《The 6th. International Conference on Discovery Science 2003》

What Kinds and Amounts of Causal Knowledge Can Be Acquired from Text by Using Connective Markers as Clues?

INUI Takashi　INUI Kentaro　MATSUMOTO Yuji
Nara Institute of Science and Technology (NAIST), Japan

Background & Research goal (1/2)

- To achieve intelligent machines, we require large amounts of several types knowledge.

- In previous research (e.g. [Lenat1995][Stork1999]), much of the knowledge is constructed manually. It is costly due to the scale of required knowledge.

- Automatic knowledge acquisition from document collections

Document collections **Knowledge base**

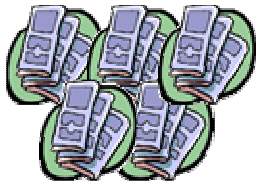
Background & Research goal (2/2)

- Knowledge acquisition of causal relations
  - Refer to an event that causes another event

Someone is thirsty → Someone Drinks water

It rains hard → Flooding occurs

- Key idea : use connective markers as clues



The laundry dried well because it was sunny.

It is sunny → The laundry dries well → KB

Outline of presentation

First part : introduction

- Example
- Typology of causal relations

Second part : technical issues

- Key idea & its problem
- Analyses & Experiments

Outline of presentation

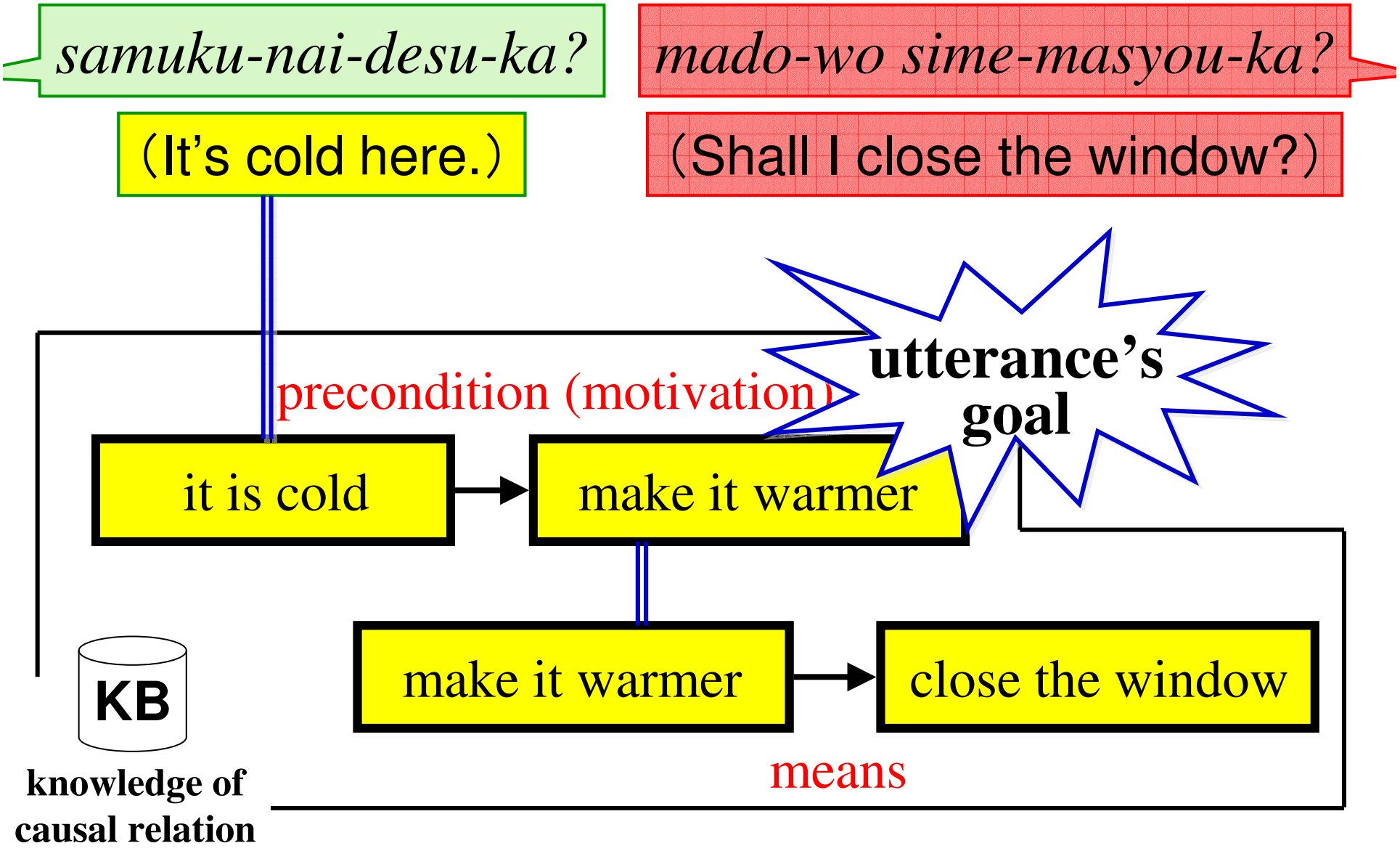
First part : introduction

- Example
- Typology of causal relations

Second part : technical issues

- Key idea & its problem
- Analyses & Experiments

Example of utterance understanding



Typology of causal relations (1/5)

- Acquire 4 types of causal relations between 2 events following [Allen 1995]

cause effect precond(ition) means

Typology of causal relations (2/5)

- *cause* relation -

non-volitional state of affairs

non-volitional state of affairs

cause ({it is sunny}, {laundry dries well})

Typology of causal relations (3/5)

- *effect* relation -

agent's volitional action

effect ({someone destroys mangrove swamps}, {flooding occurs})

non-volitional state of affairs

Typology of causal relations (4/5)

non-volitional state of affairs

precond ({someone has a driving license},
{someone drives a car})

agent's volitional action

---

agent's volitional action

means ({someone goes to the ticket office},
{someone buys a ticket})

agent's volitional action

Typology of causal relations (5/5)

Act : agent's volitional action

SOA : non-volitional state of affairs

Causal_rel(arg1, arg2)	example
cause (SOA, SOA)	***cause*** ({it is sunny}, {laundry dries well})
effect (Act, SOA)	***effect*** ({someone destroys mangrove swamps}, {flooding occurs})
precond (SOA, Act)	***precond*** ({someone has a driving license}, {someone drives a car})
means (Act, Act)	***means*** ({someone goes to the ticket office}, {someone buys a ticket})

necessary condition

sufficient conditions described later

Outline of presentation

First part : introduction

- Example
- Typology of causal relations

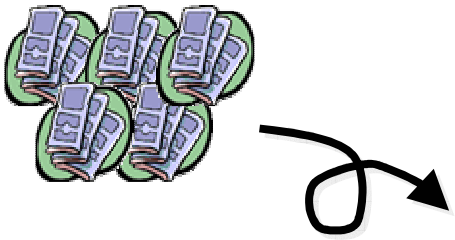
Second part : technical issues

- Key idea & its problem
- Analyses & Experiments

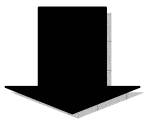
Note: this research was conducted using Japanese newspaper articles

Key idea

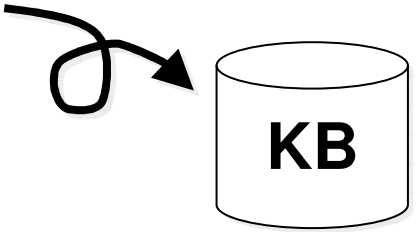
- Use connective markers as clues



The laundry dried well today because it was sunny.



cause ({it is sunny}, {laundry dries well})



Problem

- Different types of causal relations are expressed with the same marker.

The laundry dried well today because it was sunny.

➡ ***cause*** ({it is sunny}, {laundry dries well})

We need to create a computational model to identify which type of causal relation can be acquired from a given sentence.

➡ ***means*** ({Mary uses a tumble dryer}, {she dries the laundry quickly})

Procedure

Step 1 Select markers.

Step 2 Evaluate marker's effectiveness.

Step 3 Identify the causal relations automatically.

Procedure

Step 1 Select markers.

Step 2 Evaluate marker's effectiveness.

Step 3 Identify the causal relations automatically.

Step 1

Step 1

- Selected ***tame*** as our target
  - Used frequently and
  - Typically used to express causal relations

marker		freq.
ga	(but)	131 164
tame	**(because)**	**76 087**
to	(when/if)	56 549
reba	(if)	48 606
nagara	(while)	13 796

kara	(because)	10 209
node	(because)	9 994
nara	(if)	7 598
tara	(if)	6 027
noni	(but)	2 917

From Nihon Keizai Shimbun issued 1990

Example sentences

Step 1

【Ex.1】

manguroubu-wo *hakaisi-ta-**tame***

mangr ACC destroy-PAST-***tame***

subordinate clause

matrix clause

suigai-ga *hasseisi-ta*

flooding-NOM occur-PAST

【Ex.2】

kippu-wo	*kau-**tame***	*kippu-uriba-ni*	*i-tta*
ticket-ACC	buy-***tame***	to ticket office	go-PAST

Example sentences

Step 1

【Ex.1】

manguroubu-wo *hakaisi-ta-**tame***

mangrove-ACC destroy-PAST-***tame***

suigai-ga *hasseisi-ta*

flooding-NOM occur-PAST

【Ex.2】

kippu-wo *kau-**tame*** *kippu-uriba-ni* *i-tta*

ticket-ACC buy-***tame*** to ticket office go-PAST

Procedure

Step 1 Select markers.

We used "tame"

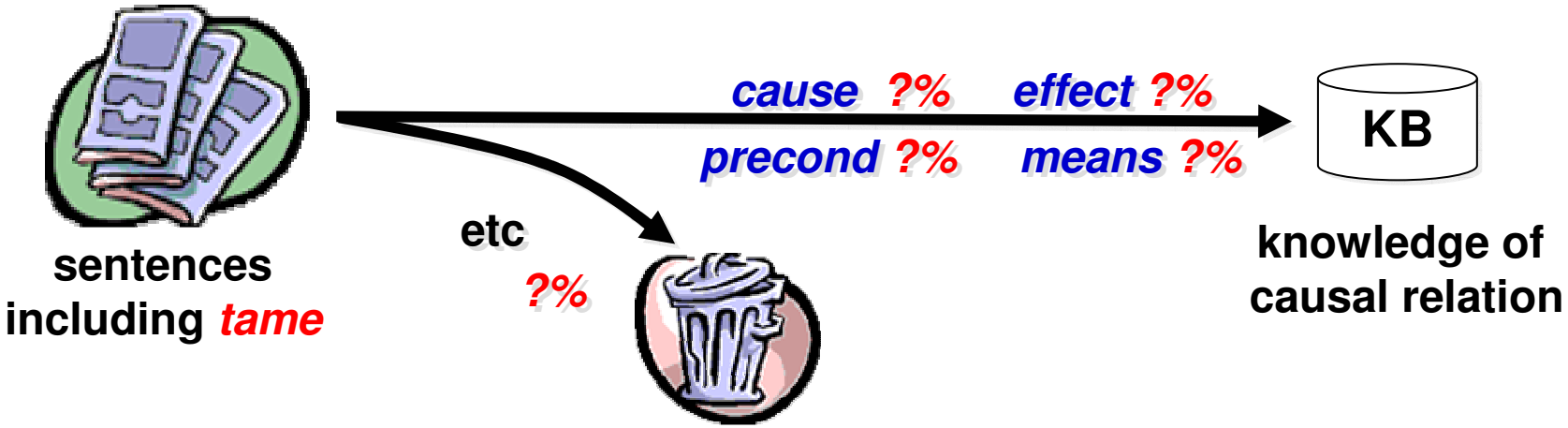
Step 2 Evaluate marker's effectiveness.

Step 3 Identify the causal relations automatically.

Step 2

Step 2

- Evaluate effectiveness of *tame*
  - How many causal relation instances are expressed by sentences including *tame*?



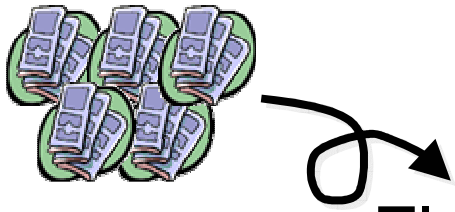
- Manually classified samples
  - Sample: about 1000 sentences including *tame*
  - Using linguistic tests

Linguistic tests (1/3)

Step 2

1. Translate sentence to two base-form items.

- Some modal information (tense, passive voice, etc.) is deleted.



Flooding occurred because mangrove swamps were destroyed.

past

passive voice

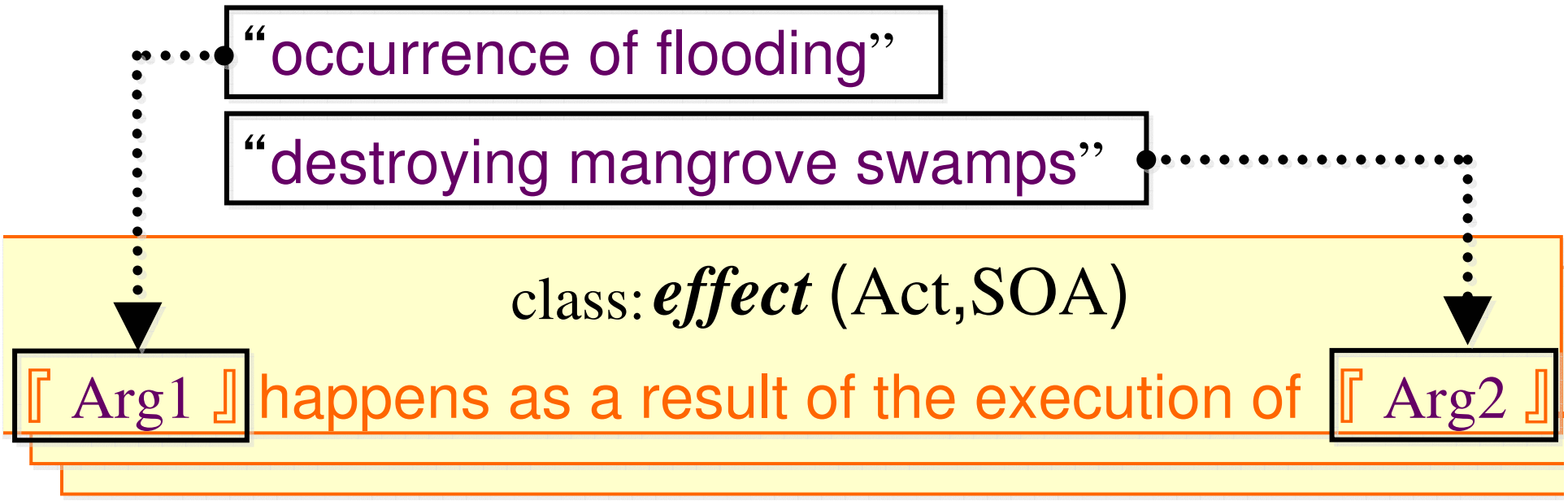
“occurrence of flooding”

“destroying mangrove swamps”

Linguistic tests (2/3)

Step 2

2. Embed items in the slot of the template to form a candidate sentence

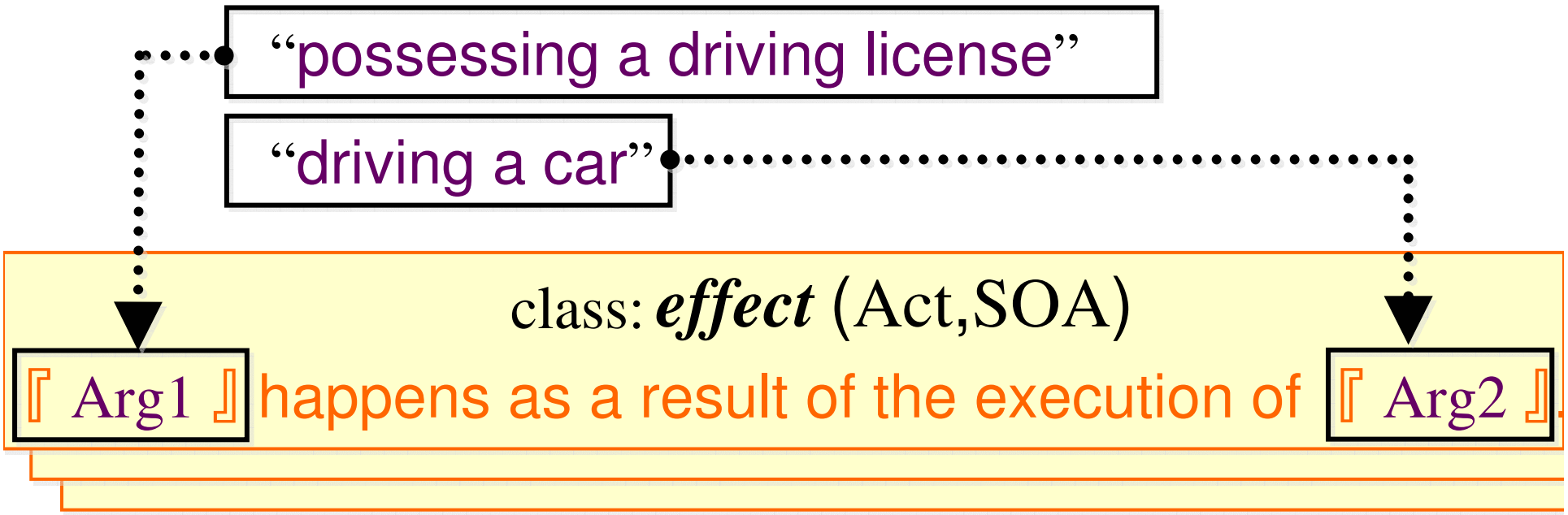


3. If the candidate sentence has the correct meaning, we identify the relation as belonging the class (***effect***)

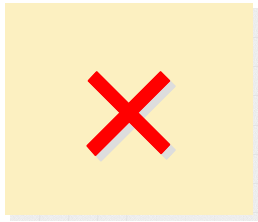
○

Linguistic tests (3/3)

Step 2

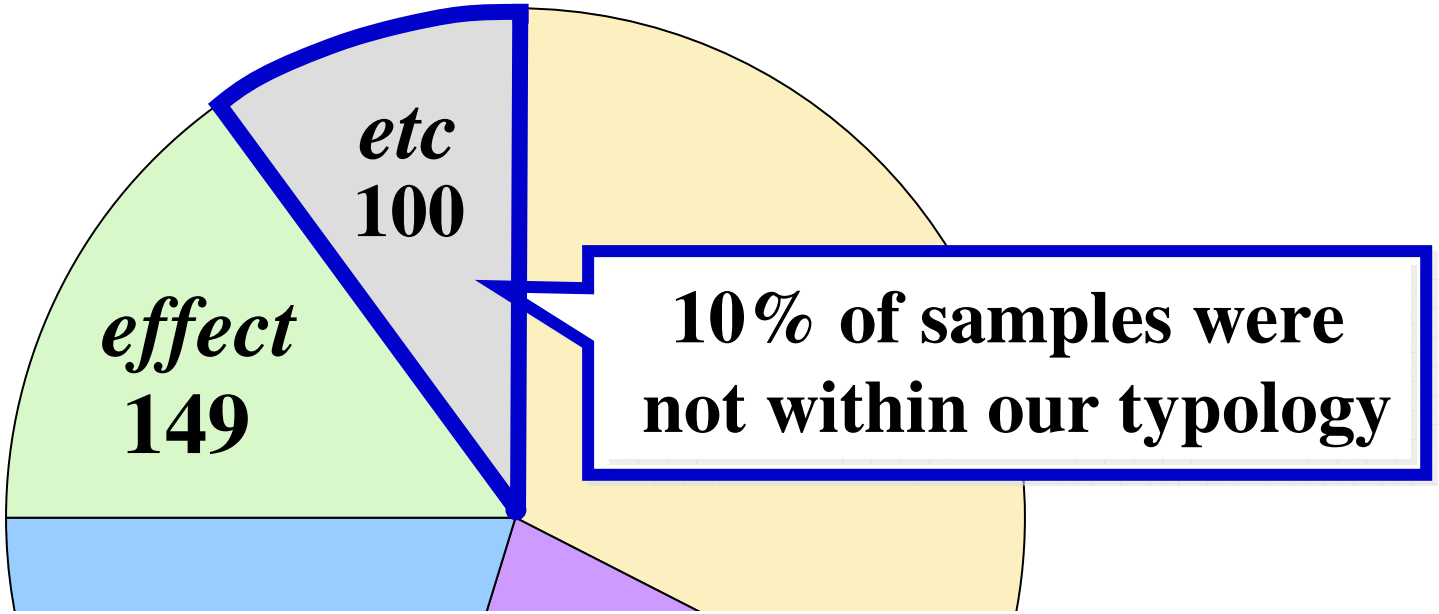


3. If the candidate sentence has the correct meaning, we identify the relation as belonging the category.



Results

Step 2



【Ex.】

hard disk-wo *kairyousuru-**tame***

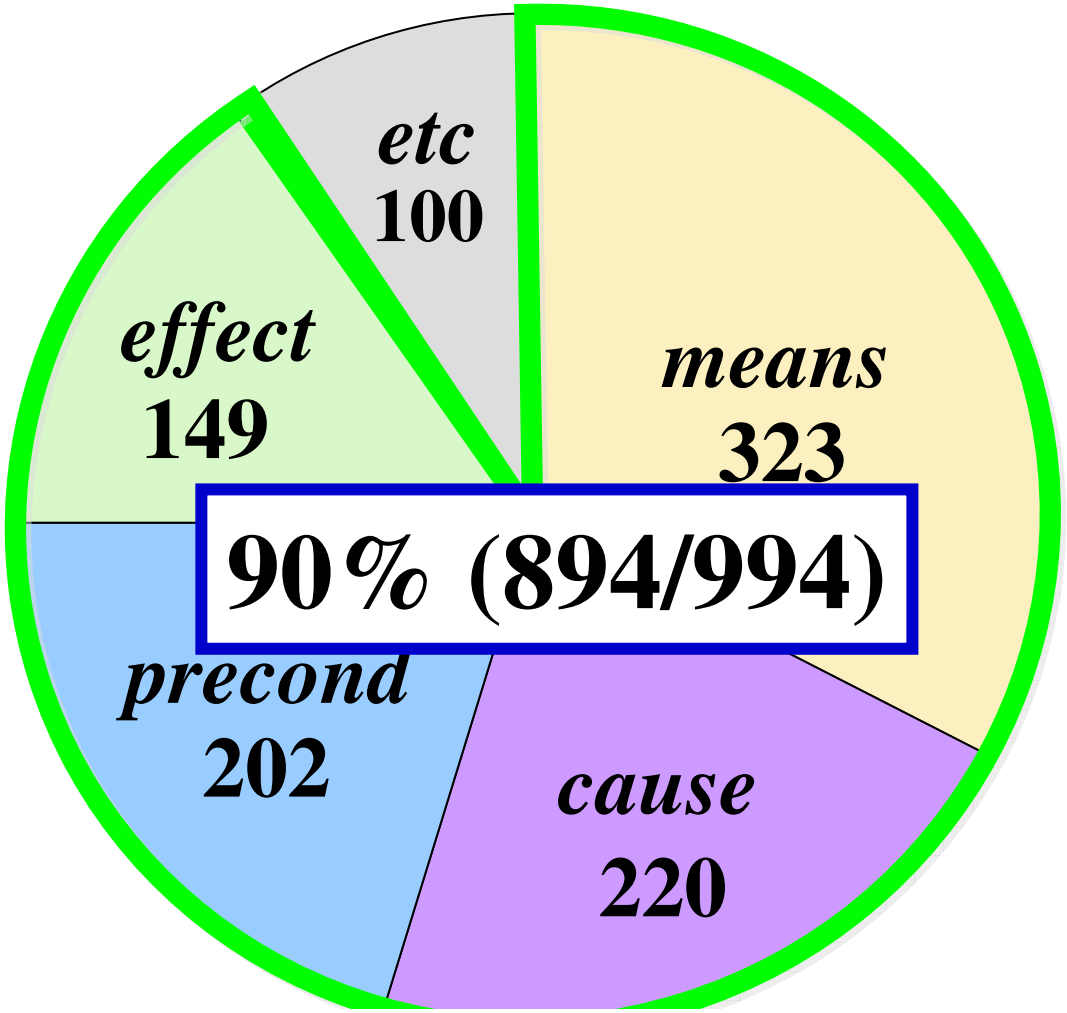
hard disk-ACC improve- ***tame***

kaihatsu kyousou-ga *gekikasuru*

development competition-NOM intensify

Results

Step 2



- Each type of causal relation appeared with relatively similar frequency

Procedure

Step 1 Select markers.

We used "tame"

Step 2 Evaluate marker's effectiveness.

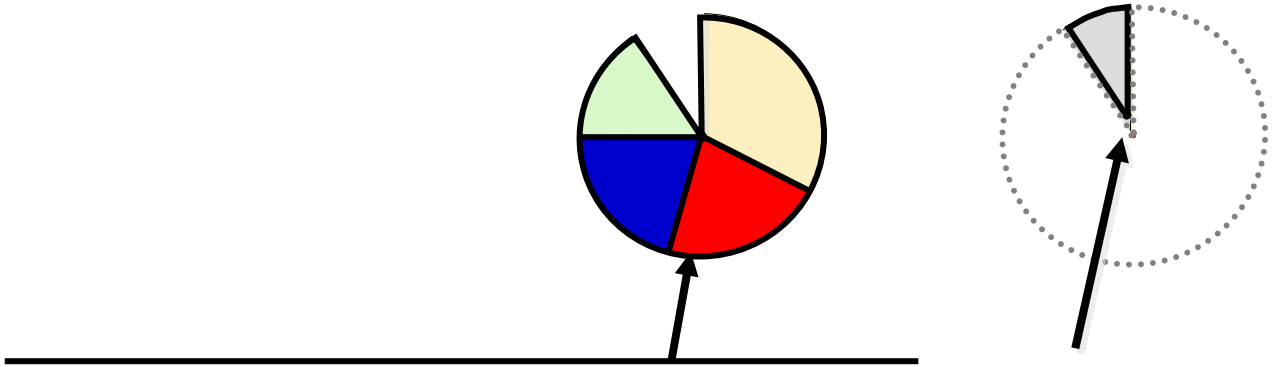
applicable of 90% of samples

Step 3 Identify the causal relations automatically.

Step 3

Step 3

- Identify the causal relations automatically
  - Experiment using a machine learning approach



  - Setting
    - 5 classes: ***cause, effect, precond, means*** and ***etc***
    - SVMs / One vs. Rest method was applied
    - Features

Features

Step 3

class		descriptions
Verb	EDR	Four conc
	ALT-J/E	A se dicti
	Goi-Taikei	Verl
Case	Marker	*"ga*
	Element	The
Modality	**Tense**	*"-ru*
	Aspect	*"-te*
	Voice	*"-re*
	Potential	*"-de*
	Negative	*"-n*
Subject		Whe
Event type		Agent's volitional action or non-volitional state of affairs

Verb = [go: transitive, ・・]

I went to a ticket office.

Tense = [past]

Event type = [Act]

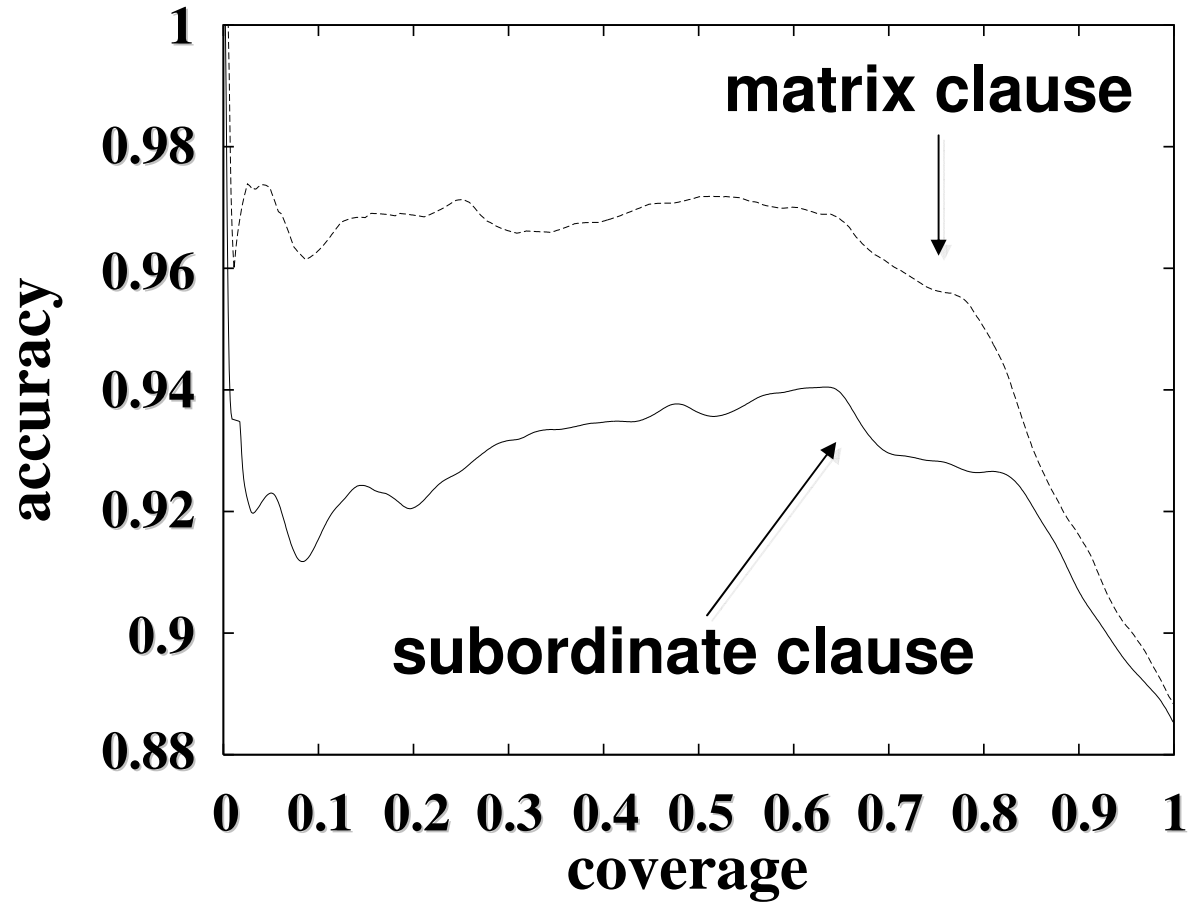
Features: (Event type estimation) Step 3

- Act: agent's volitional action
- SOA: non-volitional state of affairs
- Create

Causal_rel(arg1, arg2)
cause (SOA, SOA)
effect (Act, SOA)
precond (SOA, Act)
means (Act, Act)

necessary condit

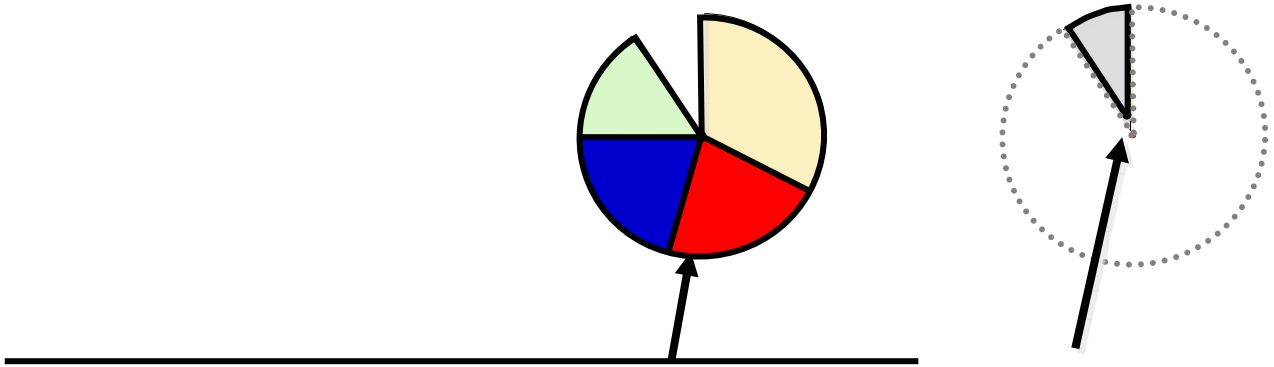
Event ty



Step 3

Step 3

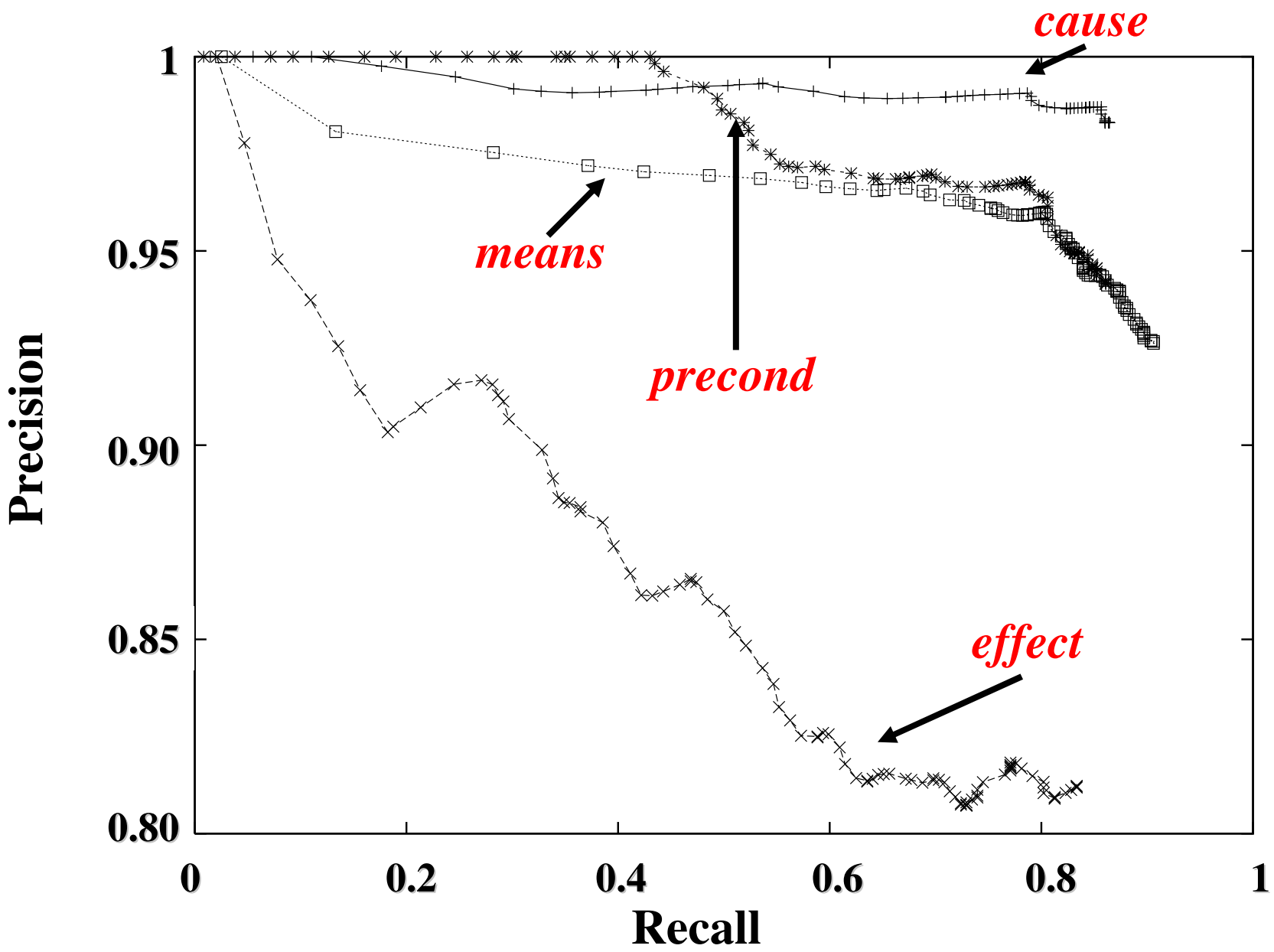
- Identify the causal relations automatically
- Experiment using a machine learning approach



  - Setting
    - 5 classes: ***cause, effect, precond, means*** and ***etc***
    - SVMs / One vs. Rest method was applied
    - Features
    - Training: about 1000 sentences
    - Evaluation: about 1000 new sentences

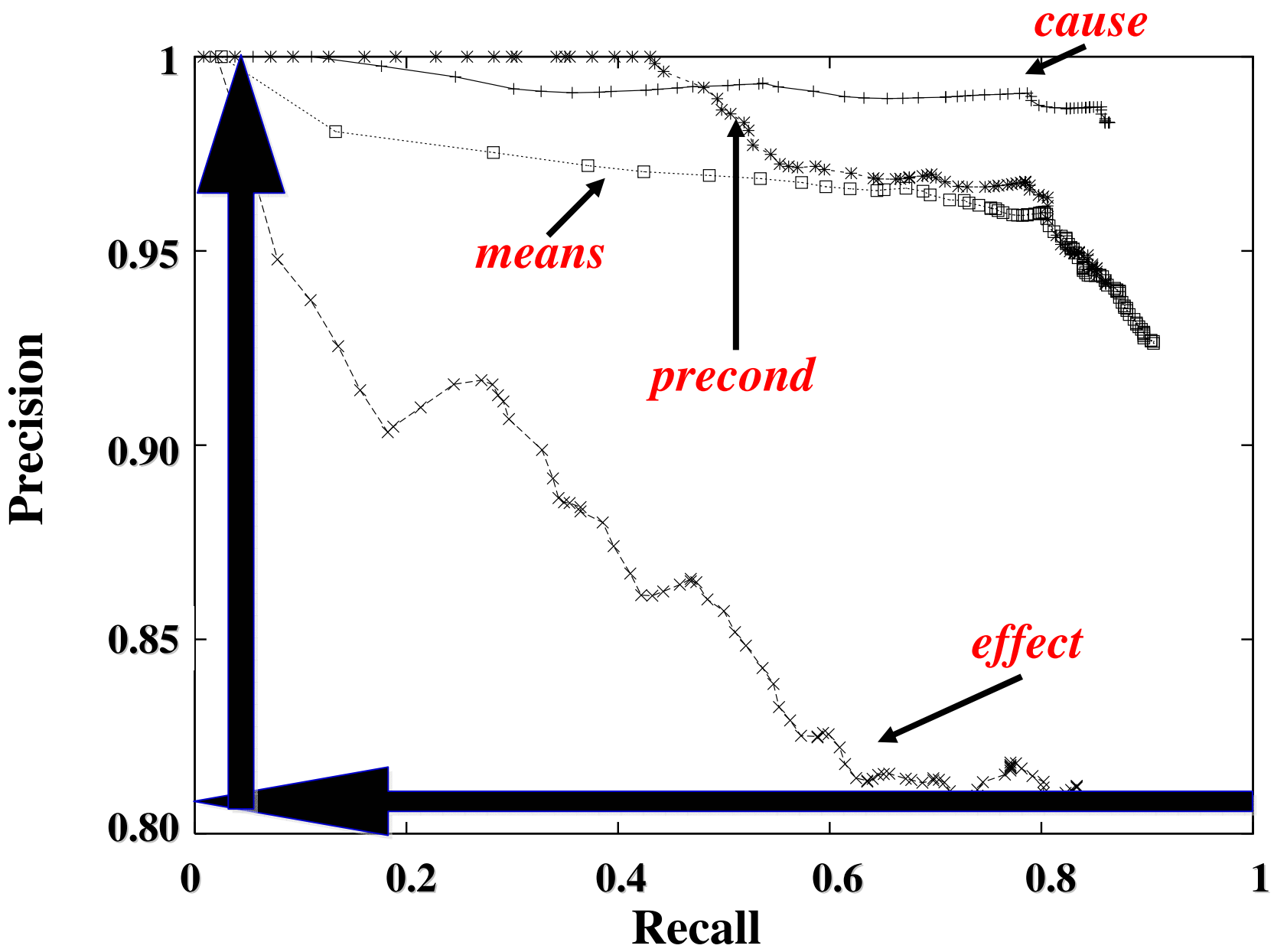
Results

Step 3



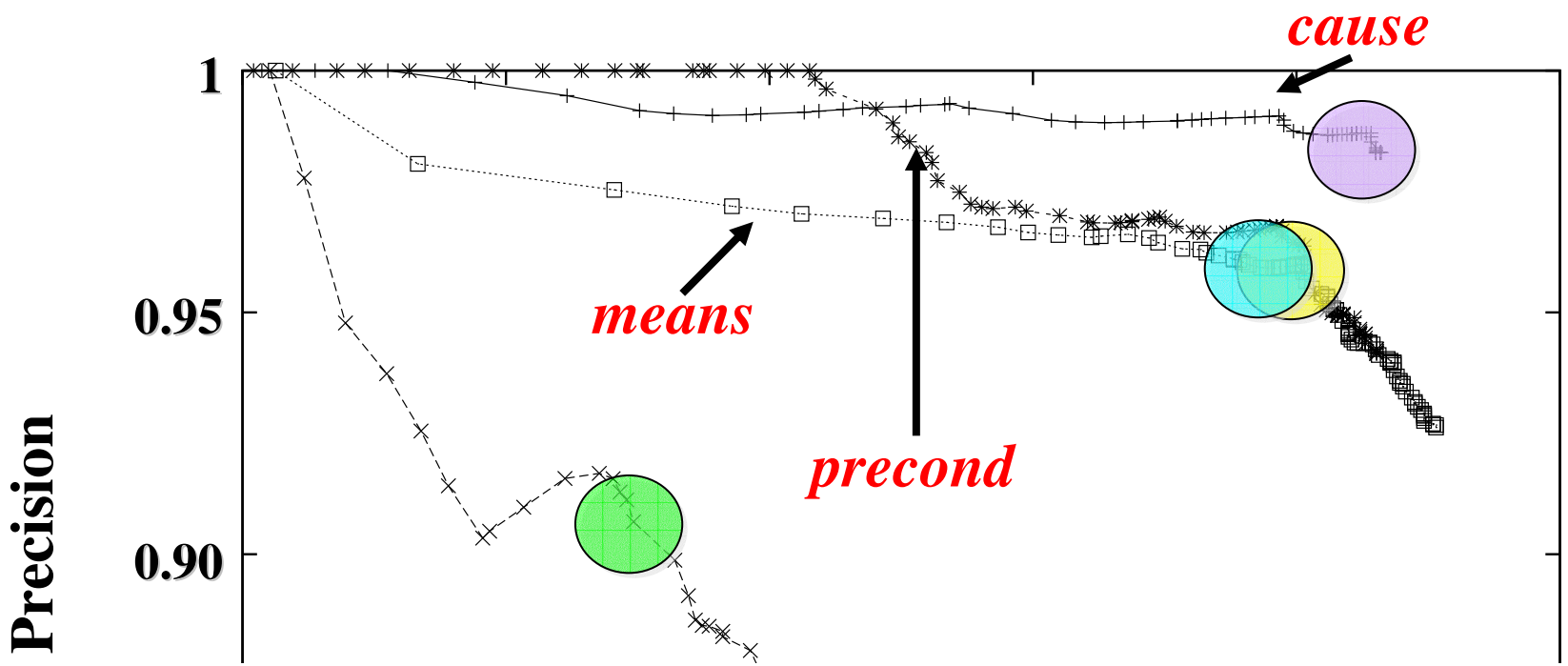
Results

Step 3



Results

Step 3



We expect to be able to acquire over 27,000 causal relation instances from one year of newspaper articles (i.e. 1.8% of all sentences).

Procedure

Step 1 Select markers. ***We used "tame"***

Step 2 Evaluate marker's effectiveness.

applicable to 90% of samples

Step 3 Identify the causal relations automatically.

Accuracy was sufficiently high

Examples of causal relation instances

cause ({temperature stays high},
{coat sales are down})

precond ({house becomes cramped},
{someone moves into a larger house})

means ({someone undertakes intensive highway},
{someone reduces the number of crashes})

Examples of causal relation instances

- Abstraction
- Ellipsis & pronoun resolution

cause ({the weather has been bad recently},
{the plan is 5 days behind schedule})

organization

numerical value

location

precond ({JBM is continuing its move into Asia},
{(ϕ) decides to produce it there})

ellipsis

pronoun

Conclusion

- We study automatic knowledge acquisition of causal relations from document collections.

Findings

Procedure

Step 1 Select markers. *we used "tame"*

Step 2 Evaluate marker's effectiveness.

applicable to 90% of samples

Step 3 Identify the causal relations automatically.

accuracy was sufficiently high

34

Future work

Examples of causal relation instances

- Abstraction
- Ellipsis & pronoun resolution

cause ({the weather has been bad recently},
{the plan is 5 days behind schedule})

organization | numerical value | location

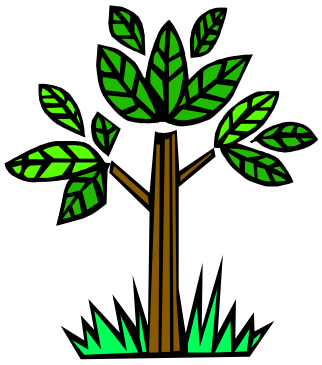
precond ({JBM is continuing its move into Asia},
{(ϕ) decides to produce it there})

ellipsis | pronoun

36

References

- [Allen1995] J. F. Allen. *Natural Language Understanding*. The Benjamin/Cummings Publishing. 1995.

- [Lenat1995] D. Lenat. Cyc: A large-scale investment in knowledge infrastructure. *Communications of the ACM*, 38(11), 1995.

- [Stork1999] D. G. Stork. Character and document research in the open mind initiative. In *Proc. of international conference on Document Analysis and Recognition*, 1999.