Extracting References to the Future from News Using Morphosemantic Patterns

*Yoko Nakajima †‡
Michal Ptaszynski †
Hirotoshi Honma ‡
Fumito Masui †

† Kitami Institute of Technology
‡ National Institute of Technology, Kushiro college

IJCAI15 WS
Chance Discovery, Data Synthesis, Curation and Data Market

background

Future References Sentences (FRS)

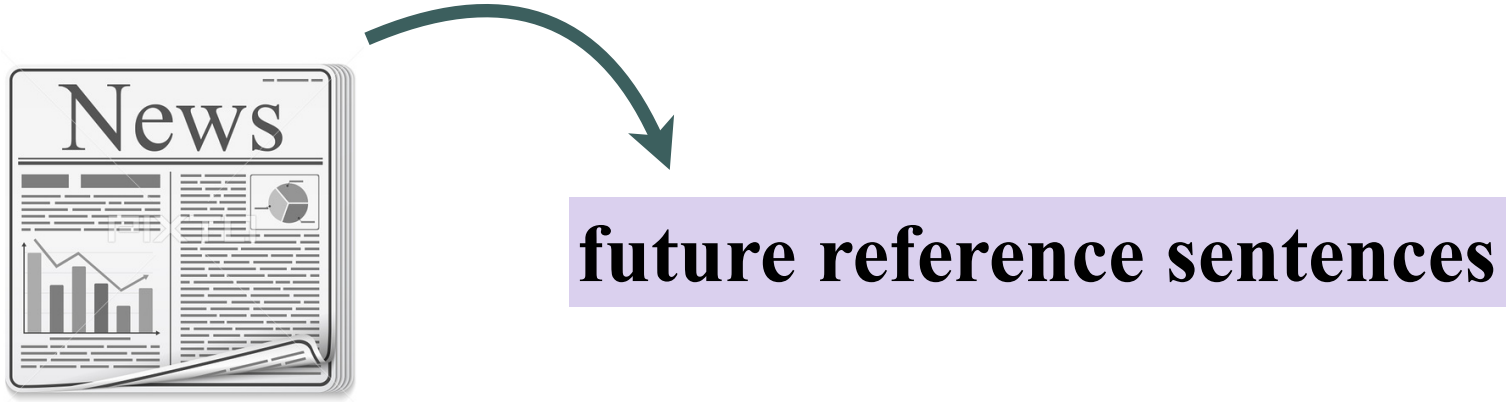
- Describe the probable future event.
- Contain comments to the future event.
- Include information on past events, background knowledge and professional views etc.

Using FRS people can decide about their action and thinking more effectively.

Research purpose

Extract **future reference sentences** from corpus for support action or thinking of people

Text corpus



Our previous work

1. Investigation of future reference expressions
2. Extract patterns of future reference expressions

Future Reference Expression : FRE

Future Reference Sentence : FRS

Our previous work

1. Investigation of FRE

Corpus: newspapers
data: 270 sentences extracted randomly

- not depend on morphology or temporal expressions.
- used variety of words in future references sentences.

Corpus: newspaper
data: 1000 sentences extracted randomly
annotated manually: one expert, two laypeople
(referring to future or not)

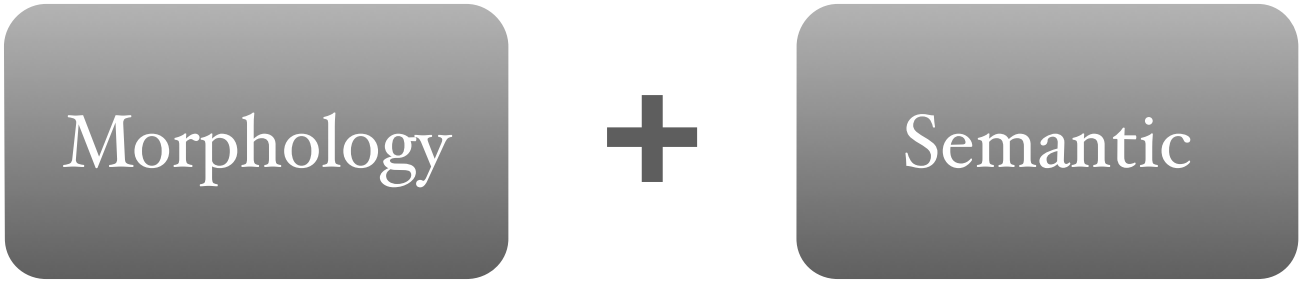
- 13% of newspaper corpus

FRE manually extracted from 270 sentences

Type	frequency	Examples
Temporal Expression	**70**	• next year • tommorow • from month M year Y • this month • next in ... etc.
verb	**141**	• mezasu (aim to) • hoshin (plan to) • - suru (do) • - iru (is/to be) ... etc.

Propose method

Extract patterns of future reference expressions



Morphosemantic
Morphosemantic Patterns : MoPs

example

- analysis of Indonesian suffix in Wordnet [*1]
- analysis of Croatian lexis [*2]

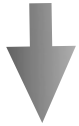
[*1] [Fellbaum et al. 2009] Christiane Fellbaum, Anne Osherson, and Peter E. Clark. 2009. Putting semantics into Word- Net's "morphosemantic" links.
[*2][Raffaelli 2013] Ida Raffaelli. 2013. The model of mor- phosemantic patterns in the description of lexical archi- tecture.

Semantic role labelling

a sentence:

"John killed Mary."

predicate argument structure



semantic role labelling

"**actor action patient**"

Argument Structure Analyzer

Thesaurus of predicate argument structure for Japanese verbs

words: 4400

semantic labels : 80

Application examples

- Construction of Japanese Frame Net [*3]
- Collection of Event Ontology [*4]

[*3] http://jfn.st.hc.keio.ac.jp/ja/

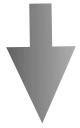
]*4] https://kaigi.org/jsai/webprogram/2014/person-1073.html

Morphological analysis

a sentence:

"John killed Mary."

morphological structure



"noun verb(past) noun"

MeCab

Standard tool morphology for Japanese [*5]

[*5] http://mecab.sourceforge.net

Additional Post-processing

- **ordering of priority taking semantic roll**
  1. Semantic role (Agent, Patient, Object, etc.)
  2. Semantic meaning (State change, etc.)
  3. Category (Dog → Living animal → Animated object)
  4. Adjunct (Time-Point, Time-Line, Location, etc.)
  5. parts of speech

- **compound word clustering**

Example:

"International Joint Conference on Artificial Intelligence"

→ Adjective Adjective Noun Preposition Adjective Noun

→ Proper **Noun**

Example of morphosemantic structure (MS)

Japanese : ニホンウナギが絶滅危惧種に指定され、完全養殖による
ウナギの量産に期待が高まっている。

Alphabet: *nihonunagi ga zetumetu kigushu ni sitei sare, kanzen yo-shoku ni yoru unagi no ryousan ni kitai ga takamatte iru.*

English : As Japanese eel has been specified as an endangered species, the expectations grow towards mass production of eel in full aquaculture.

MS : [Object] [Agent] [State change] [Action] [Noun]
[State change] [Object][State change]

Extracting morphosemantic patterns

SPEC : Sentence Pattern Extraction arChitecture [*6]

- Generate all combination from all elements of a sentence.
- Calculate occurrence frequency of combinations in a corpus.
- Frequent combinations = patterns

[*6] [Ptaszynski et al. 2011] Michal Ptaszynski, Rafal Rzepka, Kenji Araki and Yoshio Momouchi. 2011. Language combinatorics: A sentence pattern extraction architecture based on combinatorial explosion.

Generating all patterns from a sentence

J : [kinou]	[kare ha]	[watashi ni]	[tegami wo]	[okutta]
MS : [Time-Point]	[Agent]	[Patient]	[Object]	[State change]
E : [yesterday]	[he]	[me]	[a letter]	[sent]

1. [Time-Point] [Agent] [Patient] [Object] [State change]
2. [Time-Point] * [Patient] [tObject] [State change]
3. [Time-Point] [Agent] * [Object] [State change]
4. [Time-Point] [Agent] [Patient] * [State change]
5. [Time-Point] * [Object] [State change]
6. [Time-Point] [Agent] * [State change]
7. [Time-Point] * [State change]

⋮

Experiment setup

corpus

- Japan Economy Newspaper
- Asahi Newspaper (national)
- Mainichi Newspaper (national)
- Hokkaido Newspaper (regional)
- http://www.nikkei.com/
- http://www.asahi.com/
- http://www.hokkaido-np.co.jp/

- extract 1000 sentences randomly
- annotate FRE or NRE

*FRE : 13% of news paper corpus.

Training data
set50 and **set130**

FRS
(50 or130)
and
other sentences
(50 or 130)

learning by SPEC

Experiment setup

- sophisticated patterns (with disjoint elements)
  - awarding length (LA)
  - awarding length and occurrence (LOA)
  - awarding none (normalized weight, NW)
  - using all patterns (ALL)
  - erasing all ambiguous patterns (AMB)
  - erasing only those ambiguous patterns which appear in the same number in both sides (zero patterns, 0P)
  - patterns (PAT)
  - only n-grams (NGR)
- n-fold cross validation
- Results calculated in F-score, Precision, Recall
- Choose the most useful pattern

Experiment1: Extract MoPs

- Test data: set50, set130
- 10-fold cross validation

Compare to F-scores set130 and set50

sofisticated patterns	set50	set130
all_patterns	0.71	0.70
zero_deleted	0.71	0.70
ambiguous_deleted	0.70	0.70
length_awarded	0.71	0.70
length_awarded_zero_deleted	0.71	0.69
length_awarded_ambiguous_deleted	0.70	0.70

The examples of extracted MoPs

occurrence	Future Reference Patterns	occurrence	Other Patterns
26	[Action]*[State change]	5	[Place]*[Agent]
23	[Action]*[Object]	4	[Number]*[Agent]
22	[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)

Experiment 2: Extract FRS with frequent patterns

- corpus

  Mainichi Newspaper (1996)

  topics: economy, international event, energy

  270 sentences

- validation data set

  annotate manually from 270 sentences

  ・one expert ・four laypeople

  → FRS: 100 , other: 170

- frequent patterns

  out of learning set130 with length awarded

Extract with frequent patterns by pattern matching

A: first 10 patterns
B: adding 5 patterns longer than three elements to set A
C: subtracting 5 patterns from the tail of set A
D: using only first 10 patterns containing more than three elements

first 10 patterns

occurence	Frequent patterns
26	[action] * [state change]
23	[Action] * [Object]
22	[Action] * [Action]
20	[State change] * [Object]
16	[State change] * [State change]
15	[Action] * [Object] * [State change]
14	[Object] * [Action] * [State change]
13	[Object] * [Action] * [Object]
12	[State change] * [Action] * [State change]
10	[Action] * [Action] * [No state change(Activity)]
9	[Action] * [Action] * [State change] * [Object]
9	[State change] * [Noun] * [Object]
8	[Action] * [State change] * [No state change(Activity)]
8	[Object] * [Action] * [Object] * [Sate change]
5	[Action] * [State change] * [Action] * [No state change(Activity)]
5	[Action] * [State change] * [Object] * [No state change(Activity)]

Performance of extracted FRS with most frequent patterns

Pattern set	P	R	F
A: 10 patterns	0.39	0.49	**0.43**
B: 15 patterns	0.38	0.49	**0.43**
C: 5 patterns	0.35	0.35	0.35
D: 10 patterns with only over 3 elements	0.42	0.37	0.40
baseline (10 temporal expressions) [*2]	0.50	0.05	0.10

6/170 sentences

* 2 [Jatowt and Au Yeung 2011]

Experiment 3：Validation for fully optimized model

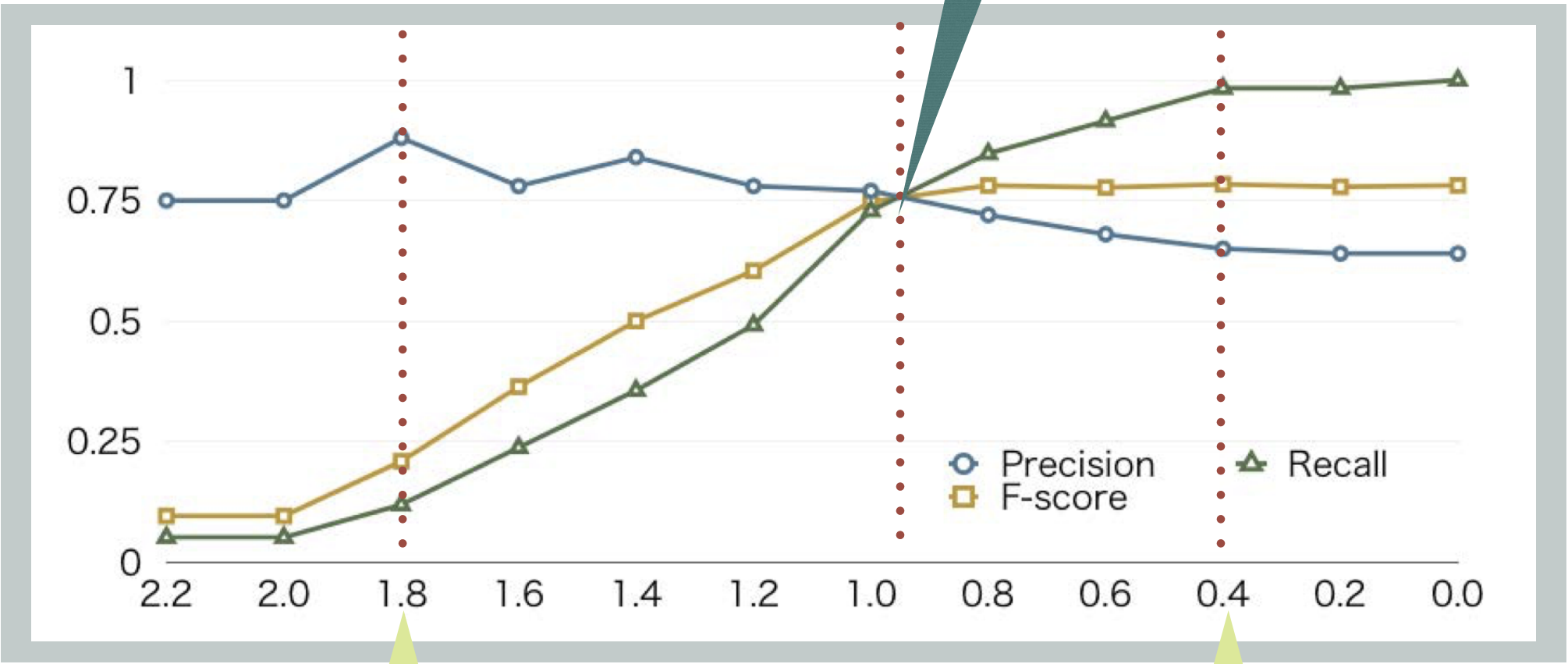
- Corpus:
  Mainichi Newspaper (1996)
  topics: economy, international event
  270 sentences (FRS:100 Non FRS:170)

- Training data:
  1 cross validation for set130
  result calculated with length-awarded

- Evaluation:
  one expert, four laypeople

Classification result

Precision, Recall, F-score



P=0.89
R=0.13
F=0.22

P=0.65
R=0.98
F=0.78

Example of extracted future referring sentence

1. score=2.27

RJ: ***Dosha*** *wa kore made,****Shigen Enerugi-Cho¯ni taishi****, , do¯ hatsudensho no* ***heisa****, kaitai ni tsuite ho¯shin o setsumei shitekitaga, kaitai ni tsuite no ho¯teki kisei wanai tame, do¯chō mo kaitai no kettei o shitatameru koto* ***ni nari-so¯da****.*

E : So far the **company** has been describing to **the Agency for Natural Resources and Energy** the policy for either **closure** or dismantling of the plant, and since there are no legal regulations found for dismantling, it is most likely that the agency **will also lean to** the decision of dismantling.

MS: **[Agent]** [Other] **[Organization]** [Action] **[State-change]** [State-change][Object][Role] [State-change] [State-change][Action][Adjective][Thing][Agent][State-change][Other] **[Verb]**

MoPs : [Agent]*[Verb],

[Agent]*[Organization]*[Verb],

[Agent]*[Action][State-change]*[Verb],

[Agent]*[Organization]*[State-change]*[Verb] .

Conclusion

- We presented a novel method for extracting references to fu- ture events.
- Based the method on automatically extracted morphosemantic patterns.
  - Represent news articles in morphosemantic structure.
  - Extract all possible morphosemantic patterns from the corpus.
- Performed a text classification experiment.
- Compared 14 different classifier versions.
- Compared to the state-of-the-art.
  - The proposed method outperformed the state-of-the-art.
- Validated the method on new dataset.
  - Final score was break even point of precision and recall = 76%.

Future Work

- Increase the size of the experimental datasets.
- Apply in practice
  - Estimating probable unfolding of events.
  - Contribute to trend prediction.

Thank you

Generating all patterns

[pattern]

1 2 3 4 ... n

number of elements : n

number of group of combination : k

in k-element : $k \subseteq n \quad 1 \leq k \leq n$

$$\binom{n}{k} = \frac{n!}{k!(n-k)!} \tag{1}$$

$$\sum_{k=1}^{n} \binom{n}{k} = \frac{n!}{1!(n-1)!} + \frac{n!}{2!(n-2)!} + \ldots + \frac{n!}{n!(n-n)!} = 2^n - 1 \tag{2}$$