Machine Translation Based on Discourse Analysis

Chengqing ZONG

Institute of Automation,
Chinese Academy of Sciences

E-mail: cqzong@nlpr.ia.ac.cn **Home Page:** http://www.nlpr.ia.ac.cn/cip/english/zong.htm
Add.: No.95, Zhong Guan Cun Dong Lu, Beijing 100190, China

Outline

1. Problems
2. Related Work
3. Motivation and Model
4. Experiments
5. Conclusions and Future Work

1. Problems

- **Word is the minimum unit that carries meanings and can be utilized independently**
- **A discourse is usually a set of sentences that are well organized and definitely express some meanings and purposes.**

Example-1: 中斐两国虽然人口数量有很大差异，但共同点很多，具有发展友好合作的良好条件和基础。

Although the population of China and Fiji is very different, there is much common ground and there exist good conditions and foundations for the development of friendly cooperation.

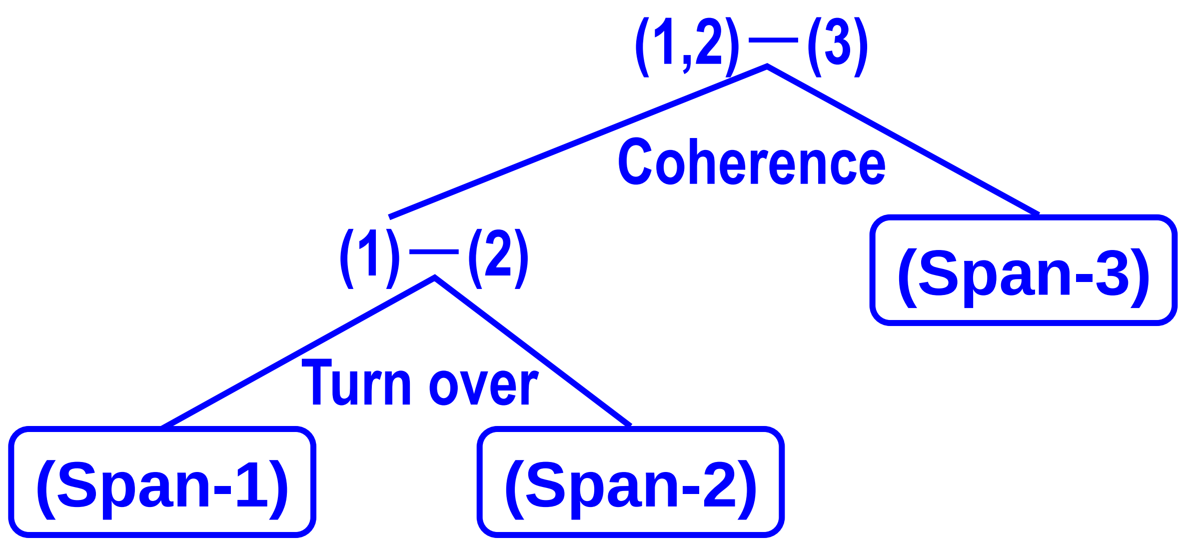
1. Problems

(1) 中斐两国虽然人口数量有很大差异(Although the population of China and Fiji is very different)， (Span-1)

(2)但共同点很多(there is much common ground)， (Span-2)

(3)具有发展友好合作的良好条件和基础(there exist good conditions and foundations for the development of friendly cooperation)。

(Span-3)



1. Problems

- **Currently, machine translation (MT) systems translate a text sentence by sentence**

  Input: 我想喝一杯饮料。

  Output: I want to have a cup of drink.

- **If the input is a discourse, some problems are often arisen**

1. Problems

① Improper or even wrong translation of the clause relations

Example-2: 他没有上过学，所以只能写到这种水平。

reason

He did not go to school, (he) can only write to this level,

Google(2013.7.24): He did not go to school, we can only write to this level.

He can only write to this level because he did not go to school.

1. Problems

②Wrong translation or missing of coreference pro-form

Example-3: 联合党党魁说，会谈已经开始，近期有望达成协议。

The leader of United Party said,

The talks have started,

It is expected to reach an agreement recently.

Google Translator(2013.7.29):

United Party leader said the talks had started recently is expected to reach an agreement.

?

1. Problems

③The tense of verbs is often inconsistent even or wrong

Example-4: 他身穿一件西装，手里拿着一本书，戴着一幅眼镜，看上去很像一个教授。

Google Translator(2013.7.29):

He was dressed in a suit, holding a book, wearing a glasses, looks like a professor.

1. Problems

In summary, we expect

(1) An MT system can correctly deal with the relations of clauses and get good translations;

(2) Discourse analysis can provide MT systems with information to properly translate coreference and anaphora;

(3) To see the translations have good coherence and cohesion.

Outline

1. Problems
2. Related Work
3. Motivation and Model
4. Experiments
5. Conclusions and Future Work

2. Related Work

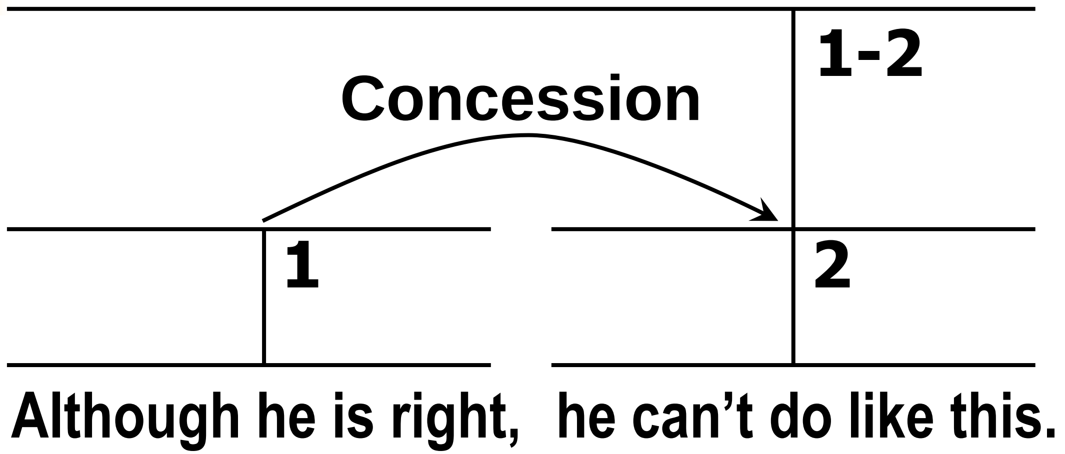
◆ RS-tree to RS-tree Translation Model

[Marcu et al., 2000]

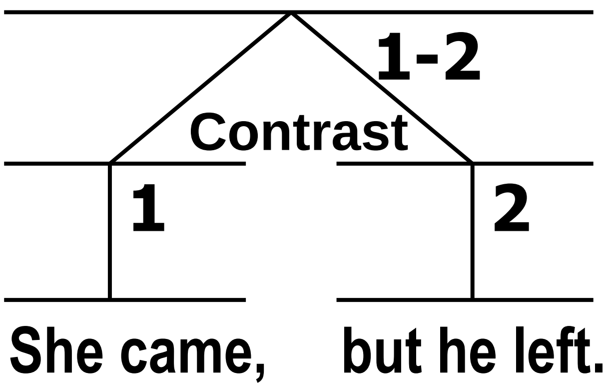
Rhetorical Structure Theory(RST)[Mann and Thompson,1987]

- A text is organized by means of relations that hold between its parts
- The structure of a text is hierarchical
- Every part of a text has a role, a function to play

2. Related Work



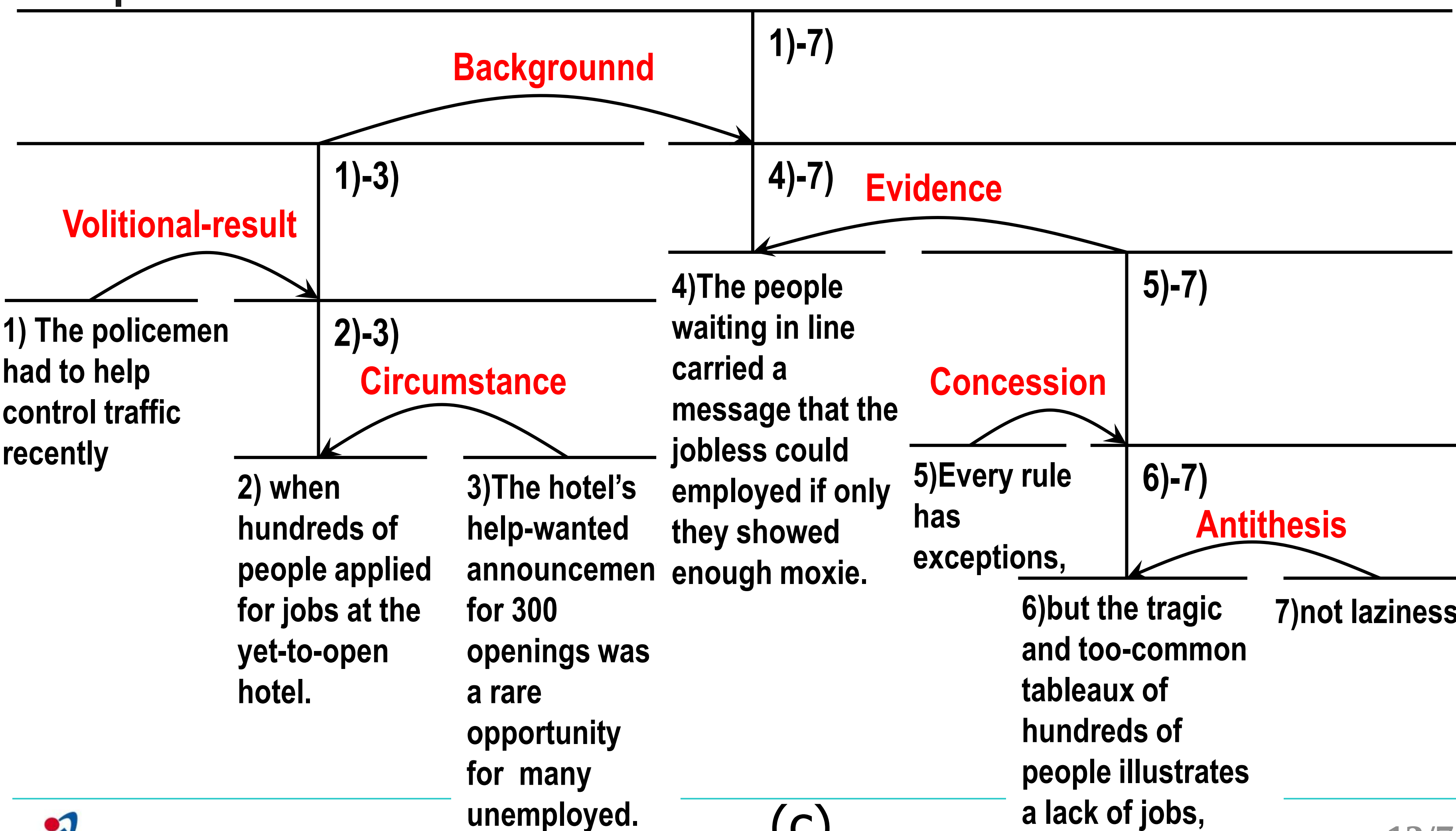
(a)



(b)

- Elementary discourse unit (EDU)
- Relations

2. Related Work



(c)

2. Related Work

In [Marcu et al., 2000], Japanese-to-English MT:

- **Japanese text is parsed into RS-tree**
  - A Japanese RS parser is employed
- **Japanese RS-tree is transferred into English RS-tree**
  - Taking elementary discourse trees as operating units
  - Operations: SHIFT, REDUCE, BREAK, CREATE-NEXT, FUSE, AWAP, ASSIGN-TYPE
- **English translation based on RS-tree**
  - A statistical translation model maps the input text into the target language

2. Related Work

The problems:

- It is unable to determine correctly how to re-package sentences into paragraphs
- How to better understand the notion of paragraph?
- How to integrate SMT model with the process to transfer RS-tree of the input text into the target language?

2. Related Work

◆ MT based on discourse division

[Shi and Chen, 2006]

$$
\begin{aligned}
P(t|s) &= P(t|s, D) && \text{(D means discourses)} \quad (1)\\
&= P(t|s, C) && \text{(C is context, and $C \subseteq D$)}\\
&= P(t|s, \langle G, E\rangle) && \text{(G is the sentence groups of C, E is the annotation information related to words in s and D.)}
\end{aligned}
$$

s means a sentence or a segment, *t* is the parsing tree of *s*.

2. Related Work

There are so many problems remained to discuss:

- How to define the boundaries of a sentence group?
- How to divide a discourse into sentence groups?
- How to determine the context?
- What tags should be annotated for a discourse?

……

2. Related Work

- **Discourse-based translation to improve coherence**
  - [Gong et al., 2011] proposes a cache-based approach to document-level translation, which utilizes three caches:

1) a dynamic cache, which stores bilingual phrase pairs from the best translation hypotheses of previous sentences in the test document;
2) a static cache, which stores relevant bilingual phrase pairs extracted from similar bilingual document pairs in the training parallel corpus;
3) a topic cache, which stores the target-side topic words related with the test document in the source-side

2. Related Work

The problems:

- How to utilize the information on relation of clauses to get the good translation structure?
- How to reflect the document divergence during training and dynamically adjust cache weights according to different documents?
- How to improve the coherence of the translation?

2. Related Work

◆ Modeling Lexical Cohesion for Document-Level MT

[Xiong et al., 2013] proposes three different models to capture lexical cohesion for document-level MT:

(a) a direct reward model to reward translation hypotheses

(b) a conditional probability model to measure the appropriateness of using lexical cohesion

(c) a mutual information trigger model to consider the lexical cohesion relation

They integrate the three models into hierarchical phrase-based machine translation

2. Related Work

The problems:

- How to construct the lexical chains?
- How to more correctly measure the appropriateness of the occurrence of a lexical cohesion device in a sentence?
- Similar to the work of word disambiguation, there are so many problems remained to further study.

2. Related Work

◆ Evaluating Text Coherence Using Discourse Relations

[Lin et al., 2011] presents a model to represent and assess the discourse coherence of text

[Wong and Kit, 2012] proposes the utilization of lexical cohesion to facilitate evaluation of machine translation at the document level

[Gimenez et al., 2010]proposes an approach to widening the scope of current automatic evaluation measures from sentence to document level

……

2. Related Work

In summary, the existing work

- more focuses on the lexical coherence of translation, but the metric BLEU can't reflect the different coherence of a text translation;
- doesn't yet implement the united MT framework based on discourse analysis, especially, the structure and semantic information including the relations of clauses in a text is not fully utilized.
- For the Chinese-to-foreign language translation, there is no RS-parser available

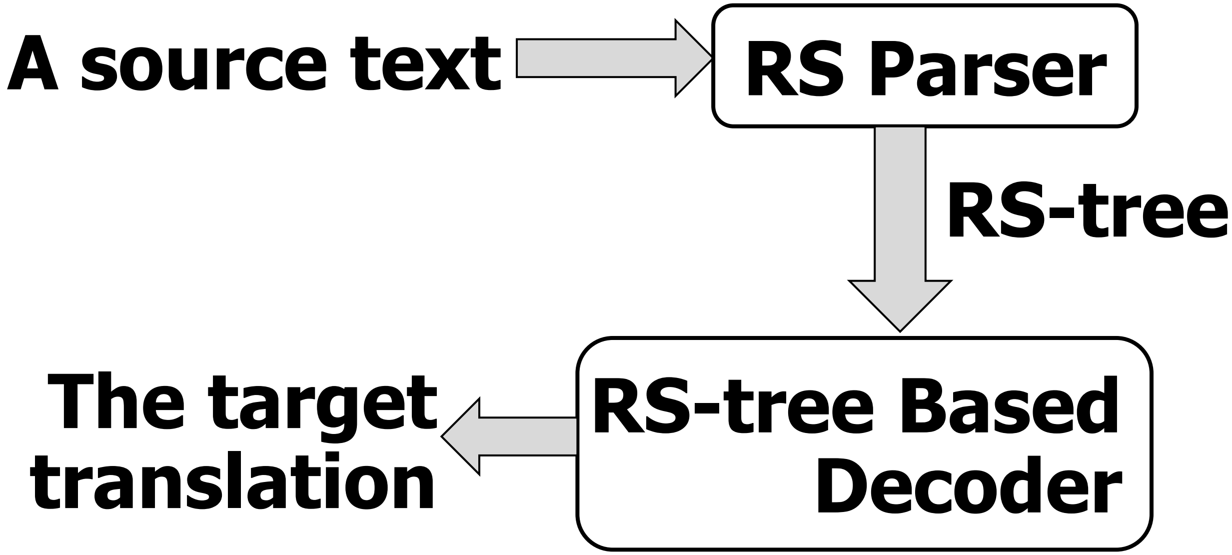
Outline

1. Problems
2. Related Work
3. Motivation and Model
4. Experiments
5. Conclusions and Future Work

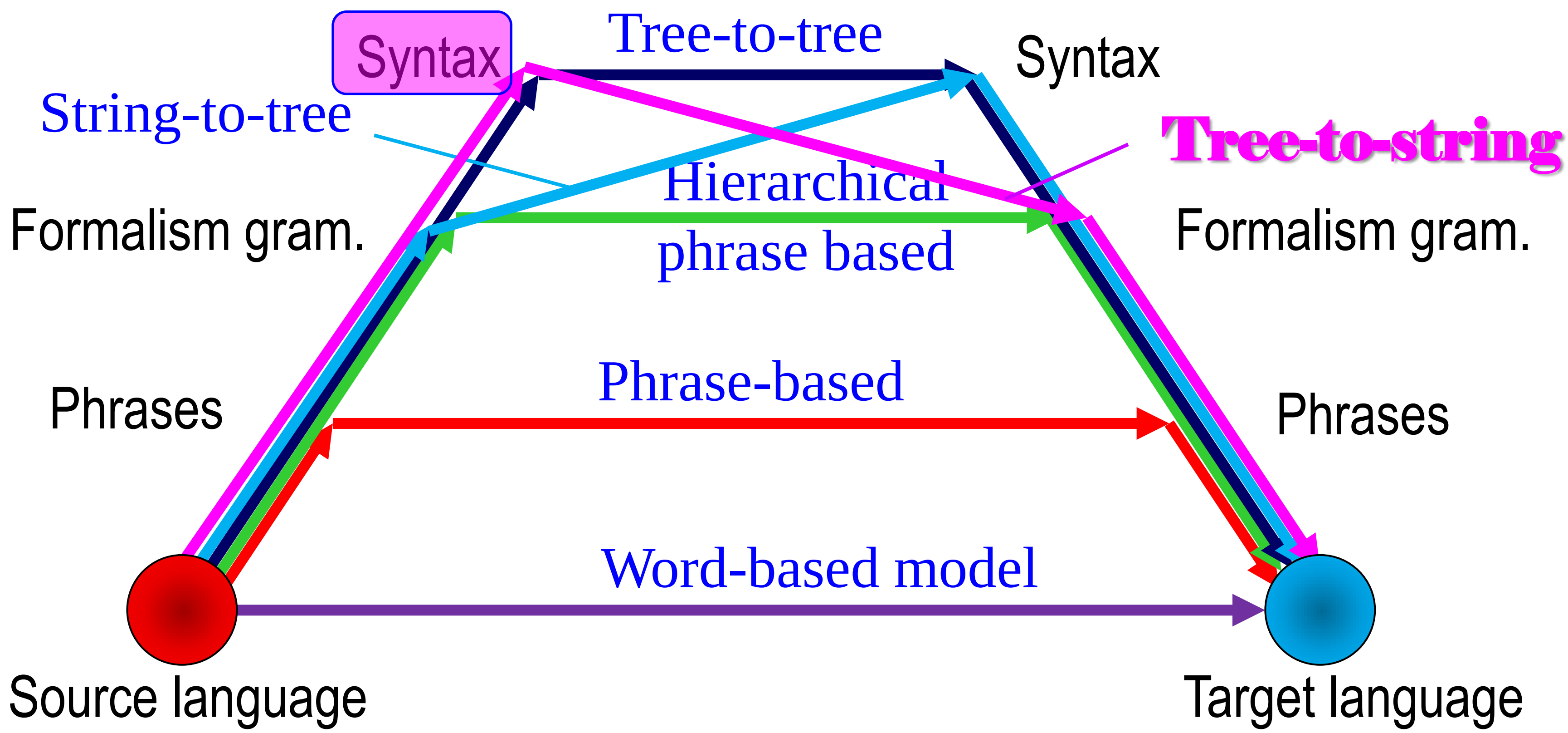
3. Motivation and Model

3.1 Motivation

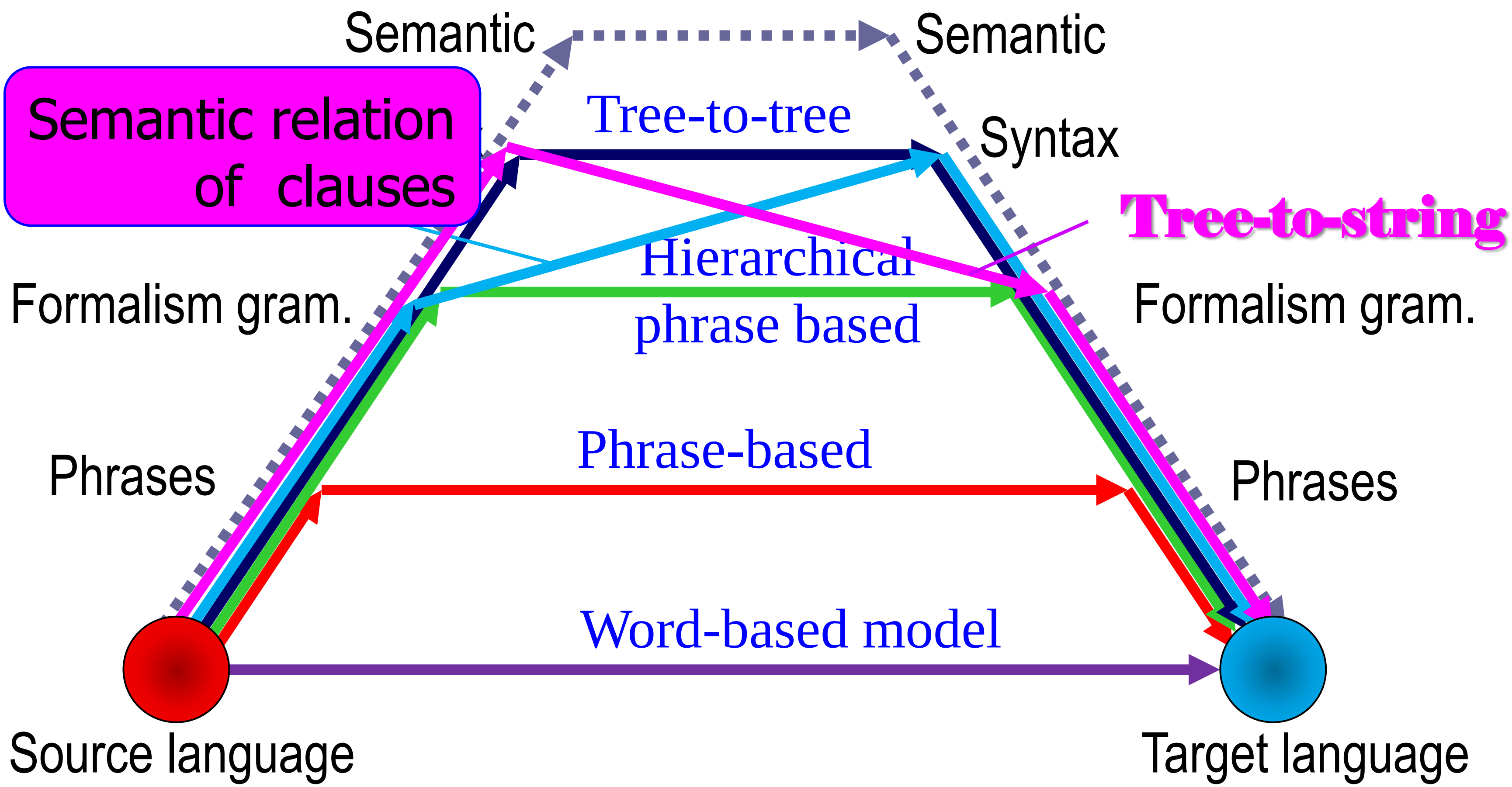
RS-tree-to-String Translation Model



3. Motivation and Model



3. Motivation and Model



3. Motivation and Model

Given a Chinese-to-English translation pair:

即使卢比对美元的名义汇率下降了，由于高通胀，其实际汇率也是上升的。

Although the rupee's nominal rate against the dollar was held down, India's real exchange rate rose because of high inflation.

3. Motivation and Model

U_1:[0, 9]

Antithesis

Jíshǐ lúbì duì meǐ yuán de míngyì huìlǜ xiàjiàng le ,

即使 卢比 对 美元 的 名义 汇率 下降 了 ，

0 1 2 3 4 5 6 7 8 9

Although the rupee’s nominal rate against the dollar was held down,

U_2:[10, 21]

U_{21}:[10, 13]

Reason

U_{22}:[14, 21]

yóuyú gāo tōngzhàng , qí shíjì huìlǜ yě shì shàngshēng de .

由于 高 通胀 ， 其 实际 汇率 也 是 上升 的 。

10 11 12 13 14 15 16 17 18 19 20 21

because of high inflation. India's real exchange rate rose

3. Motivation and Model

Why we utilize the RS-tree-to-string model?

- The semantic relations of clauses in the given text are involved and it is easy to combine with existing SMT models
- We only need a Chinese RS parser for Chinese-to-foreign language translation

3. Motivation and Model

We have to deal with the following problems:

- **Learning Chinese discourse structure**
- **Building translation model**
  - Extraction of translation rules
  - Probability estimation
  - Decoding

3. Motivation and Model

3.2 Learning Chinese discourse structure
- Chinese RS parser

- **Identify the elementary discourse units(EDUs)**
- **Recognize the relations of EDUs**

Inspired by the features used in English RST parser [Soricut and Marcu, 2003; Duverle and Prendinger, 2009; Hernault et al., 2010], we design a Bayesian model to build a joint parser for EDU identification and relation recognition simultaneously.

3. Motivation and Model

U_1:[0, 9]

Antithesis

U_2:[10, 21]

Jíshǐ lúbì duì meǐ yuán de míngyì huìlǜ xiàjiàng le ,

即使 卢比 对 美元 的 名义 汇率 下降 了 ，

0 1 2 3 4 5 6 7 8 9

Although the rupee's nominal rate against the dollar was held down,

U_{21}:[10, 13]

Reason

U_{22}:[14, 21]

yóuyú gāo tōngzhàng ,

由于 高 通胀 ，

10 11 12 13

qí shíjì huìlǜ yě shì shàngshēng de .

其 实际 汇率 也 是 上升 的 。

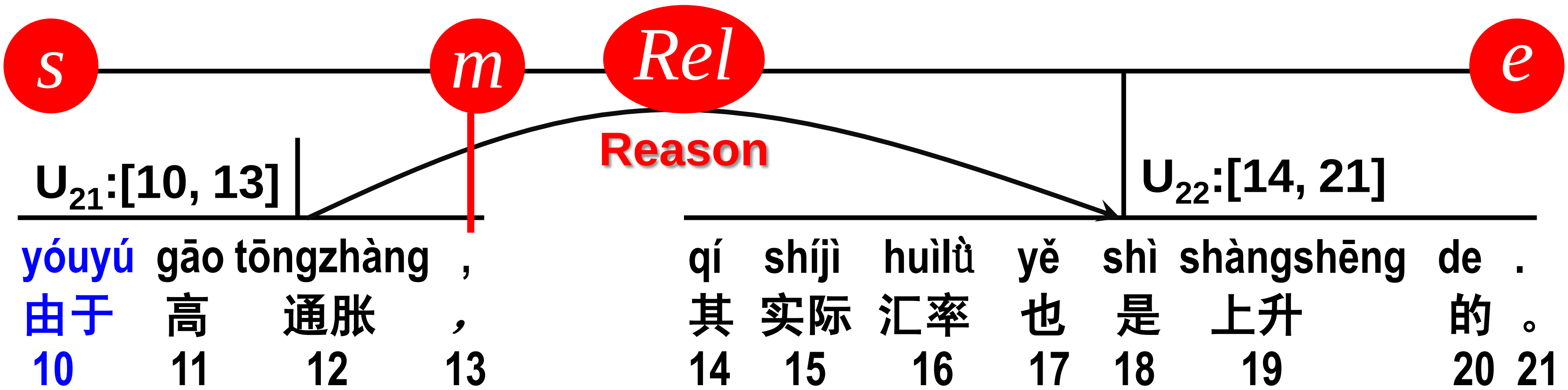
14 15 16 17 18 19 20 21

because of high inflation.

India's real exchange rate rose

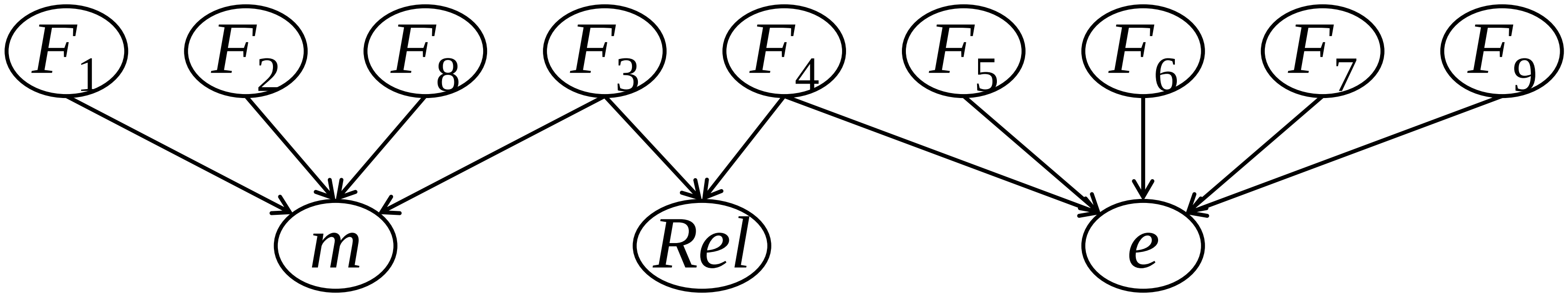
3. Motivation and Model

R-$[s, m, e]$



3. Motivation and Model

Features	Meanings
$F_1(F_6)$	left(right) child is a syntactic sub-tree?
$F_2(F_5)$	left(right) child ends with a punctuation?
$F_3(F_4)$	cue words of left (right) child.
F_7	left and right children are sibling nodes?
$F_8(F_9)$	syntactic head symbol of left(right) child.



The graph for conditional independences of 9 features

3. Motivation and Model

The conditional probability of boundary identification and relation recognition can be computed by:

$$p(m_i \mid F_1^9) = p(m_i \mid F_1^3, F_8) \quad (2)$$

$$p(e_i \mid F_1^9) = p(e_i \mid F_4^7, F_9) \quad (3)$$

$$p(Rel \mid F_1^9) = p(Rel \mid F_3^4) \quad (4)$$

F_i means the i-th feature, $F_i^{\ j}$ means the features from i to j, and *Rel* is short for the relation name.

3. Motivation and Model

```
Nodes={[]}
Parser(0,End)
Parser(s,e){                      // recursive parser function
    if s > e or e is -1: return -1;
    m = GetMaxM(s,e)              // compute m through formula (2); if no
                                     cue words found, then m=-1;
    e' = GetMaxE(s,m,e)           // compute e' through formula (3);
    if  m equals to -1 or e' equals to -1: return -1;
    Rel=GetRelation(s,m,e')  //compute relation  by formula (4)
    push [Rel,s,m,e'] into Nodes
    Parser(s,m)
    Parser(m+1,e')
    Parser(e'+1,e)
    Rel=GetRelation(s,e',e)
    push [Rel,s,e',e] into Nodes
    return e}
```

3. Motivation and Model

3.3 Building translation model

- Extraction of translation rules
- Probability estimation
- Decoding

3. Motivation and Model

U_1:[0, 9]

Antithesis

Jíshǐ lúbì duì meǐyuán de míngyì huìlǜ xiàjiàng le ,

即使 卢比 对 美元 的 名义 汇率 下降 了 ，

0 1 2 3 4 5 6 7 8 9

Although the rupee’s nominal rate against the dollar was held down,

U_2:[10, 21]

U_{21}:[10, 13]

Reason

U_{22}:[14, 21]

yóuyú gāo tōngzhàng , qí shíjì huìlǜ yě shì shàngshēng de .

由于 高 通胀 ， 其 实际 汇率 也 是 上升 的 。

10 11 12 13 14 15 16 17 18 19 20 21

because of high inflation. India's real exchange rate rose

3. Motivation and Model

- **Extraction of translation rules**
  - Phrase-based translation rules
  - RS tree-to-string rules

$$relation :: U_1(\alpha, X)/U_2(\gamma, Y) \Rightarrow U_1(tr(\alpha), tr(X)) \sim U_2(tr(\gamma), tr(Y))$$

The terminal characters α and γ represent the cue words which are optimum match for maximizing formula (4).

$$p(Rel \mid F_1^9) = p(Rel \mid F_3^4) \qquad (4)$$

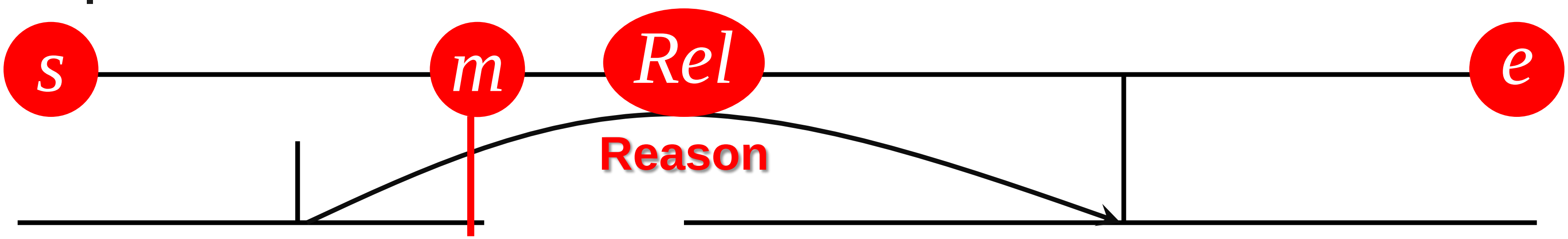
3. Motivation and Model

$$relation :: U_1(\alpha, X)/U_2(\gamma, Y) \Rightarrow U_1(tr(\alpha), tr(X)) \sim U_2(tr(\gamma), tr(Y))$$

The non-terminals *X* and *Y* represent the rest of the sequence. Function *tr*(•) means the translation of •. The operator ~ indicates the order of $tr(U_1)$ and $tr(U_2)$ is either monotone or reverse.

During the process of rule extraction, if the mean position of all words in $tr(U_1)$ precedes that in $tr(U_2)$, ~ is monotone. Otherwise, ~ is reverse.

3. Motivation and Model



yóuyú	gāo	tōngzhàng	，	qí	shíjì	huìlǜ	yě	shì	shàngshēng	de	.
由于	高	通胀	，	其	实际	汇率	也	是	上升	的	。
10	11	12	13	14	15	16	17	18	19	20	21

because of high inflation. India's real exchange rate rose

U_{21}:[10, 13] U_{22}:[14, 21]

$tr(U_{21})$ = because of high inflation.

$tr(U_{22})$ = India's real exchange rate rose

Reason::由于[*X*]/ [*Y*] => [*Y*]/ because of [X]

3. Motivation and Model

- **Probability estimation**
  - For the phrase-based translation rules, we use four common probabilities and the approach to probability estimation is the same as it is used in [Koehn et al., 2003].

3. Motivation and Model

- **Probability estimation**
  - For the RS-based translation rules, we define

$$p(r_e \mid r_c, Rel) = \frac{Count(r_e, r_c, Rel)}{Count(r_c, Rel)} \qquad (5)$$

where r_e is the target side of the rule, ignorance of the order, i.e., $U_1(tr(\alpha), tr(X))$~$U_2(tr(\gamma), tr(Y))$ with two directions; r_c is the source side, i.e., $U_1(\alpha, X)/U_2(\gamma, Y)$; and *Rel* means the relation type.

Reason::由于[*X*]/ [*Y*] => [*Y*]/ because of [*X*]

3. Motivation and Model

$$p(\tau \mid r_e, r_c, Rel) = \frac{Count(\tau, r_e, r_c, Rel)}{Count(r_e, r_c, Rel)} \qquad (6)$$

Where, $\tau \in \{\text{monotone, reverse}\}$.

3. Motivation and Model

- **Decoding**

Given the rhetorical structure c_t of a source sentence and the corresponding rule-table, translating the sentence is to select a group of best rules and then translate it into the target string e_s under the constraints of both the rules and the source sentence

3. Motivation and Model

$$\begin{aligned} &\text{argmax}_{e_s}\{p(e_s \mid c_t)\} \\ &= \text{argmax}_{e_s}\{\prod_{c_n \in c_t} p(e_{u_1}, e_{u_2}, \tau \mid c_n)\} \end{aligned} \quad (7)$$

c_t is a source RST tree combined by a set of notes c_n.
c_n consists of two edus: u_1 and u_2.

e_{u_1} and e_{u_2} are translations of u_1 and u_2 respectively.

$\tau \in \{\text{monotone, reverse}\}$

3. Motivation and Model

This is a global optimization problem. This problem can be approximately simplified to local optimization to reduce the complexity:

$$\prod_{c_n \in c_t} \operatorname{argmax}_{e_n} \{ p(e_{u_1}, e_{u_2}, \tau \mid c_n) \} \qquad (8)$$

3. Motivation and Model

- **Decoder 1**:

$$
\begin{aligned}
& p(e_{u_1}, e_{u_2}, \tau \mid c_n) \\
&= p(e_{cp}, e_X, e_Y, \tau \mid c_n) \qquad (9) \\
&= p(e_{cp} \mid c_{cp}) \times p(\tau \mid e_{cp}, c_{cp}) \times p(e_X \mid c_X) \times p(e_Y \mid c_Y) \\
&= p(r_e \mid r_c, Rel) \times p(\tau \mid r_e, r_c, Rel) \times p(e_X \mid c_X) \times p(e_Y \mid c_Y)
\end{aligned}
$$

Where, e_X and e_Y are the translation of non-terminal parts. c_{cp} and e_{cp} are cue-word pair of source and target sides.

3. Motivation and Model

After approximate simplification to local optimization, the final formulae are re-written as,

$$\operatorname{argmax}_r \{ p(r_e \mid r_c, Rel) \times p(\tau \mid r_e, r_c, Rel) \} \qquad (10)$$

$$\operatorname{argmax}_{e_X} \{ p(e_X \mid c_X) \} \qquad (11)$$

$$\operatorname{argmax}_{e_Y} \{ p(e_Y \mid c_Y) \} \qquad (12)$$

A bottom-up process can be recursively done to do the translation.

3. Motivation and Model

Reason::由于[*X*]/ [*Y*] => [*Y*]/ because of [*X*]

Reason

U_{21}:[10, 13]

U_{22}:[14, 21]

yóuyú	gāo	tōngzhàng	,	qí	shíjì	huìlǜ	yě	shì	shàngshēng	de	.
由于	高	通胀	，	其	实际	汇率	也	是	上升	的	。
10	11	12	13	14	15	16	17	18	19	20	21

high inflation.

India's real exchange rate rose

India's real exchange rate rose because of high inflation.

3. Motivation and Model

➢**Decoder 2**:

Suppose the translating process of two spans U_1 and U_2 are independent of each other, $p(e_{u1}, e_{u2}, \tau|c_n)$ can be rewritten as

$$\begin{aligned} & p(e_{u_1}, e_{u_2}, \tau \mid c_n) \\ & = p(e_{u_1}, e_{u_2}, \tau \mid c_{u_1}, c_{u_2}) \\ & = p(e_{u_1} \mid c_{u_1}) \times p(e_{u_2} \mid c_{u_2}) \times p(\tau \mid r_c, Rel) \\ & = p(e_{u_1} \mid c_{u_1}) \times p(e_{u_2} \mid c_{u_2}) \times \sum_{r_e} p(\tau \mid r_e, r_c, Rel) \times p(r_e \mid r_c, Rel) \end{aligned} \tag{13}$$

3. Motivation and Model

After approximate simplification to local optimization, the final formulae are re-written as

$$\operatorname{argmax}_{e_{u1}}\{p(e_{u_1} \mid c_{u_1})\} \qquad (14)$$

$$\operatorname{argmax}_{e_{u2}}\{p(e_{u_2} \mid c_{u_2})\} \qquad (15)$$

$$\operatorname{argmax}_{r}\{\sum_{r_e} p(\tau \mid r_e, r_c, Rel) \times p(r_e \mid r_c, Rel)\} \qquad (16)$$

Also, a bottom-up process can be recursively done to do the translation.

3. Motivation and Model

Reason

U_{21}:[10, 13]

yóuyú	gāo	tōngzhàng	,
由于	高	通胀	，
10	11	12	13

because of high inflation.

U_{22}:[14, 21]

qí	shíjì	huìlǜ	yě	shì	shàngshēng	de	.
其	实际	汇率	也	是	上升	的	。
14	15	16	17	18	19	20	21

India's real exchange rate rose

India's real exchange rate rose because of high inflation.

3. Motivation and Model

➢Comparison with decoder 1

In decoder 2, the minimal translation-unit is edu, while in decoder 1, an edu is further split into cue-word and the rest part to obtain the respective translation.

3. Motivation and Model

➢Language model

In our current system, the language model(LM) is utilized to translate edus in formula(11), (12), (14) and (15), but not to reorder the upper spans because with the combination from bottom to up, the spans become longer and harder to be judged by a traditional language model. So we only use RST rules to guide the reordering.

Outline

1. Problems
2. Related Work
3. Motivation and Model
4. Experiments
5. Conclusions and Future Work

4. Experiments

◆ Purposes

- Evaluation of Chinese RST parser
- Evaluation of RST-based C-E translation

4. Experiments

- **Evaluation of Chinese RST parser**
  - **Setup**
    - **Over 1600 annotated complicated Chinese sentences (from CTB)**
    - **1,107 sentences are utilized for training**
    - **500 sentences are utilized for testing**

4. Experiments

➢ Results of RST Parser

Tasks	Precision	Recall	F1
Segmentation	0.74	0.83	0.78
Parsing	0.71	0.78	0.75

- On average, our RS trees are 2 layers deep;
- The parsing errors are mostly resulted from the segmentation errors;
- The polysemous cue words, such as “而(but, and, thus)” may lead ambiguities for relation recognition because they can be clues for different relations.

4. Experiments

- **Evaluation of RST-based C-E translation**
  - **Setup**
    - 2.1M sentence pairs from LDC are utilized as training data
    - The word alignment with the grow-diag-final-and strategy by GIZA++
    - 5-gram language model is trained on the Xinhua portion of the English Gigaword corpus
    - NIST03 evaluation data are used as the development set

4. Experiments

- **The test data:**
  - ✓ 511 Chinese sentences (36 words in average) are from NIST04
  - ✓ 320 Chinese sentences (34 words in average) are from NIST05
  - ✓ 590 Chinese sentences (38 words in average) are from CWMT'08
- **The baseline:**
  - ✓ Moses, a phrase-based MT system
  - ✓ Xiong's work

4. Experiments

➢ Results of Translation

Testing Set	Baseline	Xiong's	Decoder1	Decoder2
NIST04	29.39	31.52	**31.34**	**31.69**
NIST05	29.86	29.80	30.28	**30.63**
CWMT08	24.31	25.24	**25.74**	**25.74**
ALL	27.85	28.49	28.66	**29.01**

Outline

1. Problems
2. Related Work
3. Motivation and Model
4. Experiments
5. **Conclusions and Future Work**

5.Conclusions and Future Work

◆ Conclusions

- RST-based translation framework is modeling semantic structures in translation model. It can maintain the semantically functional integrity and hierarchical relations of *edu*s during translating in some extent;
- This translation framework works more similarly to what human does.

5.Conclusions and Future Work

◆ Future work

- ➢ How to improve the performance of Chinese RST parser?
- ➢ How to recognize the implicit discourse relations?
- ➢ How to combine other existing translation models such as syntactic model and hierarchical model into our proposed framework?
- ➢ How to evaluate the improvement of discourse-based translation?

5.Conclusions and Future Work

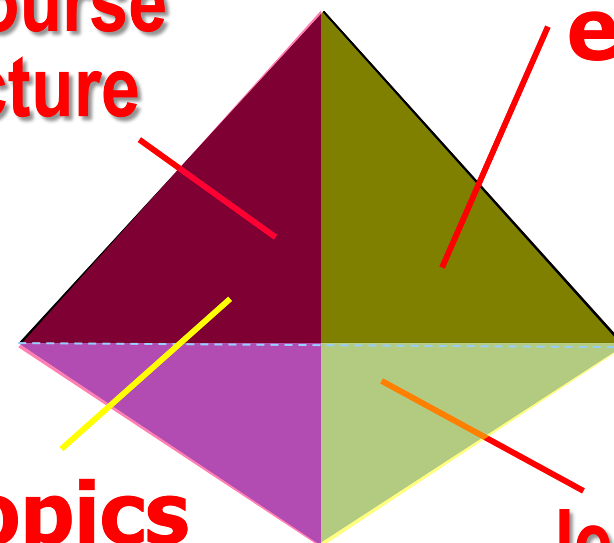
paraphrase ← repeat

topic start → change

sentence group → discourse structure

relation recognition ← edu recognition

lexicon → lexical link → Anaphora disambiguition



5.Conclusions and Future Work

- **We have much work to do for discourse parsing**
  - Building lexical links
  - Anaphora disambiguition
  - edu recognition
  - Relation recognition
  - Topic identification and tracing
  - Sentence grouping
  - Establishment of Chinese discourse analysis theory

5.Conclusions and Future Work

- How to annotate large scale discourse corpora?
- How to utilize discourse information to do MT?
- How the discourse information help Q&A?

  ……

References

[Duverle and Prendinger, 2009]David A Duverle and Helmut Prendinger. 2009. A novel discourse parser based on support vector machine classification. *Proc. ACL-IJCNLP*

[Gimenez et al., 2010]Jesus Gimenez et al. 2010. Document-Level Automatic MT Evaluation Based on Discourse Representations. *Proc. the 5th Workshop on SMT and MetricsMATR*

[Gong et al., 2011]Zhengxian Gong, Min Zhang and Guodong Zhou. 2011. Cache-based Document-level Statistical Machine Translation. *Proc. EMNLP*

[Hernault et al., 2010]Hugo Hernault, Danushka Bollegala, and Mitsuru Ishizuka. 2010. A sequential model for discourse segmentation. *Computational Linguistics and Intelligent Text Processing*, pages 315–326.

[Koehn et al., 2003]Philipp Koehn, Franz Josef Och, and Daniel Marcu. 2003. Statistical phrase-based translation. *Proc. NAACL*

[Lin et al., 2011]Ziheng Lin, Hwee Tou Ng, and Min-Yen Kan. 2011. Automatically Evaluating Text Coherence Using Discourse Relations. *Proc. ACL*

[Lin et al., 2012]Ziheng Lin et al. 2012. Combining Coherence Models and Machine Translation Evaluation Metrics for Summarization Evaluation. *Proc. ACL*

References

[Mann and Thompson, 1987]W. Mann and Sandra A. Thompson. 1987. Rhetorical Structure Theory: A Theory of Text Organization. ISI/RS-87-190

[Marcu et al., 2000]Daniel Marcu, Lynn Carlson, and Maki Watanabe. 2000. The Automatic Translation of Discourse Structures. *Proc. NAACL*

[Shi and Chen, 2006]Xiaodong Shi and Yidong Chen. 2006. Previews of Approach to Discourse-Based Machine Translation. *Proc. Frontiers of Chinese Information Processing.* (in Chinese)

[Soricut and Marcu, 2003]Radu Soricut and Daniel Marcu. 2003. Sentence level discourse parsing using syntactic and lexical information. *Proc. NAACL*

[Wong and Kit, 2012]B. T. M. Wong and C. Kit. 2012. Extending Machine Translation Evaluation Metrics with Lexical Cohesion to Document Level. *Proc. EMNLP-Computational Natural Language Learning*

[Xiong et al., 2013]Deyi Xiong, Guosheng Ben, Min Zhang, Yajuan Lv, and Qun Liu. 2013. Modeling lexical cohesion for document-level machine translation. *Proc. IJCAI*

Thanks

谢谢!