2005/6/29   **<aclws2005>**

Investigating the characteristics of the causal relations in Japanese text

Takashi Inui

Manabu Okumura

Tokyo Institute of Technology

Background & research goal (1/2)

Background   It rains hard ➡ Flooding occurs

- Some applications (e.g. dialog system and QA system) need inference rules or knowledge of causal relation.
- A large amount of text documents are used as resources in causal knowledge acquisition system
                                [Girju02，Torisawa03，Inui04].

Problem

- The characteristics of causal relation expressions in text is **unclear**.
  ➡ Low knowledge acquisition performance

Background & research goal (2/2)

Research goal

- Investigating the characteristics of causal relation expressions in text

To do this,
- We created a corpus tagged with causal relation information.
- We used Japanese text.

Table of contents

- Background & research goal
- Corpus construction

- Investigation
  - Suggestion for developing knowledge acquisition methods for causal relation

---

Table of contents

- Background & research goal
- Corpus construction

- Investigation
  - Suggestion for developing knowledge acquisition methods for causal relation

---

Target expressions

- We treat a variety of expressions with wide coverage.

Our target expressions
- explicit (e.g. [Altenberg 84])
  Flooding occurs because it rains hard.
        VP   explicit cue phrase marker   VP
- implicit
  It rains hard, and flooding occurs.
- noun phrase
  Flooding occurs because of heavy rains.
                                                NP
- between sentences
  S1: It rains hard. *S2:* Flooding occurs.

New targets

Table of contents

- Background & research goal
- Corpus construction
  - Target expressions
  - Tags
  - Annotation criteria
- Investigation
  - Suggestion for developing knowledge acquisition methods for causal relation

Tags (1/3)

- We use 4 tags
  - For events: *mod* (modifier) *head*
  - For causal relations between two events: *causal_rel*
  - For cue phrase markers: *marker*

- Tagging to *bunsetsu*-phrase (BP) chunk units
  BP: fundamental units in Japanese

  Ex.   BP

  It rains hard.   *ooame-ga*   *fu-tta*

  heavy rain-NOM   fall-PAST

Tags (2/3)

- Annotation scheme for events
  - Similar to PropBank [Palmer 05]
  - An event = one head BP + some modification BPs

  Ex.   *mod*   *head*

            ooame-ga   *fu-tta*

        heavy rain-NOM   fall-PAST

        It rains hard.

Tags (3/3)

- Annotation scheme for causal relations

  Ex.

  ooame-ga   *fu-tta*   *tame*   *kawa-ga*   *zousui-sita*

  heavy rain-NOM fall-PAST   because   river-NOM   rise-PAST
          cause   effect

 Flooding occurs because it rains hard.
        effect   cause

Tags (3/3)

- Annotation scheme for causal relations

                                        causal_rel

  Ex.
    mod cause *head* cau *marker* *mod* effect   *head* effect

  ooame-ga *fu-tta* *tame*   *kawa-ga*   *zousui-sita*

  heavy rain-NOM fall-PAST because river-NOM   rise-PAST
        cause   effect

      Flooding occurs because it rains hard.
          effect   cause

Table of contents

- Background & research goal
- Corpus construction
  - Target expressions
  - Tags
  - Annotation criteria
- Investigation
  - Suggestion for developing knowledge acquisition methods for causal relation

Annotation criteria (1/2)

- Our criteria are based on linguistic templates.
  - ◆ The templates have two slots and *a explicit cue phrase marker*.

slot → [effect] happened *as a result of* the fact that
     → [cause] happened.

Example of linguistic templates

The judgements process

1. Embedding two events in the [cause] and [effect] slots.
2. If the template is syntactically and semantically correct, a causal relation is supposed to hold between two events.

Annotation criteria (2/2)

- The alternative templates
  - ◆ One adverb (**often, usually, always**) is included.
  - ◆ Introducing an attribute (*necessity / chance*) to causal relations

- The alternative template   ➡   *necessity* relation
  The basic template   ➡   *chance* relation

[effect] **often** happened *as a result of* the fact that [cause] happened.

Table of contents

- Background & research goal
- Corpus construction
- Investigation

Table of contents

- Background & research goal
- Corpus construction
- Investigation

Table of contents

- Background & research goal
- Corpus construction
- Investigation
  - ◆ Summary of our corpus
  - ◆ Point 1: Cue phrase markers
  - ◆ Point 2: Syntactic categories
  - ◆ Point 3: The positions of events

Summary of our corpus (1/2)

- 750 newspaper article documents in [Mainichi95]
- Three annotators
- Reading docs, detecting causal relation and tagging

annotator	causal relation instances
A	2014
B	1587
C	1048

- The difference is mainly caused by missing judgements.
  - ◆ In particular, for the implicit expressions, annotators tend to miss the judgements.

Summary of our corpus (2/2)

■ Inter-annotator agreement

	#agreement	*necessity*	*chance*
no-agreement {	one (*A, B, C*)	1253(**0.64**)	997(**0.76**)
agreement {	two (*AB,BC,AC*)	429(**0.22**)	250(**0.19**)
	three (*ABC*)	270(**0.14**)	64(**0.05**)
	two & three	699(**0.36**)	314(**0.24**)

Summary of our corpus (2/2)

■ Inter-annotator agreement

	#agreement	*necessity*	*chance*
no agreement {	one (*A, B, C*)	1253 **0.64** <	997(**0.76**)
	two (*AB,BC,AC*)	*Data for investigation*	250 **0.19**
agreement {	three (*ABC*)	**14** >	64 **0.05**
	two & three	699 **0.36** >	**0.24**

■ The *necessity* causal instances tend to be in *agreement* . (statistically significant)

Table of contents

■ Background & research goal
■ Corpus construction
■ Investigation
  ◆ Summary of our corpus
  ◆ Point 1: Cue phrase markers
  ◆ Point 2: Syntactic categories
  ◆ Point 3: The positions of events

Cue phrase markers

■ Proportion of instances attached with the *marker* tag

Flooding occurs | because | it rains hard.
marker

	freq.
with marker (explicit)	219
without marker (implicit)	480
	699

■ 70% instances do **not** have a explicit cue phrase marker.

Syntactic categories

■ Proportion of the syntactic categories of the **cause** event and the **effect** event in the causal instances.

category	cause	effect
VP	365	412
NP	322	269
others	12	18

■ More than half of the events are classified as VP.
■ # of NPs is comparable to # of VPs.

The positions of events (1/6)

We investigated two types.
1. Positions of both the **cause** event and the **effect** event
2. Relative positions between the **cause** event and the **effect** event in the same causal instances
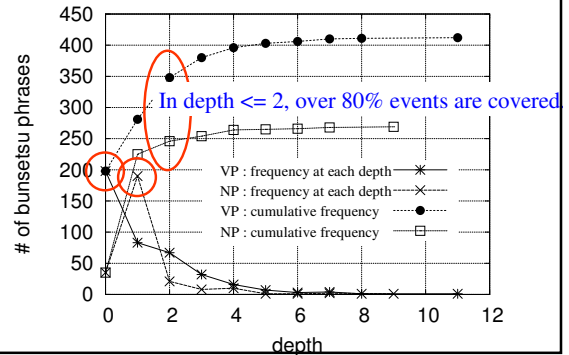
The positions of events (2/6)

1. Positions of both the **cause** event and the **effect** event
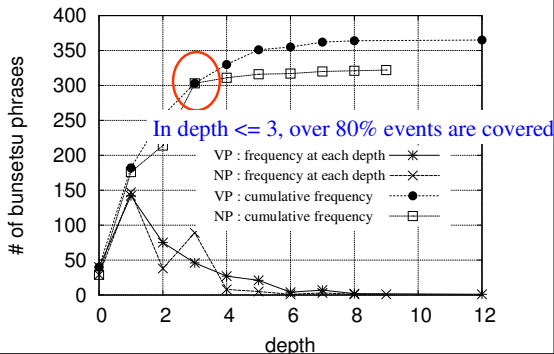
depth value = 2   2   1   0
root
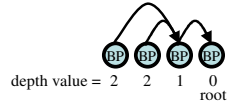
- Dependency tree structures between *bunsetus*-phrases (BPs)
  - ◆ The end of BP in a sentence is the root node, depth =0.
- We consider the depth of a BP including the *head* tag as the position of an event.

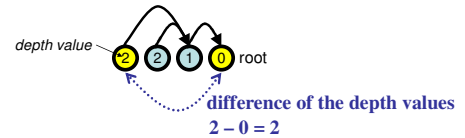## The positions of events: **effect** (3/6)

In depth <= 2, over 80% events are covered.

VP : frequency at each depth
NP : frequency at each depth
VP : cumulative frequency
NP : cumulative frequency

# of bunsetsu phrases / depth

The positions of events: **cause** (4/6)

In depth <= 3, over 80% events are covered.

VP : frequency at each depth
NP : frequency at each depth
VP : cumulative frequency
NP : cumulative frequency

# of bunsetsu phrases / depth

The positions of events (5/6)

2. Relative positions between the **cause** event and the **effect** event in the causal instances

- Difference of the depth values between the **cause** event and the **effect** event

depth value   2   2   1   0   root

difference of the depth values
2 – 0 = 2

The positions of events (6/6)

2. Relative positions between the **cause** event and the **effect** event in the causal instances

		freq.
intra-sentential	difference = 1	259
	= 2	152
	> 2	33
	no dependency	72
inter-sentential		141

Summary of investigation

- 70% of instances are expressed without markers.
- Over 50% of instances are formed by VP, but # of NPs is comparable to # of VPs.
- **cause;** depth=1,   **effect;** VP: depth=0 / NP: depth=1.
- The **cause** events directly depend on the **effect** events.

- Existing knowledge acquisition methods
  - ◆ *"VP marker VP"*
    Only some limited patterns of expressions have been treated.

Summary

- We constructed a Japanese corpus tagged with causal relation information.
- Using this corpus,
  we investigated the characteristics of causal relation expressions from some viewpoints.

- We obtain some useful suggestions for developing knowledge acquisition methods for causal relation.

Thank you!