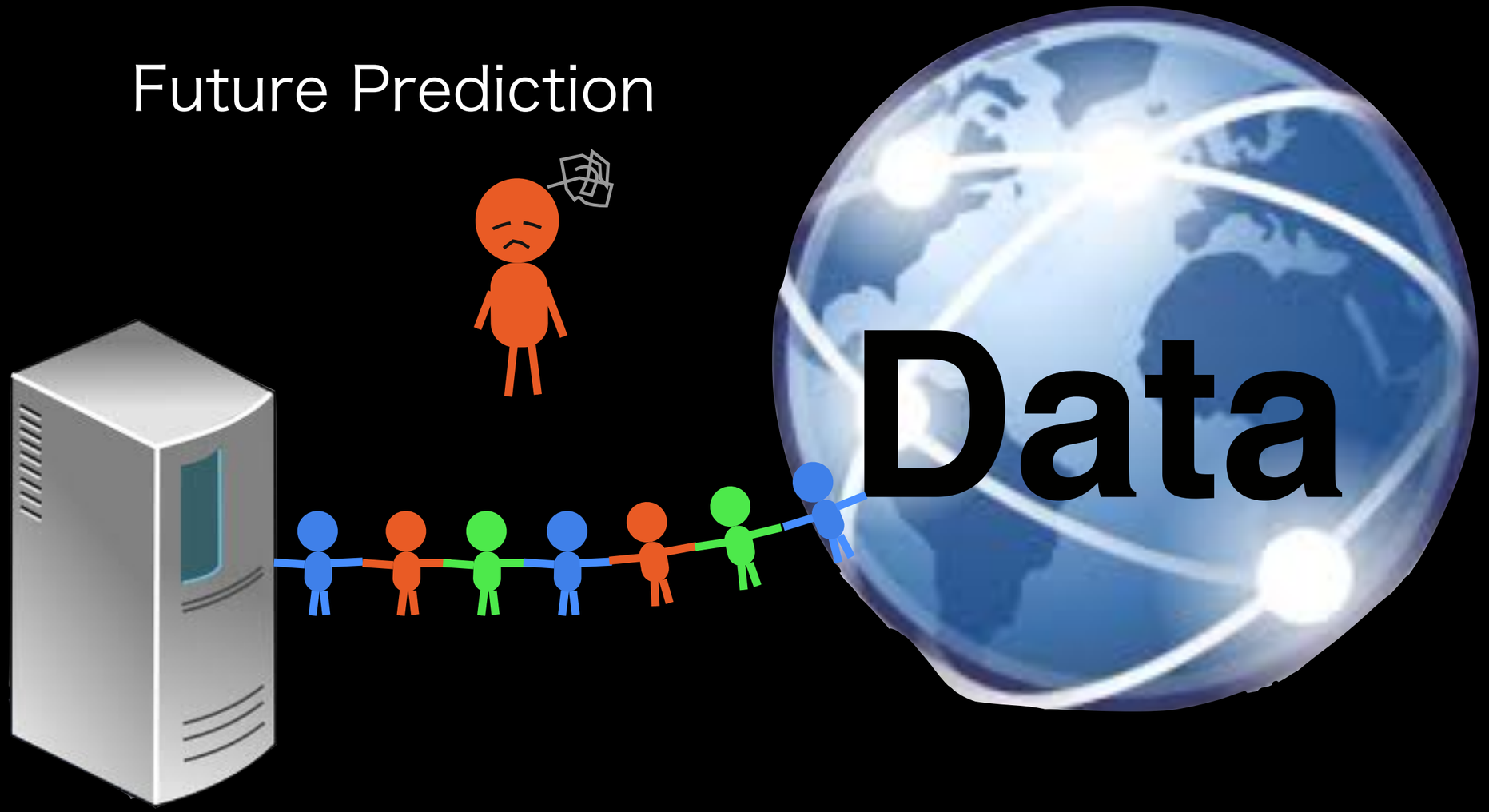
corporate
management



voting trends
in election

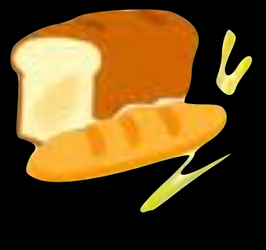
economic
trends

Future Prediction



Future reference expressions

Future reference expressions



They include the information which may happen in the future

This could be used as a hint for predicting the future

Is using future referring expressions effective?

- Investigation of future reference expressions
- Extract patterns of future reference expressions
- Experiment and validation

Is using future referring expressions effective?

- Investigation of future reference expressions
- Extract patterns of future reference expressions
- Experiment and validation

Future reference expressions manually extracted from 270 sentences

Type	frequency	Examples
time related	**70**	• 2014-Nen 3-gatsu kara (from month M year Y) • kon-getsu chujun (this month) • kongo (next in) ... etc.
future	**141**	• mezasu (aim to) • hoshin (plan to) • - suru (do) • - iru (is/to be) ... etc.

Probability of occurrence of future reference words in sentences

occurrence frequency	percentage (%)
one time	45
two times or more	55

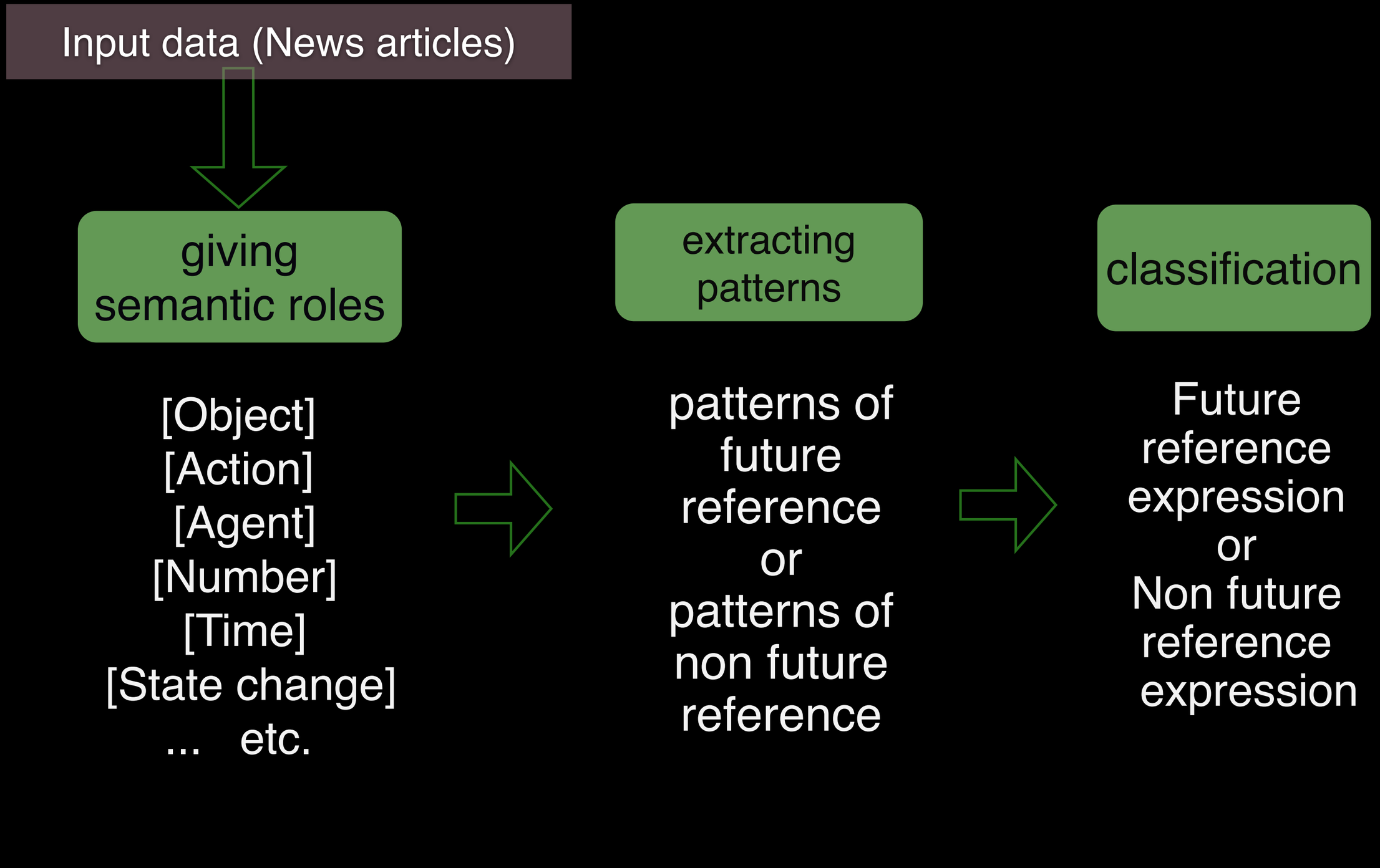
Is using future referring expressions effective?

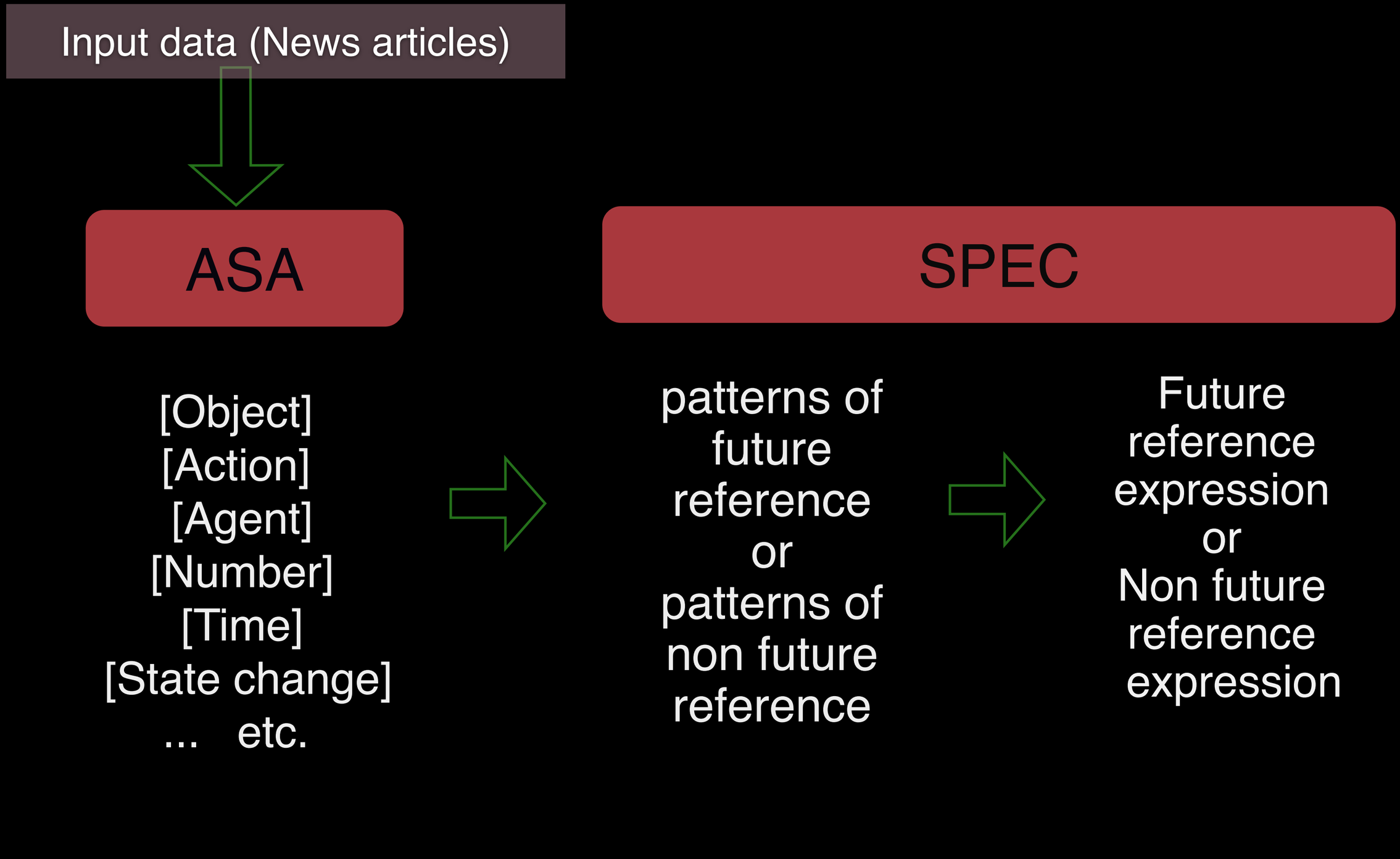
- Investigation of future reference expressions
- Extract patterns of future reference expressions
- Experiment and validation

Input data (News articles)

- Nippon Keizai Shimbun
- Asahi Shimbun
- Hokkaido Shimbun
- Mainichi Shimbun
- http://www.nikkei.com/
- http://www.asahi.com/
- http://www.hokkaido-np.co.jp/

Future
reference sentences
(130)
and
Non-Future
reference sentences
(130)


Input data (News articles)
giving semantic roles
[Object]
[Action]
[Agent]
[Number]
[Time]
[State change]
... etc.
extracting patterns
patterns of future reference or patterns of non future reference
classification
Future reference expression or Non future reference expression


Input data (News articles)
ASA
SPEC
[Object]
[Action]
[Agent]
[Number]
[Time]
[State change]
... etc.
patterns of future reference or patterns of non future reference
Future reference expression or Non future reference expression

Argument Structure Analyzer *

- Construction of Argument Structure Analyzer -

<Input >

Japanese: Kinou kare ha watashini tegami wo okutta.
English: He sent me a letter, yesterday.

<Output>

phrase	semantic role
kinou (yesterday)	: Time-Point
kare ha (he is)	: Agent
watashi ni (me)	: Patient
tegamim wo (a letter)	: Object
okutta (sent)	: State change -Change of position

*http://cl.it.okayama-u.ac.jp/study/project/asa

Additional Post-processing

ASA:

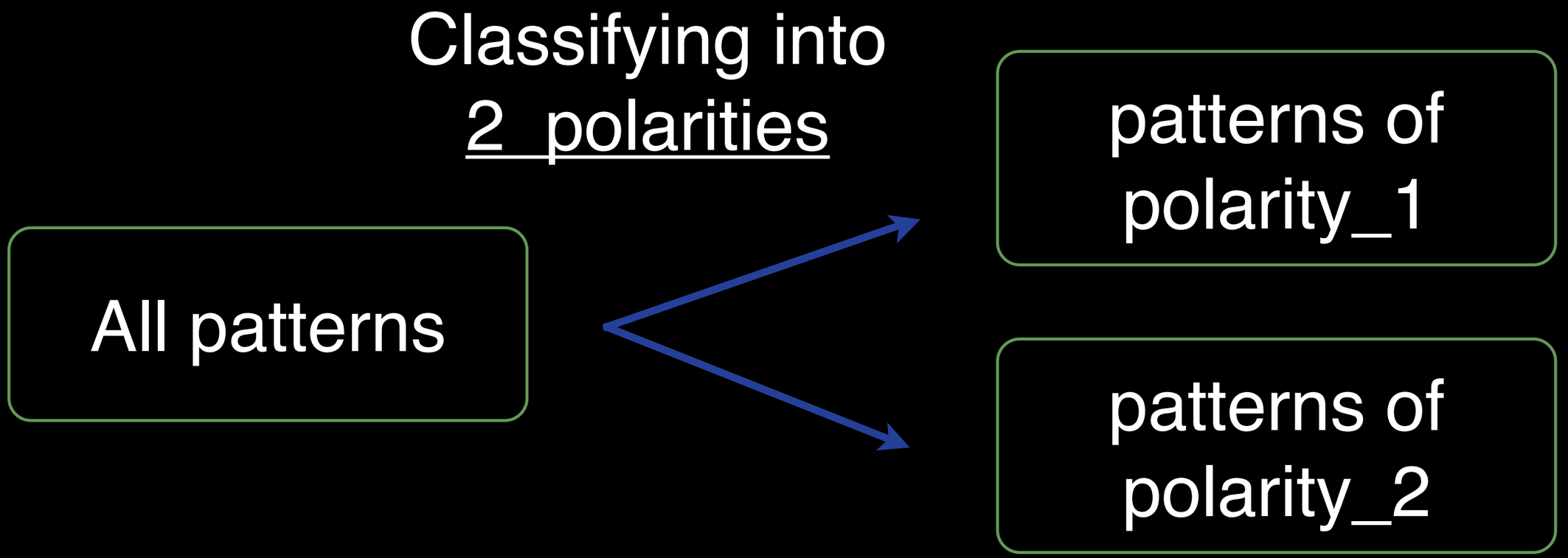
If no predicate phrase → use morphological labels ([noun], [verb], etc.)

- We additionally applied post-processing rules to deal with compound words.

Example:

AAAI Spring Symposium
= [Noun] [Noun] [Noun] → [Noun]

Sentence Pattern Extraction arChitecture



Generating all patterns

[kinou] [kare ha] [watashi ni] [tegami wo] [okutta]

[yesterday] [he] [me] [a letter] [sent]

The patterns generated

- [kinou] [kare ha] [watashi ni] [tegami wo] [okutta]
- [kinou] * [watashi ni] [tegami wo] [okutta]
- [kinou] [kare ha] * [tegami wo] [okutta]
- [kinou] [kare ha] [watashi ni] * [okutta]
- [kinou] * [tegami wo] [okutta]
- [kinou] [kare ha] * [okutta]
- [kinou] * [okutta]

Experiment Setup

- Classification using all patterns
  - 10-fold cross validation
  - sophisticated patterns (with disjoint elements)
- Choose the most useful pattern
- Results calculated in F-score, Precision, Recall

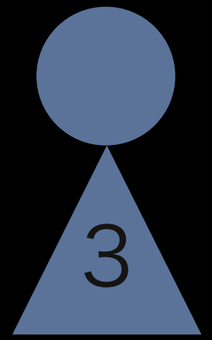
Experiment Setup

- Classification using all patterns
  - 10-fold cross validation
  - sophisticated patterns (with disjoint elements)
    I. all-patterns
    II. n-grams
      i. zero-deleted
      ii. ambiguous-deleted
        a. awarding-length
        b. awarding-length-and-occurrence
- Choose the most useful patterns
- Results calculated in: F-score, Precision, Recall

Is using future referring expressions effective?

- Investigation of future reference expressions
- Extract patterns of future reference expressions
- Experiment and validation

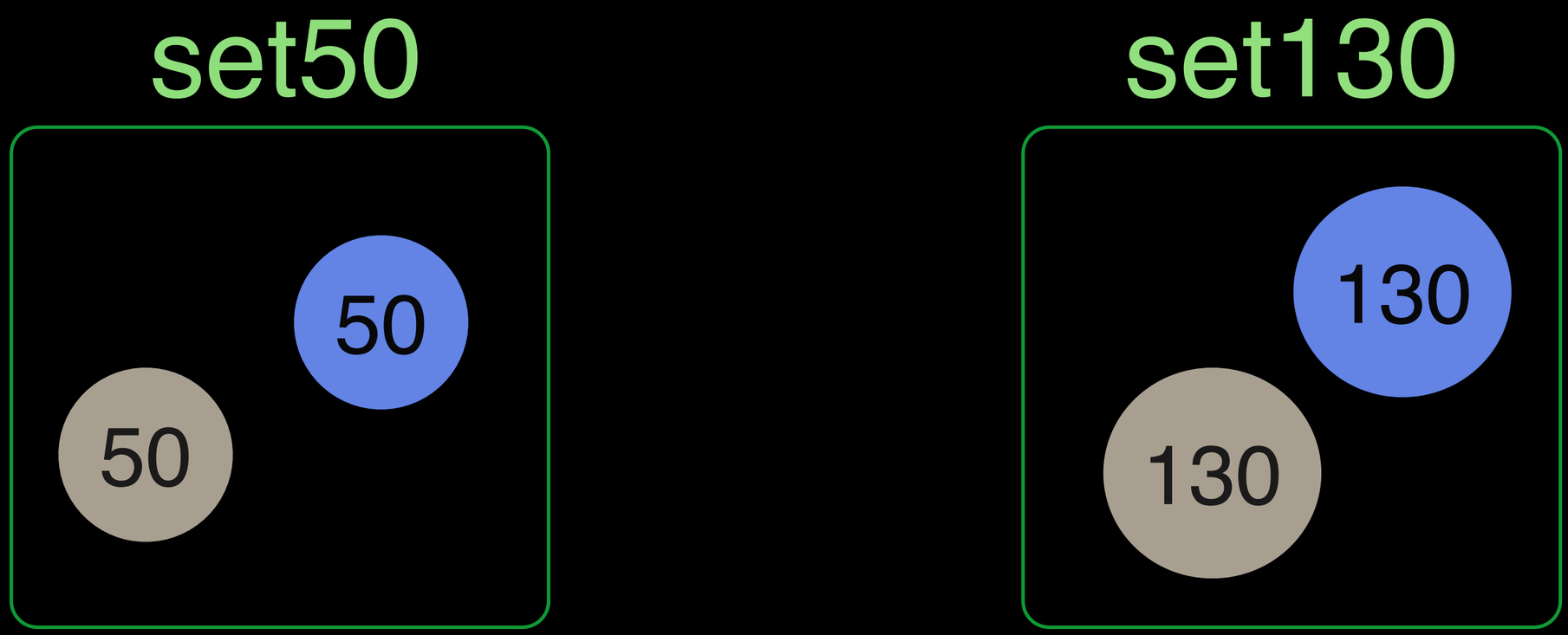
1000 Sentences
from News articles



Future reference sentences	Non-Future reference sentences	Ambiguous sentences (future or non-future)
130	540	330

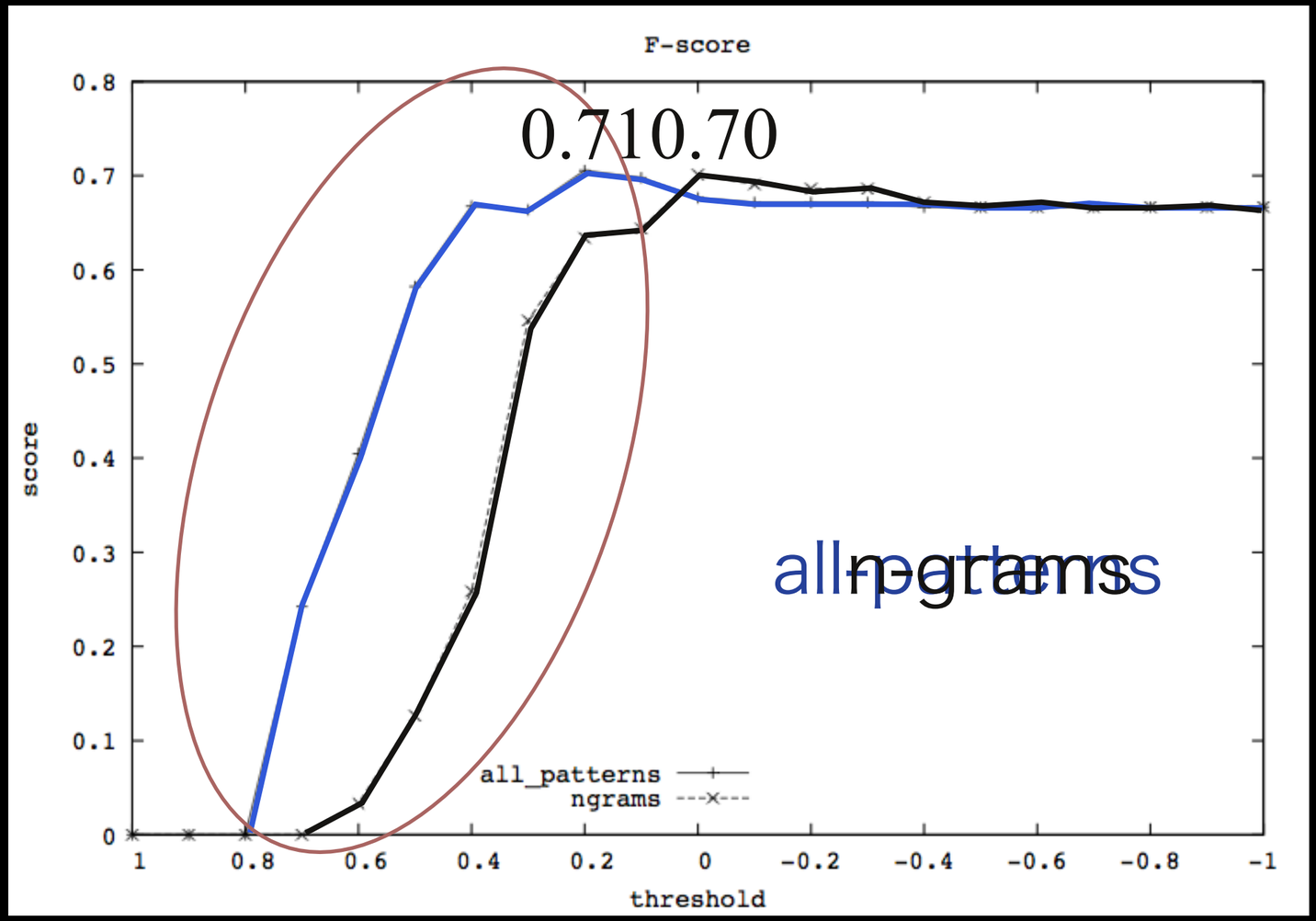
Future reference sentences

Non-Future reference sentences



The results (F-score) for set50

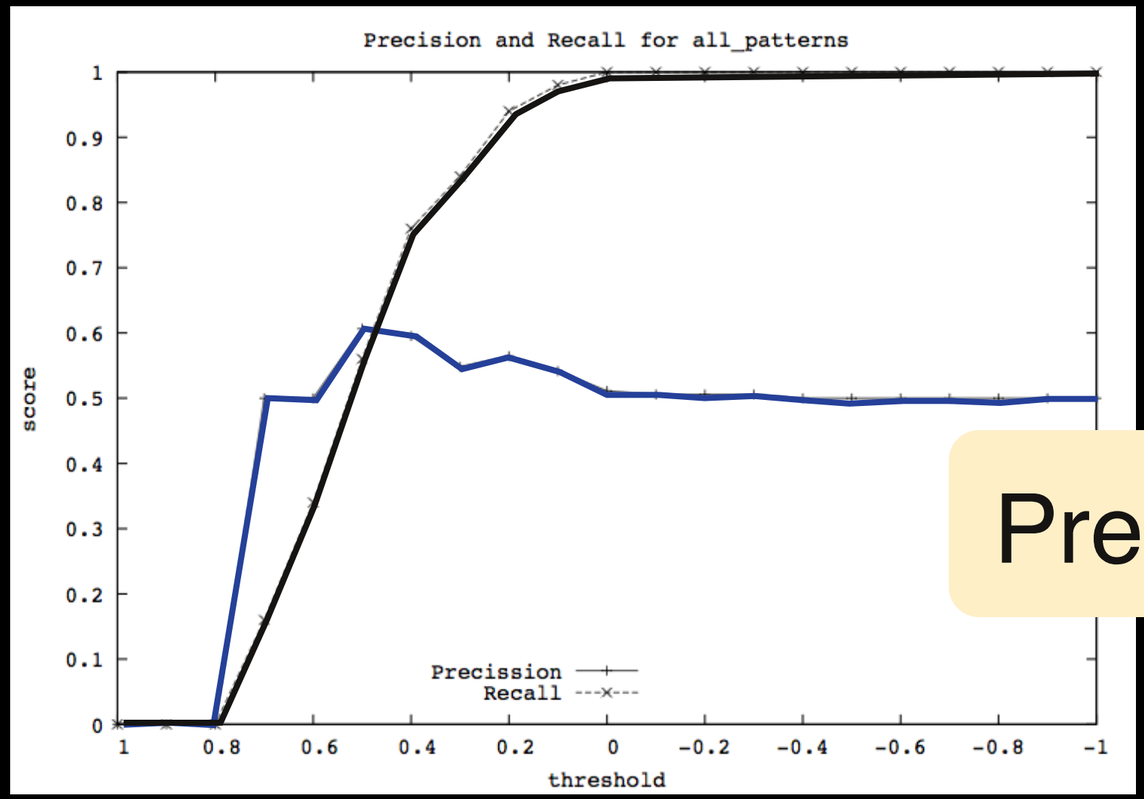
score



Threshold (1~ -1)

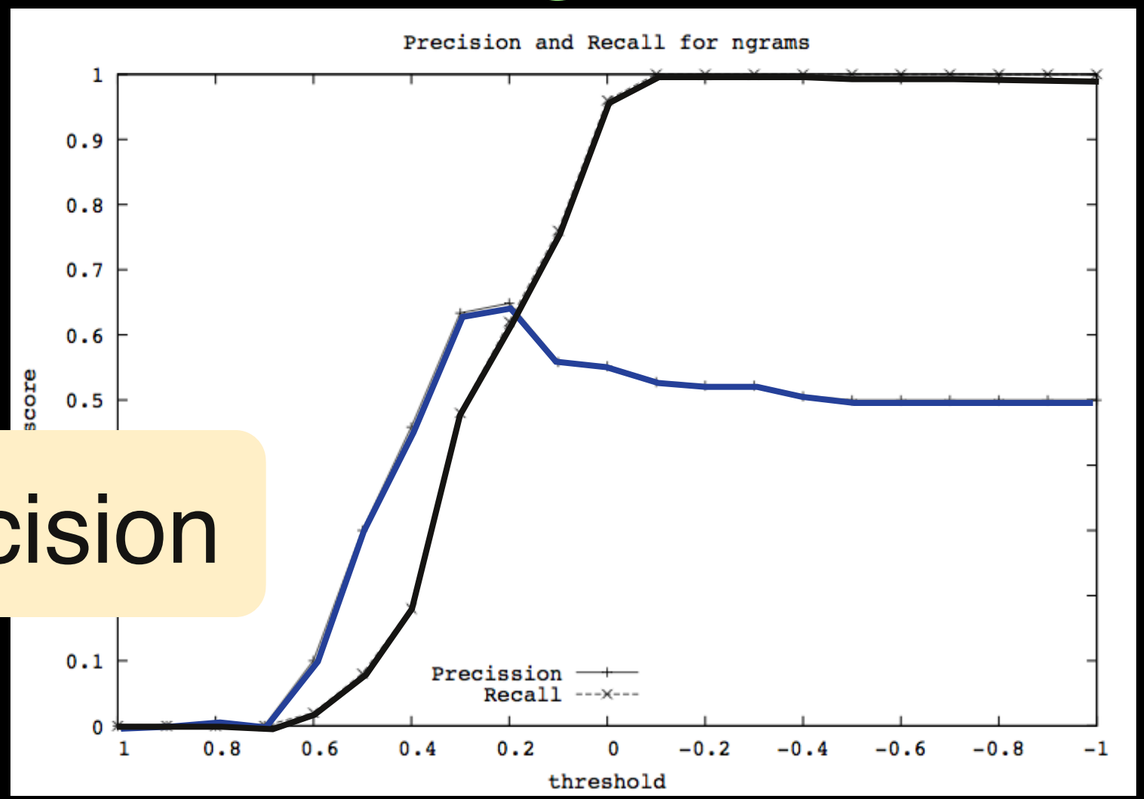
Precision and Recall for set50

for all-pattens



Threshold (1~ -1)

for n-grams



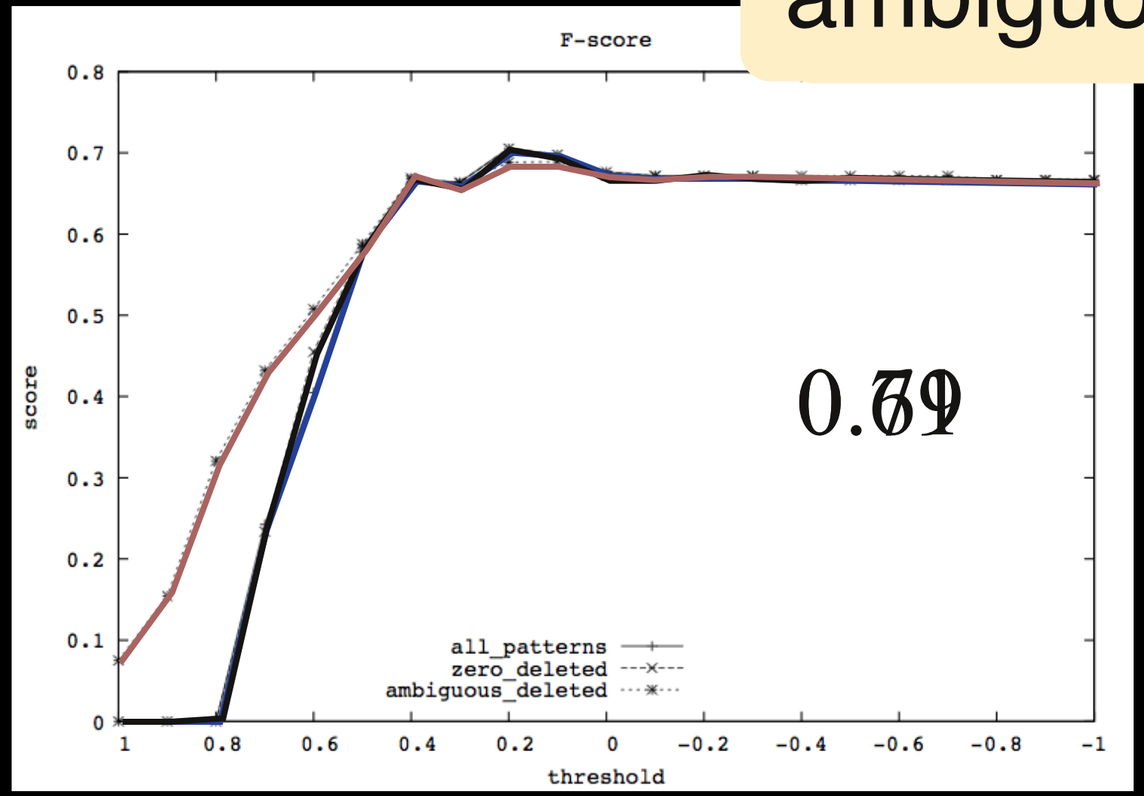
Threshold (1~ -1)

Precision

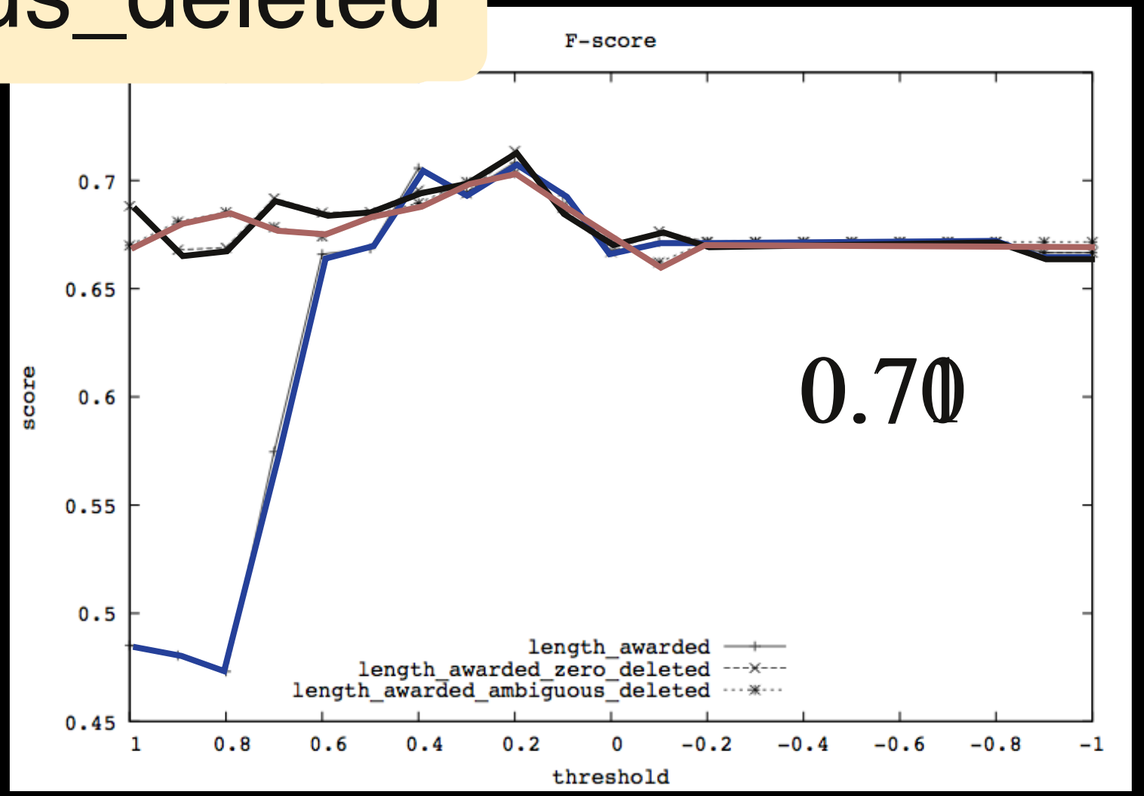
The classifier with three different versions of pattern list for set50

length awarding in weight calculation for set50

ambiguous_deleted



Threshold (1~ -1)



Threshold (1~ -1)

Compare to F-scores set50 and set130

	set50	**set130**
all_patterns	0.71	0.7
zero_deleted	0.71	0.7
ambiguous_deleted	0.7	0.7
length_awarded	0.71	0.7
length_awarded_zero_deleted	0.71	0.69
length_awarded_ambiguous_deleted	0.7	0.7

The examples of extracted patterns

occurrence	Future Reference Patterns	occurrence	Non-Future Reference Patterns
46	[Action]*[State change]	5	[Place]*[Agent]
43	[Action]*[Object]	4	[Number]*[Agent]
42	[Action]*[Action]	4	[Verb]*[Artifact]
20	[State change]*[Object]	4	[Person]*[Place]
16	[State change]*[State change]	3	[Number]*[Agent]*[Action]
15	[Action]*[Object]*[State change]	3	[Adjective]*[State change]*[State change]
15	[Action]*[State change]*[No state change(activity)]	3	[Place]*[Place]*[No state change(activity)

Example1

Future Reference Patterns

[Object]*[Action]*[State change]

Example1

Future Reference Patterns

[Object]*[Action]*[State change]

Pattern:
[Object] [Action] [Agent] [Number] [Time] [Action] [Number] [Number] [Time] [State change]

Japanese:
Nesage jisshi wa shinki kanyuryo, kihon ryokin ga 12gatsu tsuitachi kara, tsuwa ryokin ga 1996yen 3gatsu tsuitachi kara no yotei.

English:
The price cut implementation is planned to apply to the new subscription fees, for the basic rate plan from December 1, for call charges from March 1, 1996.

Example1

Future Reference Patterns

time expressions, future reference expressions

Pattern:
[Object] [Action] [Agent] [Number] [Time] [Action] [Number] [Number] [Time] [State change]

Japanese:
Nesage jisshi wa shinki kanyuryo, kihon ryokin ga *12gatsu tsuitachi kara*, tsuwa ryokin ga 1996yen *3gatsu tsuitachi kara* no *yotei*.

English:
The price cut implementation *is planned* to apply to the new subscription fees, for the basic rate plan *from December 1*, for call charges *from March 1, 1996*.

Example2

Future Reference Patterns

[Action]*[Object]*[State change]

Example2

Future Reference Patterns

[Action]*[Object]*[State change]

Pattern:
[Action] [Number] [experiencer] [Time] [Object] [State change]

Japanese:
Goukeishuturyoku wa 1man8sen-kirowatto de baidenshunyu ha nenkan 8oku5sen-manen teido wo *mikomu*.

English:
With the total output of 18 000 kilowatts, revenue from electricity sales *are expected* to be around 850 million yen per year.

Conclusions

- Future reference sentences contain characteristic expressions
- Extracted characteristic patterns from future reference sentences
- F-score equal to 71%, with Precision equal 56%, and Recall equal 98%

Conclusions

- Future reference sentences contain characteristic expressions
- Extracted characteristic patterns from future reference sentences
- F-score equal to 71%, with Precision equal 56%, and Recall equal 98%

Conclusions

- Future reference sentences contain characteristic expressions
- Extracted characteristic patterns from future reference sentences
- F-score equal to 71%, with Precision equal 56%, and Recall equal 98%

Future Works

- Improving Precision
  - increasing the experiment data, using news according to genres
- Verification of future reference patterns
- Investigation of time changes of the event of a future reference sentences
  - using a detailed topic of news for a specific period for future prediction support

Future Works

- Improving Precision
  - increasing the experiment data, using news according to genres
- Verification of future reference patterns
- Investigation of time changes of the event of a future reference sentences
  - using a detailed topic of news for a specific period for future prediction support

Future Works

- Improving Precision
  - increasing the experiment data, using news according to genres
- Verification of future reference patterns
- Investigation of time changes of the event of a future reference sentences
  - using a detailed topic of news for a specific period for future prediction support

Thank you

References

- Eiji Aramaki, Sachiko Maskawa, Mizuki Morita. 2011. Twitter catches the flu: Detecting influenza epidemics using twitter. *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, pp. 1568--1576.
- R. Baeza-Yates. 2005. Searching the Future. *SIGIR Workshop on MF/IR*.
- Adam Jatowt, Kensuke Kanazawa, Satoshi Oyama, Katsumi Tanaka. 2009. Supporting analysis of future-related information in news archives and the Web. *Proceedings of the 9th ACM/IEEE-CS joint conference on Digital libraries*, pp. 115--124.
- Adam Jatowt, Ching-man Au Yeung. 2011. Extracting collective expectations about the future from large text collections. *Proceedings of the 20th ACM international conference on Information and knowledge management*, pp. 1259--1264.
- Adom Jatowt, Hideki Kawai, Kensuke Kanazawa, Katsumi Tanaka, Kazuo Kunieda, Keiji Yamada. 2013. Multi--lingual, Longitudinal Analysis of Future-related Information on the Web. *Proceedings of the 4th International conference on Culture and Computing 2013*, IEEE Press.
- Kensuke Kanazawa, Adam Jatowt, Satoshi Oyama, Katsumi Tanaka. 2010. Exracting Explicit and Implicit future-related information from the Web(O) (in Japenese). *DEIM Forum 2010*, paper ID: A9-1.
- Kensuke Kanazawa, Adam Jatowt, Katsumi Tanaka. 2011. Improving Retrieval of Future-Related Information in Text Collections. *2011 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology (WI-IAT)*, pp. 278--283.
- Nattiya Kanhabua, Roi Blanco, Michael Matthews. 2011. Ranking related news predictions. *Proceedings of the 34th international ACM SIGIR conference on Research and development in Information Retrieval*, pp. 755-764.
- Michael Matthews, Pancho Tolchinsky, Roi Blanco, Jordi Atserias, Peter Mika, Hugo Zaragoza. 2010. Searching through time in the New York Times. *Proc. of the 4th Workshop on Human-Computer Interaction and Information Retrieval*, pp. 41--44.
- Omar Alonso, Jannik Str\"{o}tgen, Ricardo Baeza-Yates, Michael Gertz. 2011. Temporal Information Retrieval: Challenges and Opportunities. *TWAW*, Vol. 11, pp. 1--8.
- Octavian Popescu and Carlo Strapparava. 2013. Behind the Times: Detecting Epoch Changes using Large Corpora, *Proc. of IJCNLP 2013*.
- Michal Ptaszynski, Rafal Rzepka, Kenji Araki and Yoshio Momouchi. 2011. Language combinatorics: A sentence pattern extraction architecture based on combinatorial explosion. *International Journal of Computational Linguistics (IJCL)*, Vol. 2, Issue 1, pp. 24-36.
- Kira Radinsky, Sagie Davidovich and Shaul Markovitch. Learning causality for news events prediction. *Proceedings of the 21st International Conference on World Wide Web*. ACM, 2012.
- Koichi Takeuchi, Suguru Tsuchiyama, Masato Moriya, Yuuki Moriyasu. 2010. Construction of Argument Structure Analyzer Toward Searching Same Situations and Actions. *IEICE Technical Report*, Vol. 109, No. 390, pp. 1-6.