Experiments with CST-based Multi-document Summarization

Maria Lucia R. Castro Jorge
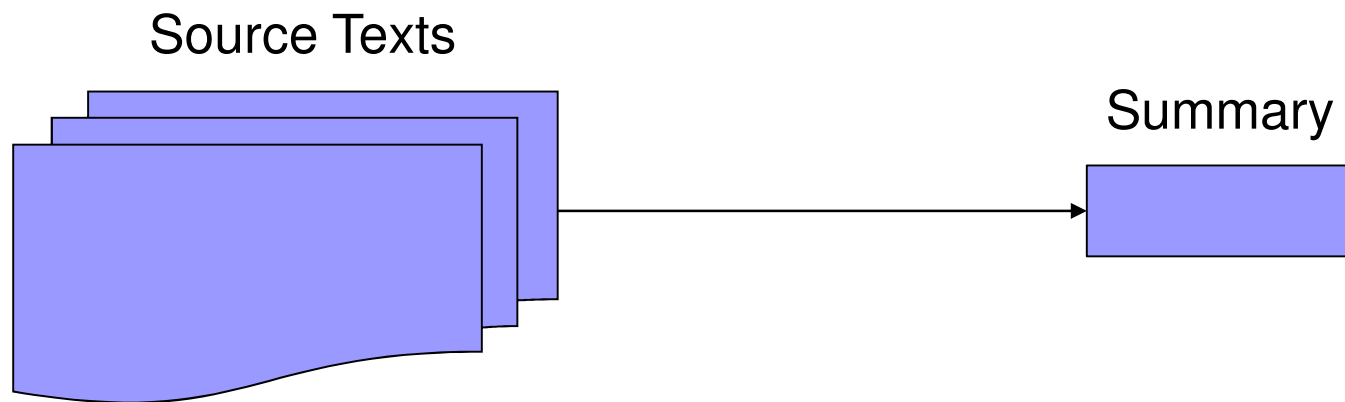mluciacj@icmc.usp.br

Thiago A.S. Pardo
taspardo@icmc.usp.br

NLP Group - University of São Paulo

Multi-document Summarization

Source Texts

Summary

- One summary from a group of texts on the same topic (Mani, 2001)

  - ➢ Relevant information according to user preferences
    - ➢ Content selection task

Content Selection Operators

(McKeown and Radev, 1995)

- **Computational artifacts**
    - process a text representation
    - produce a condensed version of it
        - Selects information considering some summarization preference
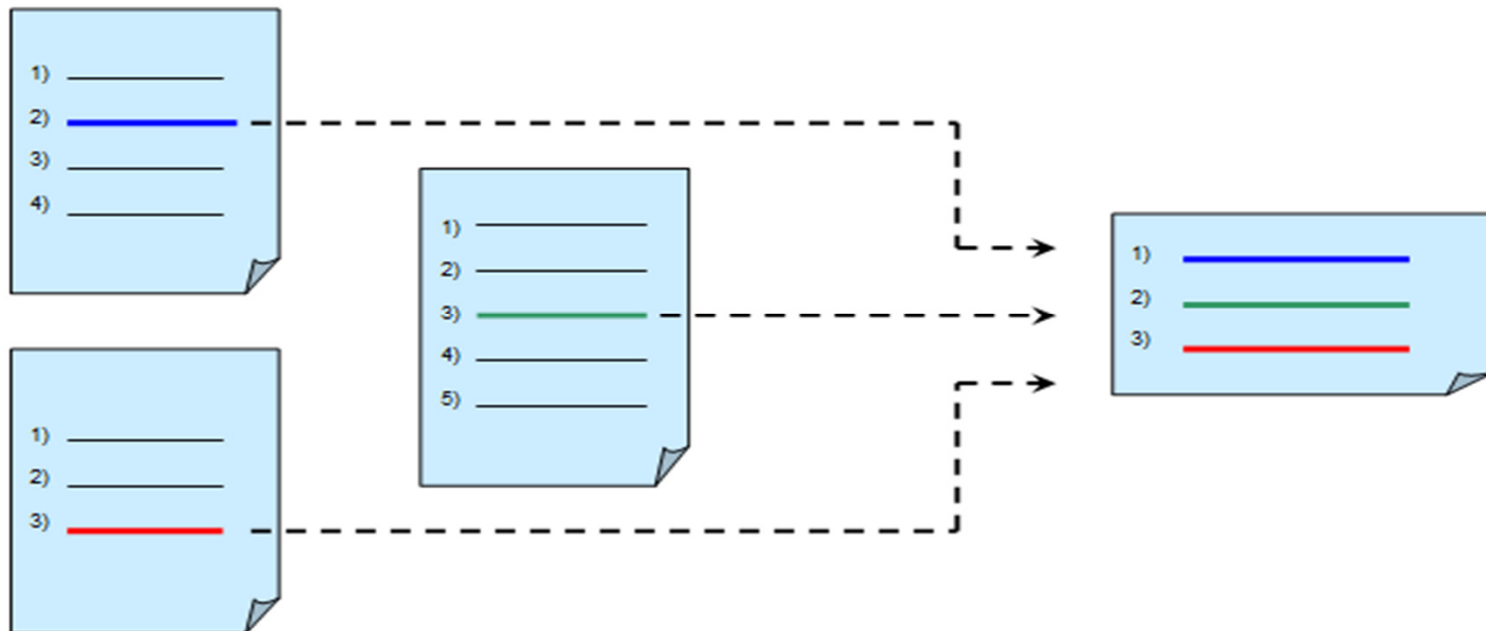
This work

- **Use of discourse information** for summarization

    ☐ Usefulness

    ☐ Impact in the task

Cross-document Structure Theory (CST)
(Radev, 2000; Zhang et al., 2003)

- <u>Discourse relations</u> among texts

- *Similarities* and *differences*
  - □ Content and writing style

- **Graph representation** of the related texts

Example

Fifteen volunteers from the French GNO Action Against Hunger (ACF) were killed in northeastern Sri Lanka today said a spokesman for the organization.

The crimes occurred in the town of Muttur, which during the last two weeks has been experiencing serious conflicts between troops of the Sri Lankan Army and the ones of the Liberation Tigers Tamil Eelam (LTTE).

"We try to send a team to Muttur to look into what is happening, but the soldiers did not allow to enter the city, which is totally blocked," said Director of ACF.

Elaboration

Fifteen volunteers from the French NGO Action Against Hunger (ACF) were killed in northeastern Sri Lanka.
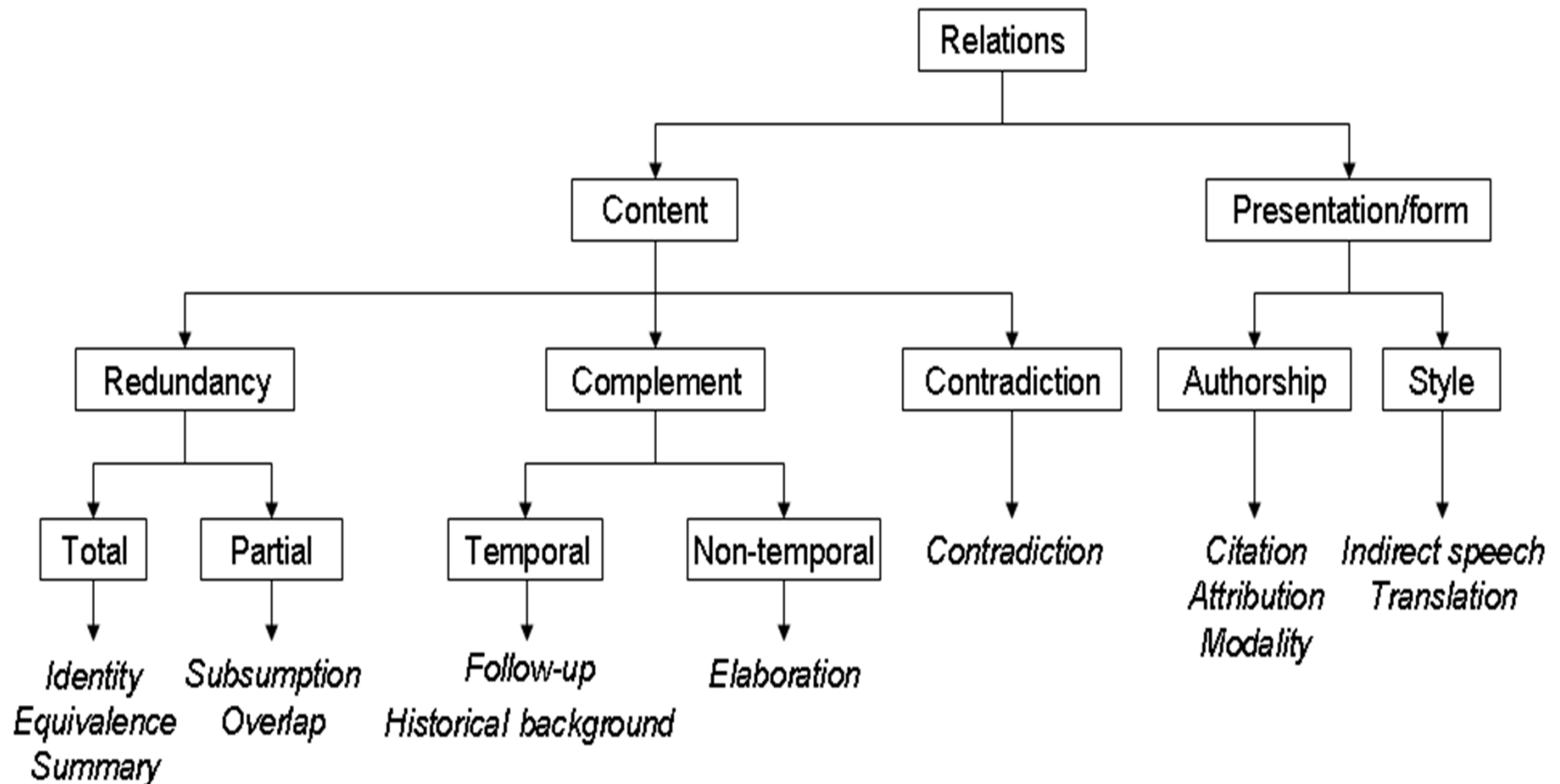
"We try to send a team to Muttur to look into what is happening, but the soldiers did not allow to enter the city, which is totally blocked," said Director of ACF.

To date, the Sri Lankan authorities did not confirm the deaths or clarified what happens in the town of Muttur.
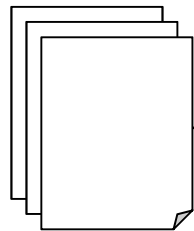
Identity

CST relation typology
(Maziero et al., 2010)

Summarization steps

Source texts

CST parsing
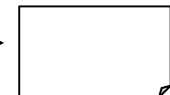(manually done
so far)

CST graph

Application of
content selection
operators

Rank of sentences

Compression
rate

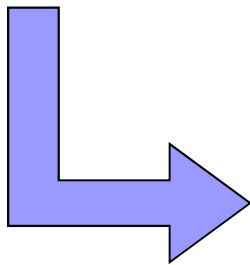Extractive
summary

1st step: CST parsing

Source texts
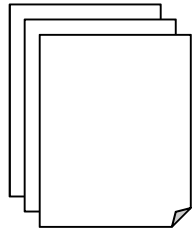
CST Graph
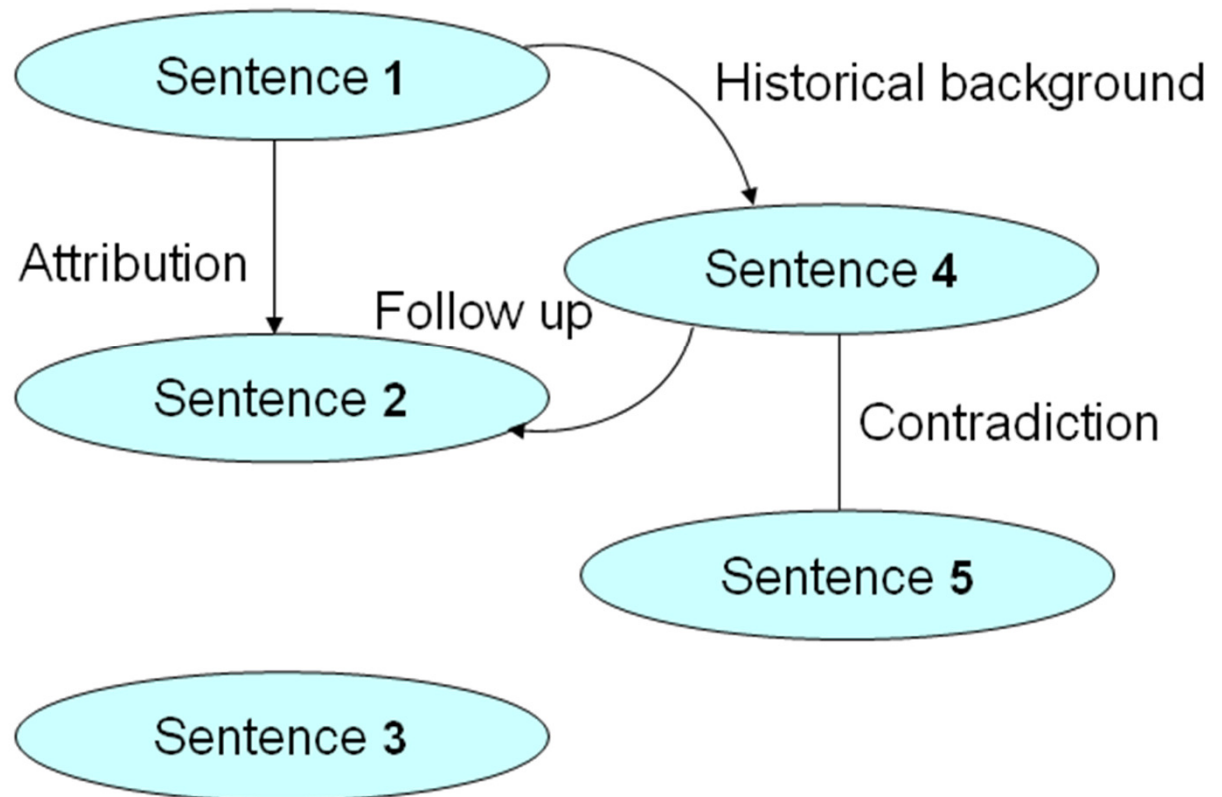
2nd step: general CS operator

CST Graph

Sentence 1 — Historical background → Sentence 4

Attribution

Follow up

Sentence 2

Contradiction

Sentence 5

Sentence 3

Initial rank

follow-up (1) Sentence 4 ← historical background

(2) Sentence 2 ←

(3) Sentence 1 ← attribution

contradiction (4) Sentence 5
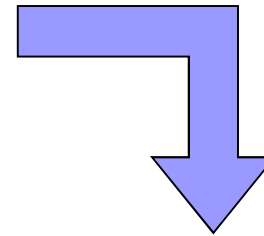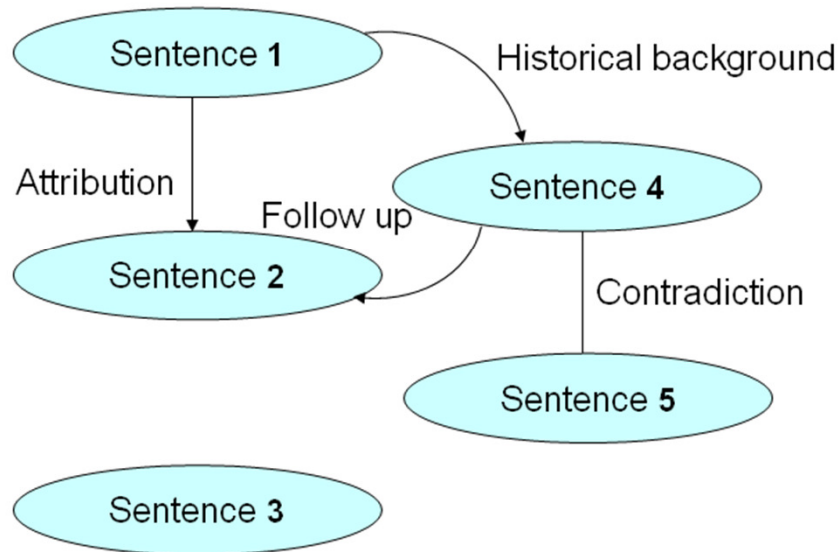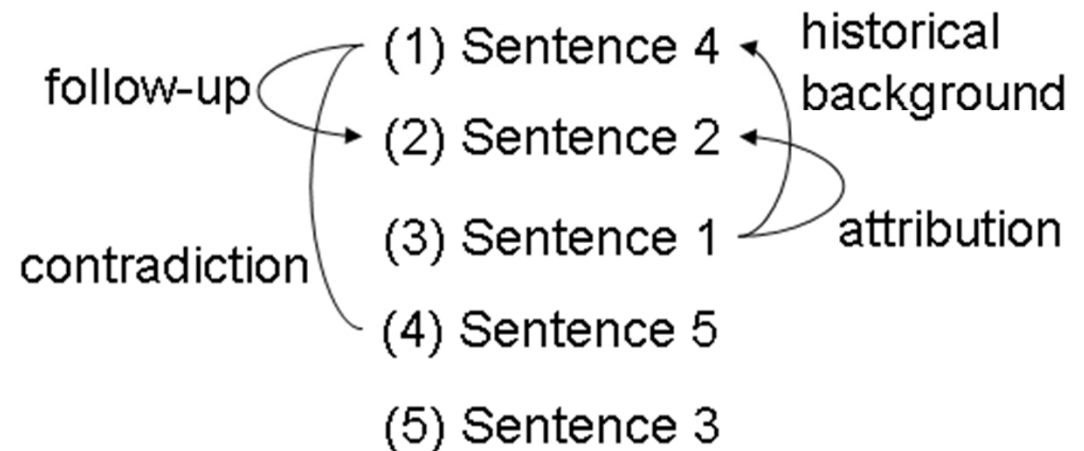
(5) Sentence 3

3rd step: preference CS operators
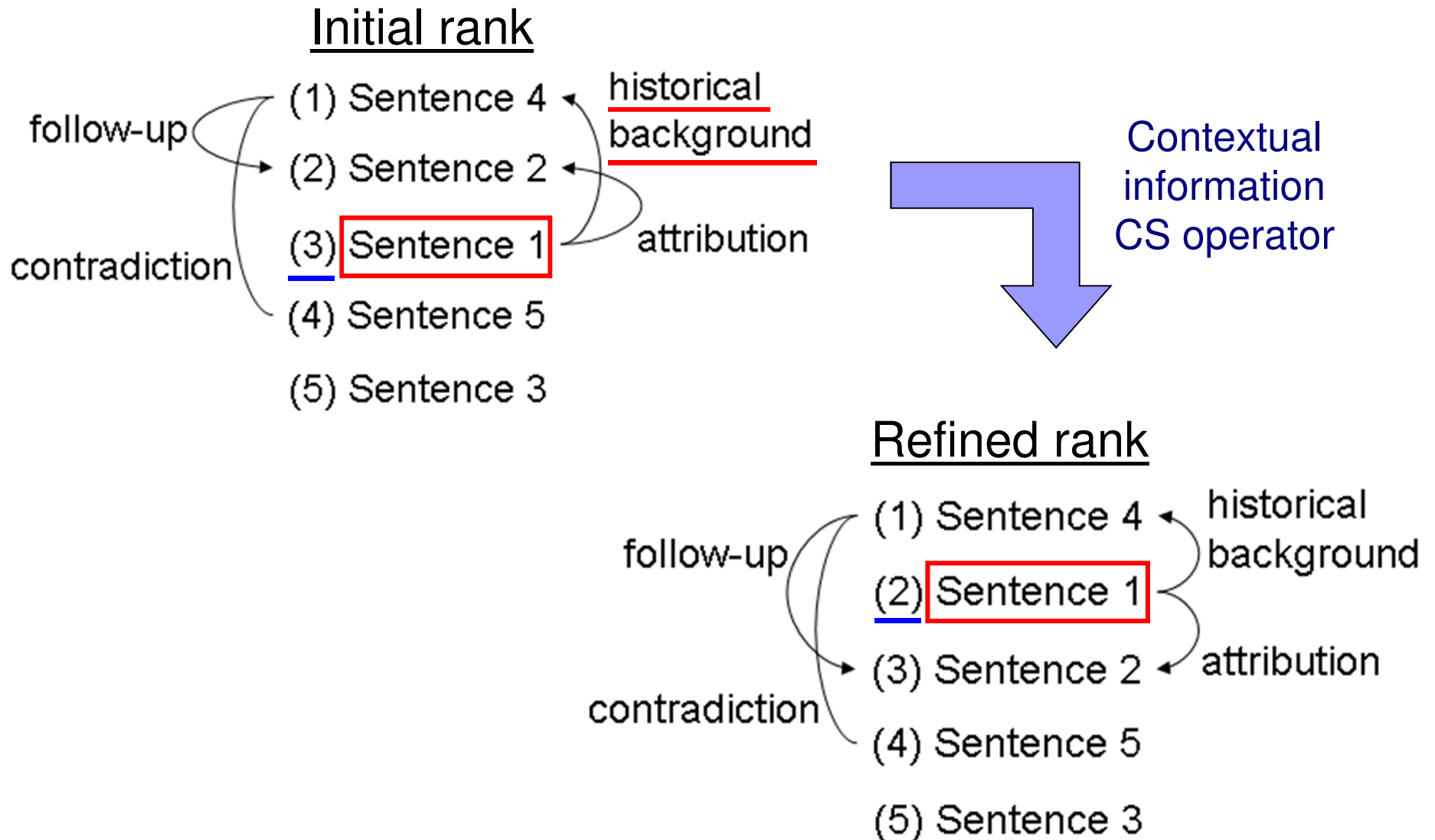
Final step: selection of sentences

Refined rank

follow-up

(1) Sentence 4 ← historical background

(2) Sentence 1

contradiction

(3) Sentence 2 ← attribution

(4) Sentence 5

(5) Sentence 3
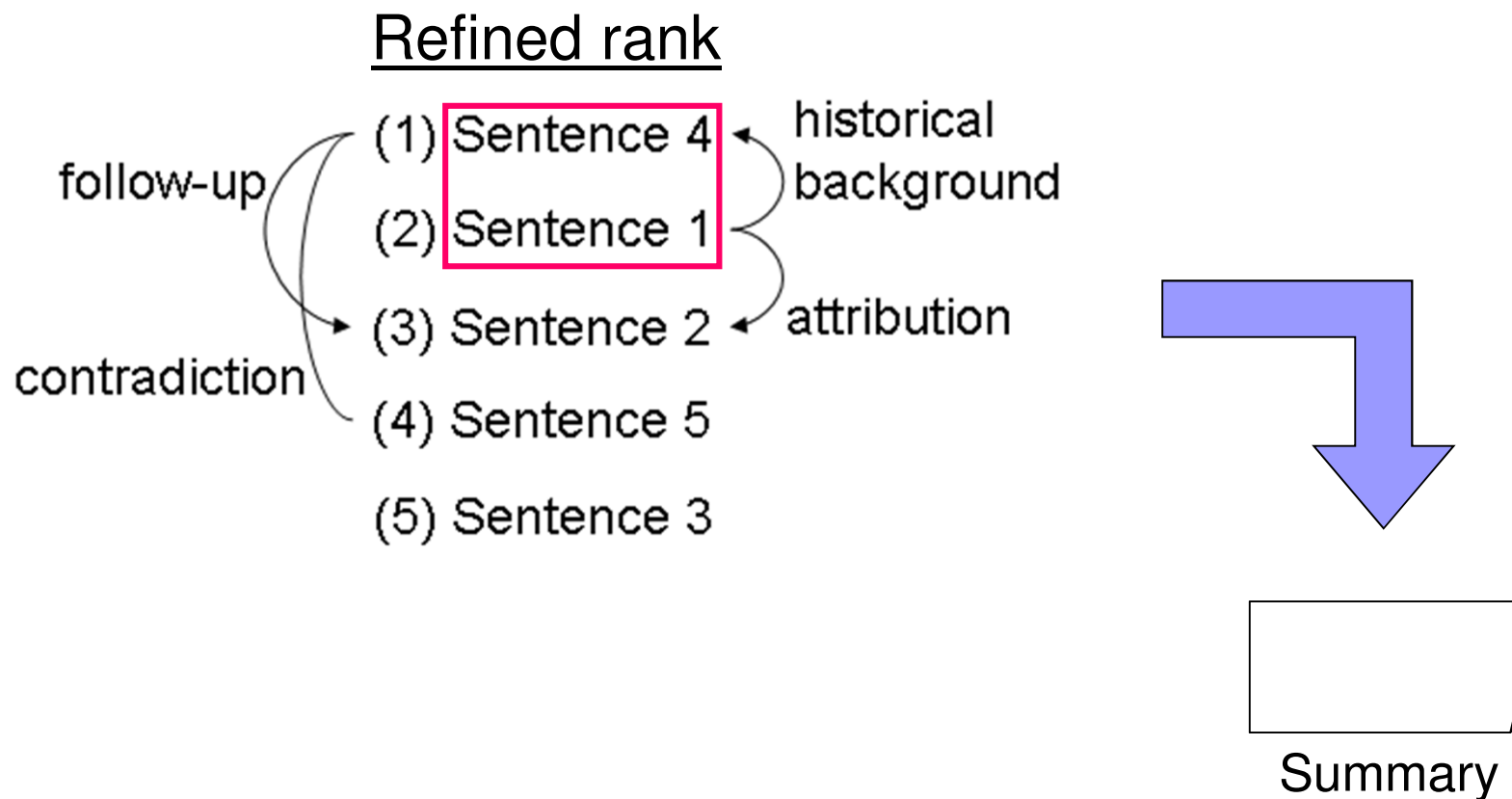
Summary

CS operators
(Jorge and Pardo, 2010)

- Each operator represents a particular summarization preference
  - General
  - Redundancy treatment
  - Context information
  - Contradictory information
  - Authorship
  - Events evolution

- Mapping of preferences to CST relations

Hybrid approach

- **CST into superficial summarizers**

  - ☐ Strategy (Zhang et al., 2002)

    New sentence score = old sentence score + number of CST relations

    - ☐ New rank of sentences

Hybrid approach

- **CST into superficial summarizers**

  - ☐ **MEAD** (Radev et al., 2000)
    - One of the most used systems
      - ☐ Criteria: sentence position and size, centroid relatedness

  - ☐ **GistSumm** (Pardo et al., 2003)
    - Simple word frequency-based summarizer

Evaluation

- **CSTNews** (Aleixo and Pardo, 2008)

  - ☐ 50 clusters of news texts in Brazilian Portuguese, with 3-4 texts per cluster

  - ☐ Human summaries, with 30% compression rate

  - ☐ Manually annotated according to CST by 4 annotators
    - Good agreement
      - ☐ Kappa ≈ 0.5-0.6
      - ☐ 80-90 percentage agreement

CSTNews

Automatic Evaluation

- ROUGE (Lin, 2004)

	Precision	Recall	F-measure
General	0.5564	0.5303	0.5356
Redundancy	0.5761	0.5065	0.5297
Contradiction	0.5503	0.5379	0.5365
Authorship	0.5563	0.5224	0.5310
MEAD with CST	0.5599	0.4988	0.5230
Event Evolution	0.5159	0.5222	0.5140
Context	0.5196	0.4938	0.4994
GISTSumm with CST	0.4945	0.5089	0.4994
MEAD	0.5242	0.4602	0.4869
GISTSumm	0.3599	0.6643	0.4599

Automatic Evaluation

- ROUGE (Lin, 2004)

	Precision	Recall	F-measure
General	0.5564	0.5303	**0.5356**
Redundancy	0.5761	0.5065	**0.5297**
Contradiction	0.5503	0.5379	**0.5365**
Authorship	0.5563	0.5224	**0.5310**
MEAD with CST	0.5599	0.4988	0.5230
Event Evolution	0.5159	0.5222	0.5140
Context	0.5196	0.4938	0.4994
GISTSumm with CST	0.4945	0.5089	0.4994
MEAD	0.5242	0.4602	0.4869
GISTSumm	0.3599	0.6643	0.4599

Automatic Evaluation

- ROUGE (Lin, 2004)

	Precision	Recall	F-measure
General	0.5564	0.5303	0.5356
Redundancy	0.5761	0.5065	0.5297
Contradiction	0.5503	0.5379	0.5365
Authorship	0.5563	0.5224	0.5310
MEAD with CST	0.5599	0.4988	0.5230
Event Evolution	0.5159	0.5222	0.5140
Context	0.5196	0.4938	0.4994
GISTSumm with CST	0.4945	0.5089	**0.4994**
MEAD	0.5242	0.4602	0.4869
GISTSumm	0.3599	0.6643	**0.4599**

Automatic Evaluation

- ROUGE (Lin, 2004)

	Precision	Recall	F-measure
General	0.5564	0.5303	0.5356
Redundancy	0.5761	0.5065	0.5297
Contradiction	0.5503	0.5379	0.5365
Authorship	0.5563	0.5224	0.5310
MEAD with CST	0.5599	0.4988	**0.5230**
Event Evolution	0.5159	0.5222	0.5140
Context	0.5196	0.4938	0.4994
GISTSumm with CST	0.4945	0.5089	0.4994
MEAD	0.5242	0.4602	**0.4869**
GISTSumm	0.3599	0.6643	0.4599

Automatic Evaluation

- ROUGE (Lin, 2004)

	Precision	Recall	F-measure
Statistically significant differences with 95% confidence	0.5564	0.5303	0.5356
	0.5761	0.5065	0.5297
Contradiction	0.5503	0.5379	0.5365
Authorship	0.5563	0.5224	0.5310
MEAD with CST	0.5599	0.4988	0.5230
Event Evolution	0.5159	0.5222	0.5140
Context	0.5196	0.4938	0.4994
GISTSumm with CST	0.4945	0.5089	0.4994
MEAD	0.5242	0.4602	0.4869
GISTSumm	0.3599	0.6643	0.4599

Human Evaluation

- **User satisfaction**
  - Random sample of texts
  - 6 evaluators
    - Coherence, cohesion, informativity and redundancy
  - Grades to the summaries
    - 0: Unacceptable
    - 1: Bad
    - 2: Regular
    - 3: Good
    - 4: Excellent

Human Evaluation

- 0: Unacceptable
- 1: Bad
- 2: Regular
- 3: Good
- 4: Excellent

Preference	Coherence	Cohesion	Informativity	Redundancy
General	3.6	3.2	3.6	1.8
Context	2.1	2.7	2.2	3.6
Authorship	3.3	2.4	3.0	2.8
Contradiction	2.4	2.7	3.7	2.5
Events evolution	2.1	2.5	3.2	2.6

Conclusions

- Use of CST allows to explore user preferences

- CST improves superficial methods

Current work

- 3 main lines

  - Automatic CST parsing

  - Machine learning of good summary CST configuration

  - Combination of CST-based CS operators with traditional summarization strategies

Experiments with CST-based Multi-document Summarization

Thank you!

www.nilc.icmc.usp.br