



# Paraphrases and Applications

Shiqi Zhao  
Baidu, Inc.

Haifeng Wang  
Baidu, Inc.

## Outline

- **Part I**
  - **Introduction**
  - Paraphrase Identification
  - Paraphrase Extraction
- **Part II**
  - Paraphrase Generation
  - Applications of Paraphrases
  - Evaluation of Paraphrases
  - Conclusions and Future work



## Definition

- Paraphrase
  - Noun
    - Alternative expressions of the same meaning
  - Verb
    - Generate paraphrases for the input expression
- “same meaning”?
  - Quite subjective
  - Different degrees of strictness
  - Depend on applications



## Paraphrase (noun): Alternative expressions of the same meaning



Korean **Kim Yuna** won gold with a world-record score in women's figure skating at the Vancouver Olympics Thursday.

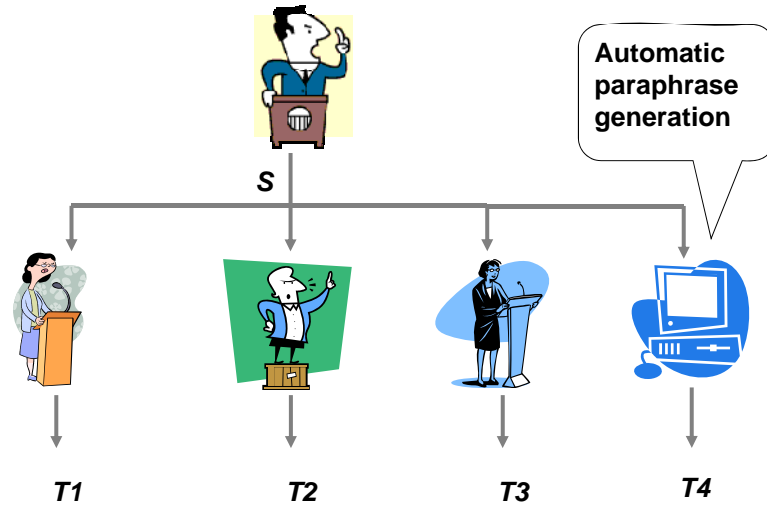
Korean figure skater **Kim Yuna** has won the gold medal of women's figure skating at the Winter Olympics in Vancouver

**Kim Yu-Na** (19) is a South Korean ice skater who took the gold medal at the Vancouver Olympics.

**Yuna Kim** of South Korea won the women's figure skating gold medal at the Vancouver Olympics in record fashion.

**Kim Yuna**, a South Korean figure skater has won the gold medal at the on-going Winter Olympics 2010.

**Paraphrase (verb): Generate paraphrases for an input S.**



## Classification of Paraphrases

- According to granularity
  - Surface paraphrases
    - Lexical level
    - Phrase level
    - Sentence level
    - Discourse level
  - Structural paraphrases
    - Pattern level
    - Collocation level

# Examples

- Lexical paraphrases (generally synonyms)
  - *solve* and *resolve*
- Paraphrase phrases
  - *look after* and *take care of*
- Paraphrase sentences
  - *The table was set up in the carriage shed.*
  - *The table was laid under the cart-shed.*
- Paraphrase patterns
  - *[X] considers [Y]*
  - *[X] takes [Y] into consideration*
- Paraphrase collocations
  - *(turn on, OBJ, light)*
  - *(switch on, OBJ, light)*

## Classification of Paraphrases

- According to paraphrase style
  - Trivial change
  - Phrase replacement
  - Phrase reordering
  - Sentence split & merge
  - Complex paraphrases



# Examples

- Trivial change
  - all the members of and all members of
- Phrase replacement
  - He said there will be major cuts in the salaries of high-level civil servants.
  - He said there will be major cuts in the salaries of senior officials.
- Phrase reordering
  - Last night, I saw Tom in the shopping mall.
  - I saw Tom in the shopping mall last night.
- Sentence split & merge
  - He bought a computer, which is very expensive.
  - (1) He bought a computer. (2) The computer is very expensive.
- Complex paraphrases
  - He said there will be major cuts in the salaries of high-level civil servants.
  - He claimed to implement huge salary cut to senior civil servants.

## Applications of Paraphrases

- Machine Translation (MT)
  - Simplify input sentences
  - Alleviate data sparseness
  - Parameter tuning
  - Automatic evaluation
- Question Answering (QA)
  - Question reformulation
- Information Extraction (IE)
  - IE pattern expansion
- Information Retrieval (IR)
  - Query reformulation
- Summarization
  - Sentence clustering
  - Automatic evaluation
- Natural Language Generation (NLG)
  - Sentence rewriting
- Others
  - Changing writing style
  - Text simplification
  - Identifying plagiarism
  - Text steganography
  - .....



## Research on Paraphrasing

- Paraphrase identification
  - Identify (sentential) paraphrases
- Paraphrase extraction
  - Extract paraphrase instances (different granularities)
- Paraphrase generation
  - Generate (sentential) paraphrases
- Paraphrase applications
  - Apply paraphrases in other areas



## Textual Entailment – A Similar Direction

- Textual entailment:
  - A directional relation between two text fragments
    - $T$ : the entailing text
    - $H$ : the entailed hypothesis
  - $T$  entails  $H$  if, typically, a human reading  $T$  would infer that  $H$  is most likely true.
  - Compare entailment with paraphrase
    - Paraphrase is bidirectional entailment



## Text Entailment – A Similar Direction

- Recognizing Textual Entailment Track (RTE)
  - RTE-1 (2004) to RTE-5 (2009)
  - RTE-6 (2010) is in progress
- Example:
  - **T**: A shootout at the Guadalajara airport in May, 1993, killed Cardinal Juan Jesus Posadas Ocampo.
  - **H**: Juan Jesus Posadas Ocampo died in 1993.



## Outline

- Part I
  - Introduction
  - **Paraphrase Identification**
  - Paraphrase Extraction
- Part II
  - Paraphrase Generation
  - Applications of Paraphrases
  - Evaluation of Paraphrases
  - Conclusions and Future work



## Paraphrase Identification

- Specially refers to sentential paraphrase identification
  - Given any pair of sentences, automatically identifies whether these two sentences are paraphrases
- Paraphrase identification is not trivial

Susan **often** goes to see movies with her boyfriend.  
Susan **never** goes to see movies with her boyfriend.

He **said there will be major cuts in the salaries of high-level** civil servants.  
He **claimed to implement huge salary cuts to senior** civil servants.

## Overview

- Classification based methods
  - Reviewed as a binary classification problem, i.e., input  $s_1$  and  $s_2$  to a classifier and output 0/1
  - Compute the similarities between  $s_1$  and  $s_2$  at different levels, which are then used as classification features
- Alignment based methods
  - Align  $s_1$  and  $s_2$  first, and score the sentence pair based on the alignment results
    - Alignment based on ITG
    - Alignment based on quasi-synchronous dependency grammars



## Classification based Methods

- Brockett and Dolan, 2005

- Features:

- String similarity features
  - Sentence length, word overlap, edit distance, ...
- Morphological variants
  - Word pairs with the same stem
- WordNet lexical mappings
  - Synonym pairs / word-hypernym pairs from WordNet
- Word association pairs
  - Automatically learned synonym pairs

orbit | orbital

operation | procedure

vendors | suppliers

- Classifier

- SVM classifier



## Classification based Methods (cont')

- Finch et al., 2005

- Using MT evaluation techniques to compute sentence similarities, which are then used as classification features

- WER, PER, BLEU, NIST
- Feature vector  $vec(\mathbf{s}_1, \mathbf{s}_2)$ 
  - $vec1(\mathbf{s}_1, \mathbf{s}_2)$ :  $\mathbf{s}_1$  as reference,  $\mathbf{s}_2$  as MT system output;
  - $vec2(\mathbf{s}_1, \mathbf{s}_2)$ :  $\mathbf{s}_2$  as reference,  $\mathbf{s}_1$  as MT system output;
  - $vec(\mathbf{s}_1, \mathbf{s}_2)$ : average of  $vec1(\mathbf{s}_1, \mathbf{s}_2)$  and  $vec2(\mathbf{s}_1, \mathbf{s}_2)$ :

- Classifier

- SVM classifier



## Classification based Methods (cont')

- Malakasiotis, 2009
  - Combining multiple classification features
    - String similarity (various levels)
      - Tokens, stems, POS tags, nouns only, verbs only, ...
    - Different measures
      - Edit distance, Jaro-Winkler distance, Manhattan distance...
    - Synonym similarity
      - Treat synonyms in two sentences as identical words
    - Syntax similarity
      - Dependency parsing of two sentences and compute the overlap of dependencies
  - Classifier
    - Maximum Entropy classifier



## Alignment based Methods

- Wu, 2005
  - Conduct alignment based on **I**nversion **T**ransduction **G**rammars (ITG)
    - Sensitive to the differences in sentence structures
    - Without using any thesaurus to deal with lexical variation
  - Performance is comparable to the classification based methods
  - Also performs well in recognizing textual entailment

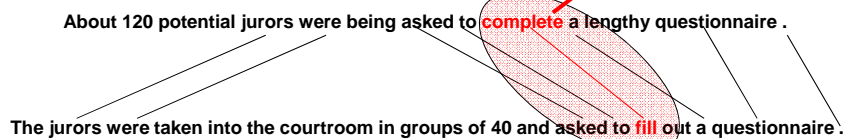


## Alignment based Methods (cont')

- Das and Smith, 2009
  - Conduct alignment based on Quasi-Synchronous Dependency Grammar (QG)
    - Alignment between two dependency trees
    - Assumption: the dependency trees of two paraphrase sentences should be aligned closely
  - Why does it work?

About 120 potential jurors were being asked to **complete** a lengthy questionnaire .

The jurors were taken into the courtroom in groups of 40 and asked to **fill** out a questionnaire .



- Performs competitively with classification based methods



## A Summary

- Classification based method is still the mainstream method, since:
  - Binary classification problem is well defined;
  - Classification algorithms and tools are readily available;
  - It can combine various features in a simple way;
  - It achieves state-of-the-art performance.



## References

- Brockett and Dolan. 2005. Support Vector Machines for Paraphrase Identification and Corpus Construction.
- Finch et al. 2005. Using Machine Translation Evaluation Techniques to Determine Sentence-level Semantic Equivalence.
- Wu. 2005. Recognizing Paraphrases and Textual Entailment using Inversion Transduction Grammars.
- Malakasiotis. 2009. Paraphrase Recognition Using Machine Learning to Combine Similarity Measures.
- Das and Smith. 2009. Paraphrase Identification as Probabilistic Quasi-Synchronous Recognition.

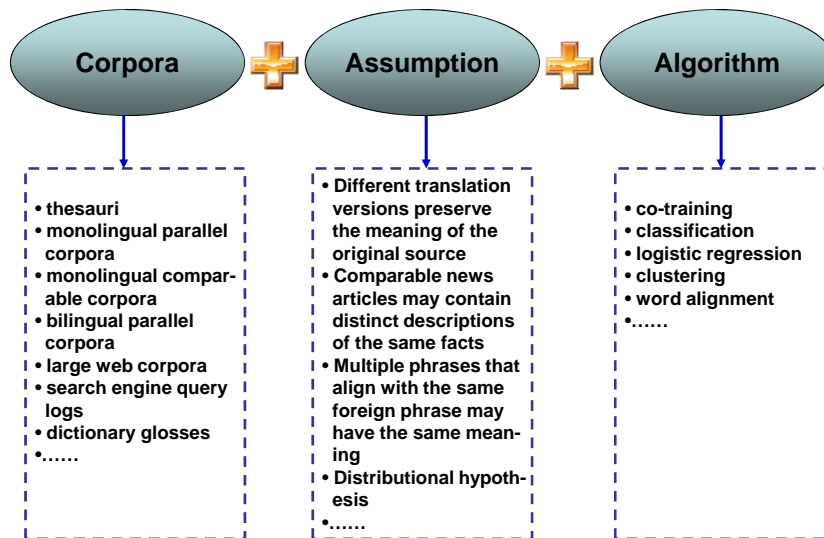


## Outline

- **Part I**
  - Introduction
  - Paraphrase Identification
  - **Paraphrase Extraction**
- **Part II**
  - Paraphrase Generation
  - Applications of Paraphrases
  - Evaluation of Paraphrases
  - Conclusions and Future work



## Three Elements for Paraphrase Extraction



## Outline

### • Part I

- Introduction
- Paraphrase Identification
- **Paraphrase Extraction**
  - **From Thesauri**
  - From Monolingual Parallel Corpora
  - From Monolingual Comparable Corpora
  - From Bilingual Parallel Corpora
  - From Large Web Corpora
  - From Other Resources

## Method Overview

- Extract words with specific semantic relations as paraphrases
  - Most common: synonyms
  - Other relations: hypernyms, hyponyms...
- Widely used thesauri
  - In English
    - WordNet
  - In other languages
    - E.g., HowNet, Tongyici Cilin in Chinese



## Pros and Cons

- Pros
  - Existing resources
  - High quality
    - Thesauri are hand crafted
- Cons
  - Language limitation
    - Thesauri are not available in many languages
  - Difficult to update



## Outline

- **Part I**

- Introduction
- Paraphrase Identification
- **Paraphrase Extraction**
  - From Thesauri
  - **From Monolingual Parallel Corpora**
  - From Monolingual Comparable Corpora
  - From Bilingual Parallel Corpora
  - From Large Web Corpora
  - From Other Resources



## Method Overview

- Corpus
  - Multiple translations of the same foreign literary work
- Assumption
  - Different translation versions preserve the meaning of the original source, but may use different expressions

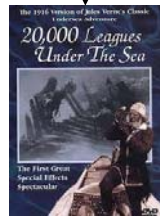
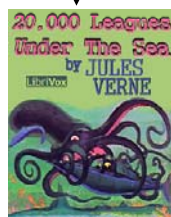


## Example



Vingt mille lieues sous les mers  
(in French)

20000 Leagues Under the Sea  
(different English translation versions)



.....

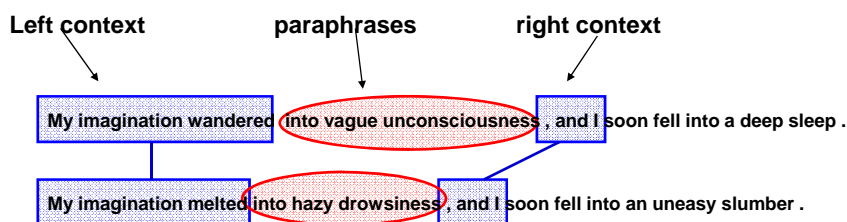
## Sentence Alignment and Preprocessing

- Barzilay and McKeown, 2001
  - Collected 11 English translations for 5 foreign novels
    - E.g., *Madame Bovary*, *Fairy Tale*, *Twenty Thousand Leagues under the sea...*
  - Sentence alignment
    - A dynamic programming algorithm
    - Produced 44,562 pairs of parallel sentences
    - Precision is 94.5%
  - Other preprocessing
    - POS tagging and chunking
    - Phrases are the atomic units in paraphrase extraction



## Paraphrase Phrase Extraction

- Barzilay and McKeown, 2001 (cont')
  - Extracting paraphrase phrases
    - Assumption: phrases in aligned sentences which appear in similar contexts are paraphrases
    - Method: co-training
      - Iteratively learn contexts and paraphrases



## Pros and Cons

- Pros
  - Easy to align monolingual parallel sentences
- Cons
  - Domain limitation
    - Limited in literary works
  - Scale limitation
    - The size of the corpus is relatively small
  - Context dependence
    - E.g., “*John said*” and “*he said*”



## Outline

- **Part I**

- Introduction
- Paraphrase Identification
- **Paraphrase Extraction**
  - From Thesauri
  - From Monolingual Parallel Corpora
  - **From Monolingual Comparable Corpora**
  - From Bilingual Parallel Corpora
  - From Large Web Corpora
  - From Other Resources

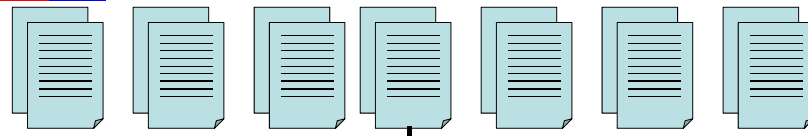


## Method Overview

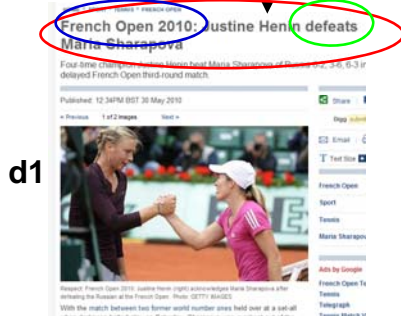
- **Corpus**
  - News articles that report the same event within a brief period of time
    - Produced by different news agencies
- **Assumption**
  - Comparable news articles may contain distinct descriptions of the same facts



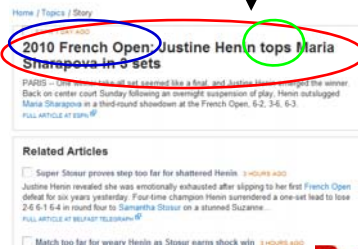
## Example



Comparable documents



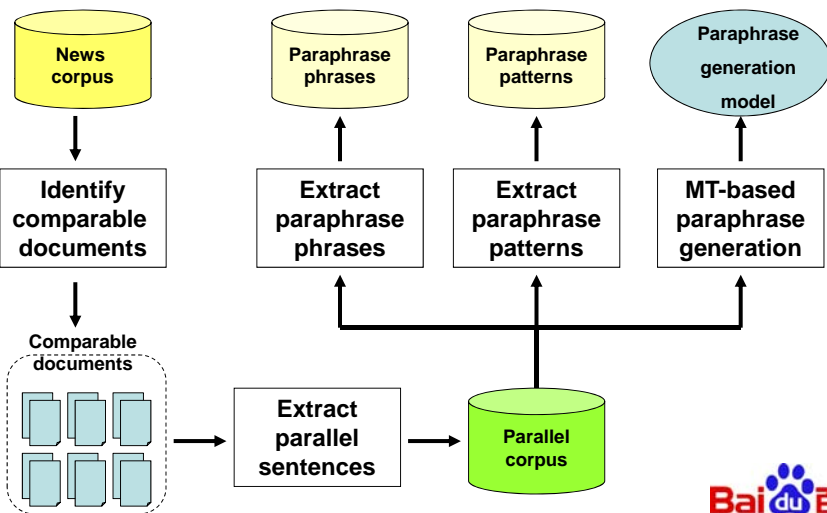
d1



d2



## Procedure



## Identify Comparable Documents

- Input
  - News articles from different news agencies
    - E.g., CNN, New York Times, Washington Post...
- Processing
  - **Method-1**: Retrieve documents on a given topic or event
    - Needs predefined topics or events
  - **Method-2**: Cluster documents
    - Content similarity; time interval
- Output
  - Corpus of comparable documents



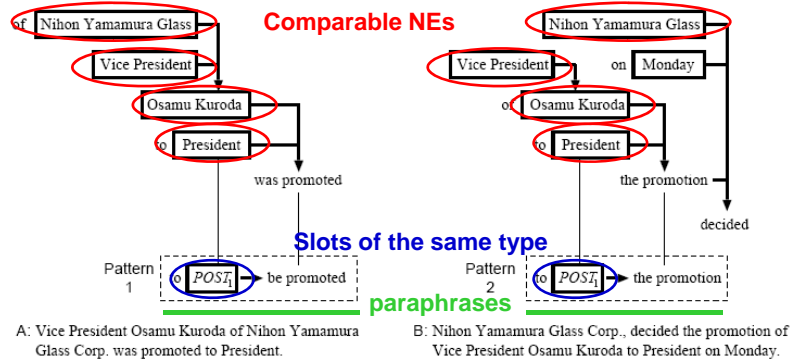
## Extract Parallel (Paraphrase) Sentences

- Input
  - Corpus of comparable documents
- Processing
  - Sentence clustering
    - **Method-1**: based on an assumption: first sentences of a news article usually summarize its content
    - **Method-2**: based on computing the content similarity
- Output
  - Corpus of parallel (paraphrase) sentences



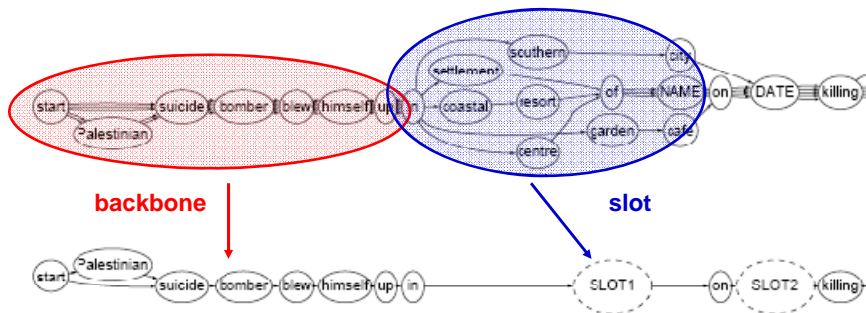
## Extract Paraphrase Patterns

- Using NEs as anchors
  - Shinyama et al., 2002
  - Basic idea: paraphrase sentences should contain comparable NEs



## Extract Paraphrase Patterns

- Multiple-sequence alignment
  - Barzilay and Lee, 2003



## Pros and Cons

- Pros
  - Language-independent
    - Comparable news can be found in many languages
- Cons
  - Domain-dependent
    - Paraphrases are extracted from specific domains or topics
  - Sentence clustering
    - Either too strict or too loose



## Outline

- **Part I**
  - Introduction
  - Paraphrase Identification
  - **Paraphrase Extraction**
    - From Thesauri
    - From Monolingual Parallel Corpora
    - From Monolingual Comparable Corpora
    - **From Bilingual Parallel Corpora**
    - From Large Web Corpora
    - From Other Resources

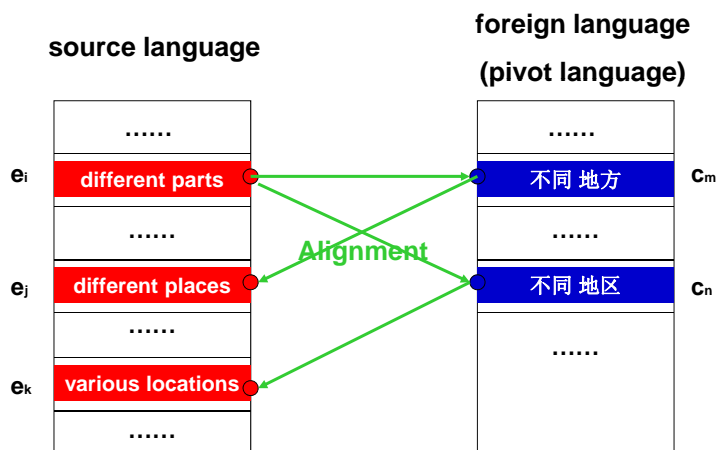


## Method Overview

- Corpus
  - A parallel corpus of the source language and a foreign language
- Assumption
  - Multiple phrases that align with the same foreign phrase may have the same meaning
- The method is also termed as “*pivot approach*”



## Example

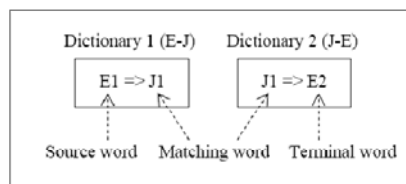


## A Simple Version

- Takao et al., 2002

- Basic idea:

- Generating lexical paraphrases using 2-way dictionaries
    - English word  $e_1$  can be translated to a Japanese word  $j$  with an E-J dic.  $D_1$ , and then  $j$  can be translated back to an English word  $e_2$  with a J-E dictionary  $D_2$ .  $e_1$  and  $e_2$  are extracted as paraphrases



## Extracting Paraphrase Phrases

- Bannard and Callison-Burch, 2005

- Word alignment and phrase extraction

- Basic assumption:

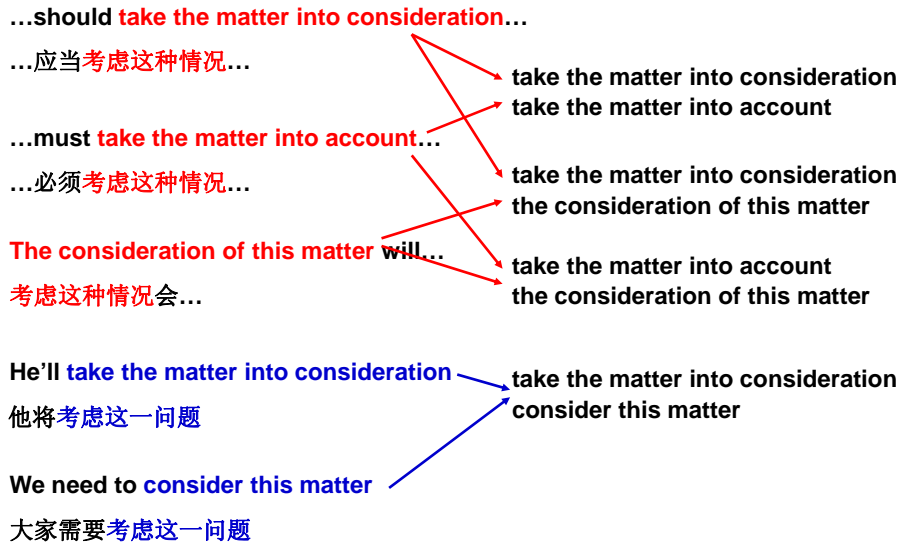
- If two English phrases  $e_1$  and  $e_2$  can be aligned with the same foreign phrase  $f$ ,  $e_1$  and  $e_2$  are likely to be paraphrases.

- Paraphrase probability:

$$\hat{e}_2 = \arg \max_{e_2 \neq e_1} p(e_2 | e_1)$$
$$= \arg \max_{e_2 \neq e_1} \underbrace{\sum_f p(f | e_1)}_{\text{Translation probability}} \underbrace{p(e_2 | f)}_{\text{Pivot in a foreign language}}$$



## Bannard & Callison-Burch (2005) 's results:



## Add Syntactic Constraints

- Callison-Burch, 2008

- Basic idea:

- Two paraphrase phrases should have the same syntactic type.

- Paraphrase probability:

$$\hat{e}_2 = \arg \max_{e_2: e_2 \neq e_1 \wedge s(e_2) = s(e_1)} p(e_2 | e_1, s(e_1))$$

→ given the syntactic type

$$= \arg \max_{e_2: e_2 \neq e_1 \wedge s(e_2) = s(e_1)} \sum_f p(f | e_1, s(e_1)) p(e_2 | f, s(e_1))$$

- Syntactic constraints are also used when substituting paraphrases in sentences



## Callison-Burch (2008) 's results:

...should **take the matter into consideration**...

...应当**考虑**这种情况...

...**must take the matter into account**...

...**必须**考虑这种情况...

**The consideration of this matter will**...

**考虑**这种情况会...

He'll **take the matter into consideration**

他将**考虑**这一问题

We need to **consider this matter**

大家需要**考虑**这一问题

take the matter into consideration  
take the matter into account

take the ~~matter~~ into consideration  
the ~~consideration~~ of this matter

take the ~~matter~~ into account  
the ~~consideration~~ of this matter

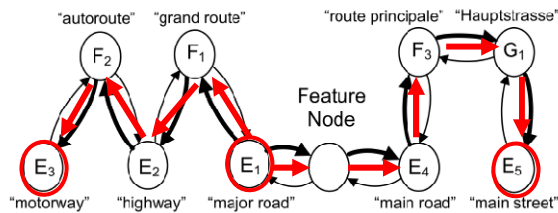
take the matter into consideration  
consider this matter

## Learning Paraphrases from Graphs

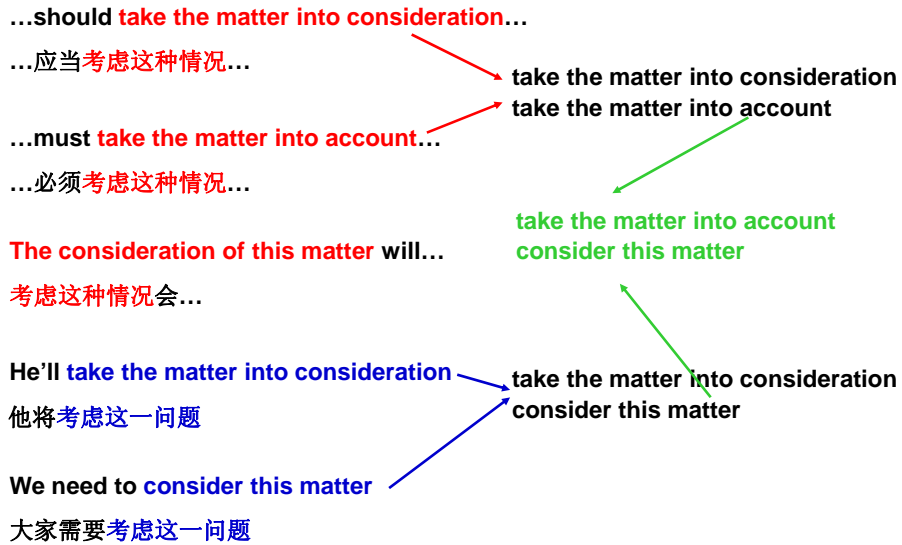
- Kok and Brockett, 2010

— Basic idea:

- Convert aligned phrases into a graph, extract paraphrases based on random walks and hitting times



## Kok and Brockett (2010) 's results:



## Extracting Paraphrase Patterns

- Zhao et al., 2008

- Basic idea:

- Generate paraphrase patterns that include part-of-speech slots.

- Paraphrase probability:

$$score(e_2 | e_1) = \sum_c \exp\left[\sum_{i=1}^N \lambda_i h_i(e_1, e_2, c)\right]$$

$$h_1(e_1, e_2, c) = score_{MLE}(c | e_1)$$

$$h_2(e_1, e_2, c) = score_{MLE}(e_2 | c)$$

$$h_3(e_1, e_2, c) = score_{LW}(c | e_1)$$

$$h_4(e_1, e_2, c) = score_{LW}(e_2 | c)$$

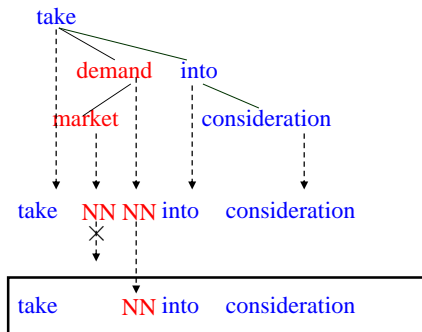
} Based on maximum likelihood estimation

} Based on lexical weighting

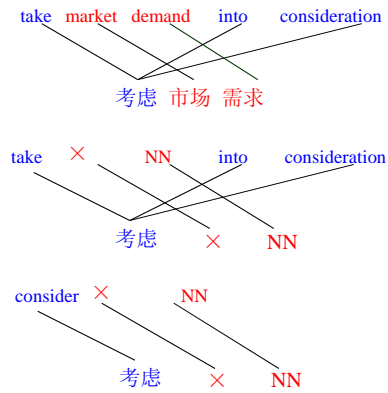


## Example

### Inducing English patterns



### Inducing Chinese patterns



### Extract paraphrase patterns

take NN into consideration & consider NN

## Zhao et al (2008) 's results:

...should **take the matter into consideration**...

...应当**考虑**这种情况...

...**must take the matter into account**...

...**必须考虑**这种情况...

**The consideration of this matter will**...

**考虑**这种情况会...

He'll **take the matter into consideration**

他将**考虑**这一问题

We need to **consider this matter**

大家需要**考虑**这一问题

take [NN] into consideration  
 take [NN] into account

take [NN] into consideration  
 the consideration of [NN]

take [NN] into account  
 the consideration of [NN]

take [NN] into consideration  
 consider [NN]

## Pros and Cons

- Pros
  - The method proves effective, hence it's widely used
    - High precision
    - Large scale
- Cons
  - Language limitation
    - Cannot work where the large-scale bilingual parallel corpora are not available



## Outline

- **Part I**
  - Introduction
  - Paraphrase Identification
  - **Paraphrase Extraction**
    - From Thesauri
    - From Monolingual Parallel Corpora
    - From Monolingual Comparable Corpora
    - From Bilingual Parallel Corpora
    - **From Large Web Corpora**
    - From Other Resources

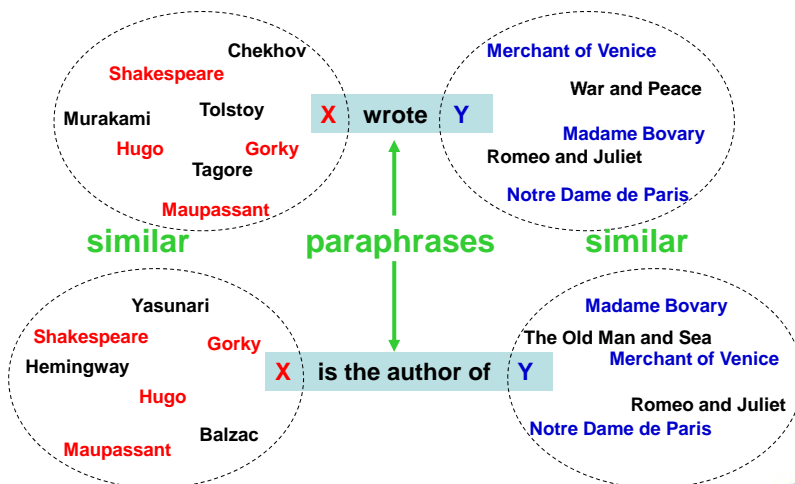


## Method Overview

- Corpus
  - Large corpus of web documents
  - Or directly based on web mining
- Assumption
  - Distributional hypothesis
    - If two words / phrases / patterns often occur in similar contexts, their meanings tend to be similar



## Example



## Extracting Lexical Paraphrases (Word Clustering)

- Lin, 1998

- Basic idea

- Measure words' similarity based on the distributional pattern of words

- Corpus

- A (dependency) parsed corpus

- Word similarity

$$sim(w_1, w_2) = \frac{\sum_{(r,w) \in T_r(w_1) \cap T_r(w_2)} (I(w_1, r, w) + I(w_2, r, w))}{\sum_{(r,w) \in T_r(w_1)} I(w_1, r, w) + \sum_{(r,w) \in T_r(w_2)} I(w_2, r, w)}$$

Mutual information



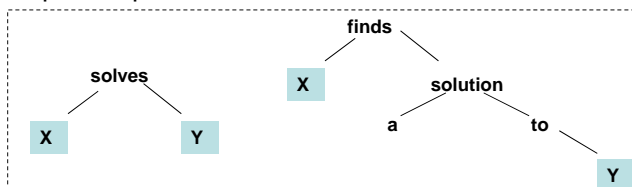
## Extracting Syntactic Paraphrase Patterns

- Lin and Pantel, 2001

- Basic idea: extended distributional hypothesis

- Corpus: a large corpus of parsed monolingual sentences

- pattern pairs



- Pattern similarity

$$sim(p_1, p_2) = \sqrt{sim(SlotX_1, SlotX_2) \times sim(SlotY_1, SlotY_2)}$$

Similarity of the slot fillers



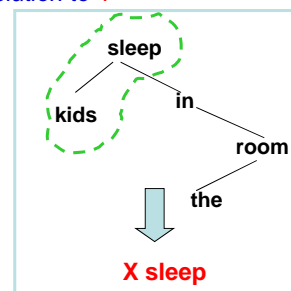
## Extracting Surface Paraphrases

- Bhagat and Ravichandran, 2008
  - Basic idea is the same as the above work
  - Corpus:
    - a large corpus of monolingual sentences without parsing
      - 150GB, 25 billion words
  - Surface paraphrases
    - Pairs of n-grams
      - E.g., “*X acquired Y*” and “*X completed the acquisition of Y*”
  - Techniques
    - Apply locality sensitive hashing (LSH) to speed up the computation



## Learning Unary Paraphrase Patterns

- Szpektor and Dagan, 2008
  - Binary paraphrase patterns (most of the previous work)
    - Each pattern has two slots at both ends
      - E.g., “*X solves Y*” and “*X found a solution to Y*”
  - Unary paraphrase patterns
    - Each pattern has a single slot
      - E.g., “*X take a nap*” and “*X sleep*”
  - Method
    - The same with the above works
      - Based on distributional hypothesis



## Extracting Paraphrases based on Web Mining

- Ravichandran and Hovy, 2002
  - Basic idea
    - Learn paraphrase patterns with search engines
  - Corpus
    - The whole internet
  - Method
    - Extract paraphrase patterns for each type, e.g., “*BIRTHDAY*”
    - Provide hand-crafted seeds, e.g., “*Mozart, 1756*”
    - Retrieve sentences containing the seeds from the web with a search engine
    - Extract patterns, e.g.,
      - *born in <ANSWER> , <NAME>*
      - *<NAME> was born on <ANSWER> ,*
      - .....



## Pros and Cons

- Pros
  - Language independent
- Cons
  - For methods based on large web corpora
    - Computation complexity is high
      - Needs to process an extremely large corpus
      - Needs to compute pairwise similarity for all candidates
  - For methods based on web mining
    - Extract paraphrase patterns type by type
    - Needs to prepare seeds beforehand



## Outline

- **Part I**

- Introduction
- Paraphrase Identification
- **Paraphrase Extraction**
  - From Thesauri
  - From Monolingual Parallel Corpora
  - From Monolingual Comparable Corpora
  - From Bilingual Parallel Corpora
  - From Large Web Corpora
  - **From Other Resources**



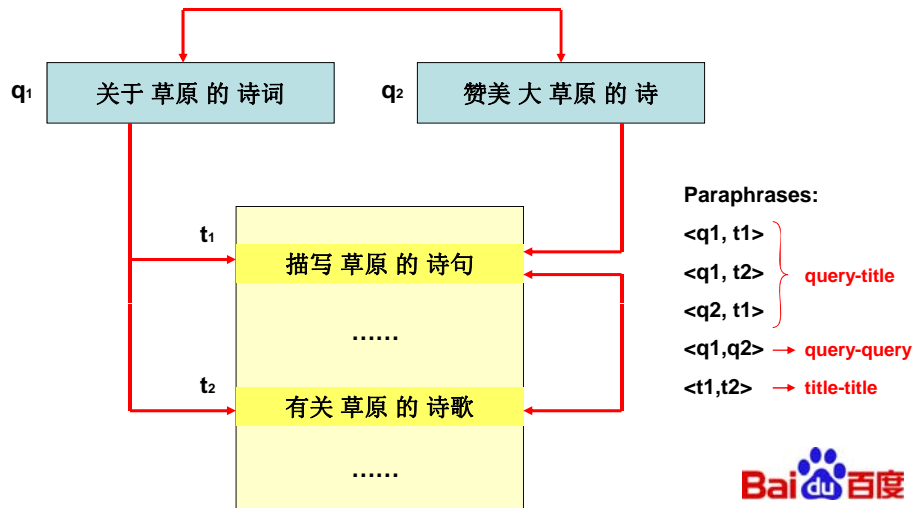
## Paraphrasing with Search Engine Query Logs

- Zhao et al., 2010

- Corpus
  - Query logs (queries and titles) of a search engine
- Assumption
  - If a query  $q$  hits a title  $t$ , then  $q$  and  $t$  are likely to be paraphrases
  - If queries  $q_1$  and  $q_2$  hit the same title  $t$ , then  $q_1$  and  $q_2$  are likely to be paraphrases
  - If a query  $q$  hits titles  $t_1$  and  $t_2$ , then  $t_1$  and  $t_2$  are likely to be paraphrases



## Example



## Method

- Step-1: extracting  $\langle q, t \rangle$  paraphrases
  - Extracting candidate  $\langle q, t \rangle$  pairs from query logs
  - Paraphrase validation based on binary classification
    - Combining multiple features
- Step-2: extracting  $\langle q, q \rangle$  paraphrases
  - Extracting candidate  $\langle q, q \rangle$  from  $\langle q, t \rangle$  paraphrases
  - Paraphrase validation based on binary classification
- Step-3: extracting  $\langle t, t \rangle$  paraphrases
  - Extracting candidate  $\langle t, t \rangle$  from  $\langle q, t \rangle$  paraphrases
  - Paraphrase validation based on binary classification

## Pros and Cons

- Pros
  - No scale limitation
    - Query logs keep growing
    - A large volume of paraphrases can be extracted
  - Query logs reflect web users' real needs
- Cons
  - Query logs data are only available in IR companies
  - User queries are noisy
    - Spelling mistakes, grammatical errors...



## Extracting Paraphrases from Dictionary Glosses

- Corpus
  - Glosses of dictionaries
- Assumption
  - A word and its definition (gloss) in the dictionary have the same meaning



## Example (Encarta Dictionary)

A B C D E F G H I J K L M N O P Q R S T U V W X Y Z

hurling  
hurly-burly  
Huron (1)  
Huron (2)  
Huron, Lake  
hurrah  
hurray  
Hurrian  
► **hurricane**  
hurricane deck  
hurricane lamp  
hurried  
hurry  
hurry-come-up  
hurry-scurry  
hurry sickness  
hurst

hurricane

hurricane

**hur-ri-cane** [ hŭri kâyn ] (*plural*  
hur-ri-canes)

noun

**Definition:**

1. **severe storm**: a severe tropical storm with torrential rain and extremely strong winds. Hurricanes originate in areas of low pressure in equatorial regions of the Atlantic or Caribbean, and then strengthen, traveling northwest, north, or northeast.

2. **high wind**: a wind of above 119 km (74 mi) per hour, classified as force 12 or above on the Beaufort scale

3. **fast and forceful person or thing**: somebody or something resembling a violent storm in force, speed, or effect

Also available:

World English Dictionary  
Dictionnaire Français

hurricane

severe storm GOOD

high wind GOOD

fast and forceful person or thing BAD

Baidu 百度

## Method

- Prune and reformulate the definitions
  - For a verb *v*, extracts the head of the definition (*h*) and *h*'s adverb modifier *m* as *v*'s paraphrase
    - Kaji et al., 2002
  - Rule based method for extracting the appropriate part from the definition
    - Higashinaka and Nagao, 2002
    - E.g., *w* should not be in *def*; ignore contents in parentheses in *def*; avoid double negation...

Baidu 百度



## References (cont')

- From bilingual parallel corpora
  - Takao et al. 2002. Comparing and Extracting Paraphrasing Words with 2-Way Bilingual Dictionaries.
  - Bannard and Callison-Burch. 2005. Paraphrasing with Bilingual Parallel Corpora.
  - Callison-Burch. 2008. Syntactic Constraints on Paraphrases Extracted from Parallel Corpora.
  - Kok and Brockett. 2010. Hitting the Right Paraphrases in Good Time.
  - Zhao et al. 2008. Pivot Approach for Extracting Paraphrase Patterns from bilingual corpora.



## References (cont')

- From large web corpora
  - Lin. 1998. Automatic Retrieval and Clustering of Similar Words.
  - Lin and Pantel. 2001. Discovery of Inference Rules for Question Answering.
  - Bhagat and Ravichandran. 2008. Large Scale Acquisition of Paraphrases for Learning Surface Patterns.
  - Szepektor and Dagan. 2008. Learning Entailment Rules for Unary Templates.
  - Ravichandran and Hovy. 2002. Learning Surface Text Patterns for a Question Answering System.



## References (cont')

- From other resources
  - Zhao et al. 2010. Paraphrasing with Search Engine Query Logs.
  - Kaji et al. 2002. Verb Paraphrase based on Case Frame Alignment.
  - Higashinaka and Nagao. 2002. Interactive Paraphrasing Based on Linguistic Annotation.



Coffee Break!



## Outline

- **Part 2**
  - **Paraphrase Generation**
    - **Rule based Method**
      - Thesaurus based Method
      - NLG based Method
      - MT based Method
      - Pivot based Method
    - Applications of Paraphrases
    - Evaluation of Paraphrases
    - Conclusions and Future work

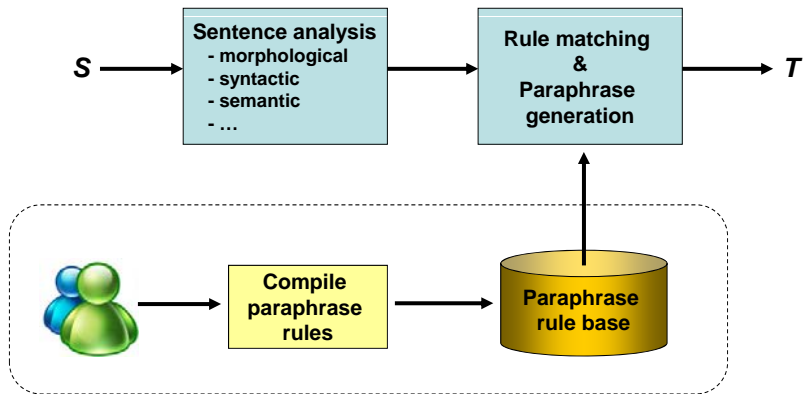


## Rule based Method

- Two types:
  - Based on hand-crafted rules
    - Widely used in early studies of paraphrase generation
    - McKeown, 1979; Zong et al., 2001; Tetsuro et al., 2001; Zhang and Yamamoto, 2002.....
  - Based on automatically extracted rules
    - Extract paraphrase patterns from corpora
    - Barzilay and Lee, 2003, Zhao et al., 2009.....



## Based on Hand-crafted Rules




## Based on Hand-crafted Rules

- Examples of paraphrase rules
  - Change the positions of adverbials
    - *He booked a single room in Beijing **yesterday**.* =>
      - ***Yesterday**, he booked a single room in Beijing.*
  - Split a compound sentence into a group of simple sentences
    - *He booked a single room in Beijing yesterday =>*
      - *He booked a single room in Beijing.*
      - *He booked a single room yesterday.*
      - *He booked a room.*
  - Rewrite a sentence using hand-crafted patterns
    - ***Can** I have a cup of tea? =>*
      - ***May** I have a cup of tea?*
      - ***I would like** a cup of tea, please.*
      - ***Give me** a cup of tea.*



## Based on Automatically Extracted Rules

- Studies on paraphrase patterns extraction has been introduced above
- Some of them have tried to apply the extracted paraphrase patterns in paraphrase generation
  - Complex paraphrase patterns
    - Barzilay and Lee, 2003
    - E.g., 
  - Short and simple paraphrase patterns
    - Zhao et al., 2009
    - E.g., *consider [NN]* and *take [NN] into consideration*



## Pros and Cons

- Methods based on hand-crafted rules
  - Pros
    - Can design paraphrase rules for specific applications and requirements
  - Cons
    - It is time-consuming to construct paraphrase rules
    - Problem of rules conflict
    - Coverage of paraphrase rules is limited
- Methods based on automatically extracted rules
  - Pros
    - Can generate paraphrases with structural changes
  - Cons
    - Coverage of paraphrase rules is limited



## References

- McKeown. 1979. Paraphrasing Using Given and New Information in a Question-Answer System.
- Zong et al. 2001. Approach to Spoken Chinese Paraphrasing Based on Feature Extraction.
- Tetsuro et al.. 2001. KURA: A Transfer-Based Lexico-Structural Paraphrasing Engine.
- Zhang and Yamamoto. 2002. Paraphrasing of Chinese Utterances.
- Barzilay and Lee. 2003. Learning to Paraphrase - An Unsupervised Approach Using Multiple-Sequence Alignment.
- Zhao et al. 2009. Application-driven Statistic Paraphrase Generation.



## Outline

- **Part 2**
  - **Paraphrase Generation**
    - Rule based Method
    - **Thesaurus based Method**
    - NLG based Method
    - MT based Method
    - Pivot based Method
  - Applications of Paraphrases
  - Evaluation of Paraphrases
  - Conclusions and Future work



## Thesaurus based Method

- Also known as lexical substitution
  - Substitute words in a sentence with their synonyms that fit in the given context
  - SemEval-2007: English lexical substitution task
  - SemEval-2010: Cross-lingual lexical substitution
  - Example:
    - *There will be major cuts in the **salaries** of high-level civil servants.*
    - *There will be major cuts in the **wages** of high-level civil servants.*



## Thesaurus based Method

- Include two stages
  - **Stage-1:** extract candidate substitutes from predefined inventories.
    - E.g., WordNet
  - **Stage-2:** find substitutes that fit in the given context
    - Using language model or web data (e.g., Google 5-gram) for evaluating the fitness in the context
    - Disambiguation may also be useful



## Stage-1: Candidate Extraction

- Various thesauri have been tried
  - WordNet:
    - the most commonly used
  - Others:
    - Encarta, Roget, Oxford American Writer's Thesaurus...
- Extracting different information as candidates
  - Synsets (all synsets vs. best synset)
  - Hypernyms, similar-to, also-see...
  - Words in glosses



### Example: WordNet

different  
synsets

WordNet 2.1 Browser

File History Options Help

Search Word: bright Redisplay Overview

Searches for bright: Adjective Adverb Senses:

11 senses of bright

Sense 1  
**bright** (vs. dull) -- (emitting or reflecting light readily or in large amounts; "the sun was bright and hot"; "a bright sunlit room")

=> agleam, gleaming, nitid -- (bright with a steady but subdued shining; "from the plane we saw the city below agleam with lights"; "the gleaming brass on the altar"; "Nereids beneath the nitid moon")

=> aglow(predicate), lambent, lucent, luminous -- (softly bright or radiant; "a house aglow with lights"; "glowing embers"; "lambent tongues of flame"; "the lucent moon"; "a sky luminous with stars")

=> aglitter(predicate), coruscant, fulgid, glinting, glistering, glittering, glittery, scintillant, scintillating, sparkly -- (having brief brilliant points or flashes of light; "bugle beads all aglitter"; "glinting eyes"; "glinting water"; "his glittering eyes were cold and malevolent"; "shop window full of glittering Christmas trees"; "glittery costume jewelry"; "scintillant mica"; "the scintillating stars"; "a dress with sparkly sequins"; "glistering" is an archaic term)

=> beady, beadlike, buttony, buttonlike -- (small and round and shiny like a shiny bead or button; "bright beady eyes"; "black buttony eyes")

=> beaming, beamy, effulgent, radiant, refulgent -- (radiating or as if radiating light; "the beaming sun"; "the effulgent daffodils"; "a radiant sunrise"; "a refulgent sunset")

=> blazing, blinding, dazzling, fulgent, glaring, glary -- (shining intensely; "the blazing sun"; "blinding headlights"; "dazzling snow"; "fulgent patterns of sunlight"; "the glaring sun")

=> bright as a new penny(predicate) -- ((metaphor) shining brightly)

"Synonyms/Related Nouns" search for adjective "bright"

Example:

Encarta

The screenshot shows the Encarta Thesaurus interface. At the top, there are tabs for 'Dictionary', 'Thesaurus', and 'Translations'. Below the tabs is an alphabetical index 'A B C D E F G H I J K L M N O P Q R S T U V W X Y Z'. A search box on the left contains a list of words, with 'bright (adj)' selected and highlighted. To the right of the search box, the word 'bright (adj)' is circled in green. Below this, there are three sections of synonyms and antonyms. The first section lists synonyms: brilliant, vivid, intense, dazzling, light, clear. A red arrow points from the text 'definition of the synset' to the underlined word 'brilliant'. A blue arrow points from the text 'synset' to the underlined word 'brilliant'. The second section lists synonyms: intelligent, clever, smart, brainy, quick, sharp-witted. The third section lists synonyms: cheerful, happy, lively, optimistic, positive, upbeat, sunny, perky. Each section also includes an antonym: dark, unintelligent, and gloomy.

## Stage-2: Substitute Selection

- Rank the candidates and select the one fits best in the given context
- Context constraints
  - Semantic constraints
    - Select substitutes with the correct meaning wrt the given context
  - Syntactic constraints
    - The sentence generated after substitution should keep grammatical



## SubFinder: A Lexical Substitution System

- SubFinder
  - University of North Texas
  - Performs well in SemEval-2007 English lexical substitution task
- Candidate extraction
  - WordNet
  - Encarta
  - Others
    - Prove to be useless



## SubFinder: A Lexical Substitution System

- Substitute selection (**5 ranking methods R1~R5**)
  - Language model
    - Google 1T 5-gram (**R1**)
    - Query search engine (**R2**)
  - Latent semantic analysis (LSA) (**R3**)
    - Rank a candidate by its relatedness to the context sentence
  - Word sense disambiguation (WSD) (**R4**)
    - Disambiguate the target word and select the synset of the right sense
  - Pivot approach (**R5**)
    - Check whether a candidate substitute can be generated via a 2-way translation



## SubFinder: A Lexical Substitution System

- Combine R1~R5:
  - Voting mechanism

$$score(c_i) = \sum_{m \in rankings} \lambda_m \frac{1}{r_{c_i}^m}$$

→ Ranks according to R1-R5

- Contribution of each ranking method is not analyzed ☹️



## Pros and Cons

- Pros
  - Based on existing inventories
- Cons
  - Cannot generate structural paraphrases
  - Language limitation
- Question
  - *How to merge different thesauri?*
    - Thesauri have different forms of synset clustering



## References

- McCarthy and Navigli. 2007. SemEval-2007 Task 10: English Lexical Substitution Task.
- Hassan et al. 2007. UNT: SubFinder: Combining Knowledge Sources for Automatic Lexical Substitution.
- Yuret. 2007. KU: Word Sense Disambiguation by Substitution.
- Giuliano et al. 2007. FBK-irst: Lexical Substitution Task Exploiting Domain and Syntagmatic Coherence.
- Martinez et al. 2007. MELB-MKB: Lexical Substitution System based on Relatives in Context.
- Kauchak and Barzilay. 2006. Paraphrasing for Automatic Evaluation.



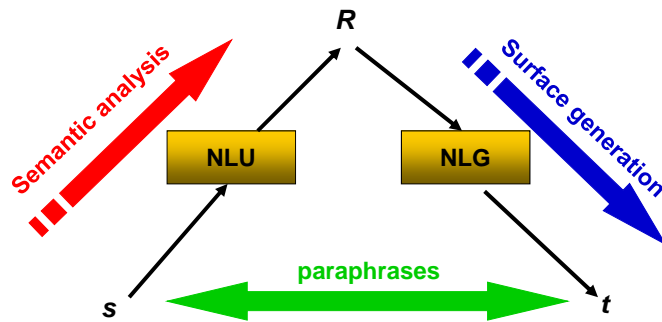
## Outline

- **Part 2**
  - **Paraphrase Generation**
    - Rule based Method
    - Thesaurus based Method
    - **NLG based Method**
    - MT based Method
    - Pivot based Method
  - Applications of Paraphrases
  - Evaluation of Paraphrases
  - Conclusions and Future work



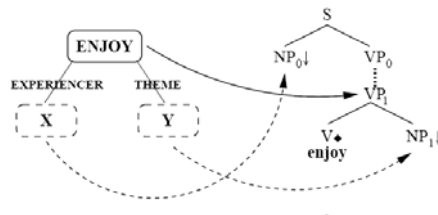
## Overview

- Two steps
  - (1) analysis and (2) generation



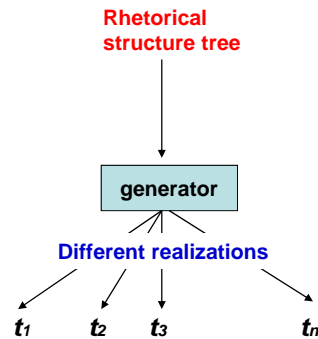
## NLG based Methods

- Kozlowski et al., 2003
  - Generate single-sentence paraphrases
  - Input: predicate/argument structure
    - Not natural language sentences ⊕
  - Based on lexico-grammatical resources
    - Map elementary semantic structures with syntactic realization



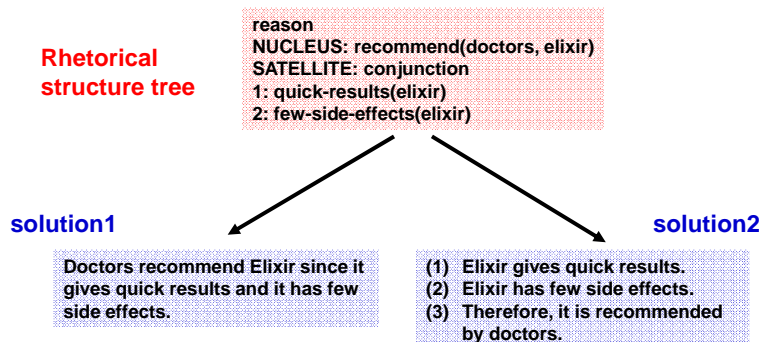
## NLG based Methods (cont')

- Power and Scott, 2005
  - Concerning *larger-scale* paraphrases
    - Paraphrases of multiple sentences or even the whole text
    - Paraphrases vary not only at lexical and syntactic levels, but also in document structure and layout
  - Problem:
    - The input is not natural language textsⓂ



## NLG based Methods (cont')

- Power and Scott, 2005 (cont')
  - Example:



## NLG based Methods (cont')

- Fujita et al., 2005
  - Paraphrase *light-verb constructions* (LVC) in sentences
    - LVC: consists of a light-verb that syntactically governs a deverbal noun
  - Semantic representation
    - LCS: Lexical Conceptual Structure
  - Procedure
    - Semantic analysis
    - Semantic transformation
    - Surface generation



## Pros and Cons

- Pros
  - It simulates human being's behavior when generating paraphrases:
    - Step-1: understand the meaning of a sentence
    - Step-2: generate a new sentence expressing the meaning
- Cons
  - Both deep analysis of sentences and NLG are difficult to realize



## References

- Kozlowski et al. 2003. Generation of single-sentence paraphrases from predicate/argument structure using lexico-grammatical resources.
- Power and Scott. 2005. Automatic generation of large-scale paraphrases.
- Fujita et al. 2005. Exploiting Lexical Conceptual Structure for Paraphrase Generation.

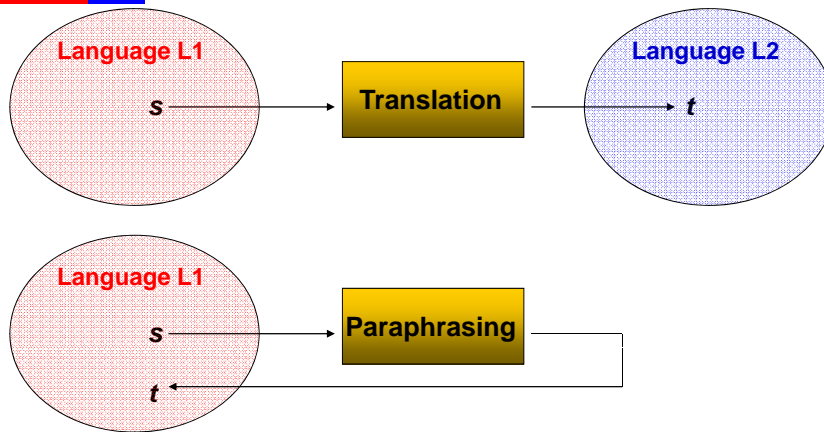


## Outline

- **Part 2**
  - **Paraphrase Generation**
    - Rule based Method
    - Thesaurus based Method
    - NLG based Method
    - **MT based Method**
    - Pivot based Method
  - Applications of Paraphrases
  - Evaluation of Paraphrases
  - Conclusions and Future work



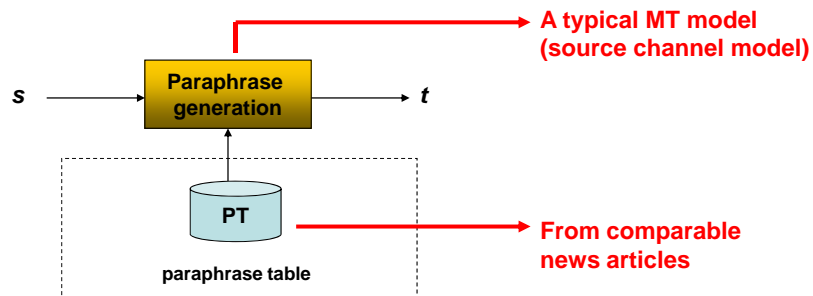
## Machine Translation vs. Paraphrase Generation



For both machine translation and paraphrase generation:  
(1)  $t$  should preserve the meaning of  $s$   
(2)  $t$  should be a fluent sentence

## Paraphrase Generation as Machine Translation

- Quirk et al., 2004
  - First recast paraphrase generation as a monolingual machine translation task



## Paraphrase Generation as Machine Translation (cont')

- Model

- Source channel model

$$\begin{aligned} t^* &= \arg \max_t p(t | s) \\ &= \arg \max_t p(s | t) p(t) \end{aligned}$$

→ Language model

↓  
"Translation" model  
(based on a phrasal paraphrase table)



## Paraphrase Generation as Machine Translation (cont')

- Paraphrase table

- Monolingual parallel sentences
  - Extracted from comparable news articles
  - 139K pairs
- Word alignment & phrase pair extraction
  - With Giza++

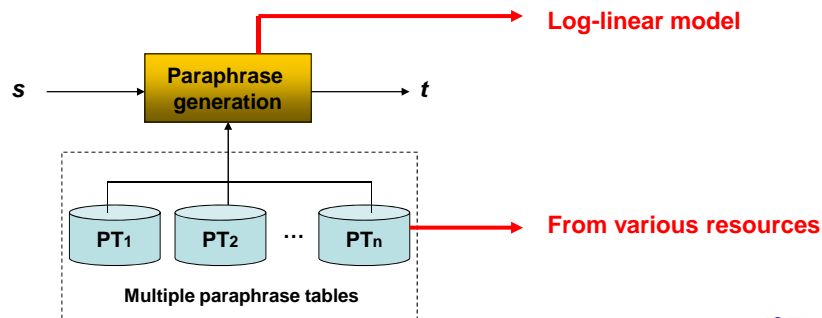
- Limitation

- Lack of monolingual parallel corpora to train the paraphrase table!!!



## Paraphrase Generation as Machine Translation (cont')

- Zhao et al., 2008
  - Combine multiple resources to train the paraphrase table



## Paraphrase Generation as Machine Translation (cont')

- Model
  - Log-linear model

$$t^* = \arg \max_t \left\{ \sum_{i=1}^N \lambda_{TM\_i} h_{TM\_i}(t, s) + \lambda_{LM} h_{LM}(t, s) \right\}$$

N paraphrase tables, each feature corresponds to a paraphrase table

Language model

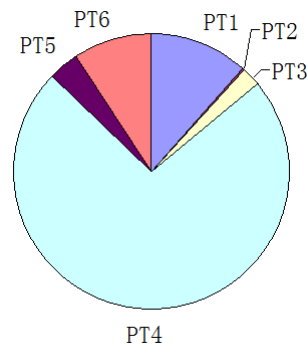
$$h_{TM\_i}(t, s) = \log \prod_{k=1}^{K_i} score_i(t_k, s_k)$$



## Paraphrase Generation as Machine Translation (cont')

- Paraphrase tables
  - **PT1**: from word clusters (Lin, 1998)
  - **PT2**: from monolingual parallel corpora
  - **PT3**: from monolingual comparable corpora
  - **PT4**: from bilingual parallel corpora
  - **PT5**: from Encarta dictionary glosses
  - **PT6**: from clusters of similar user queries

- Volumes of the PTs:



Proves most useful!



## Paraphrase Generation vs. Machine Translation

- Differences between machine translation and paraphrase generation (Zhao et al., 2009):

### Machine Translation (MT)

### Paraphrase Generation (PG)

MT has a unique purpose

PG has distinct purposes in different applications

In MT, all words in a sentence should be translated

In PG, not all words need to be paraphrased

In MT, the bilingual parallel data are easy to collect

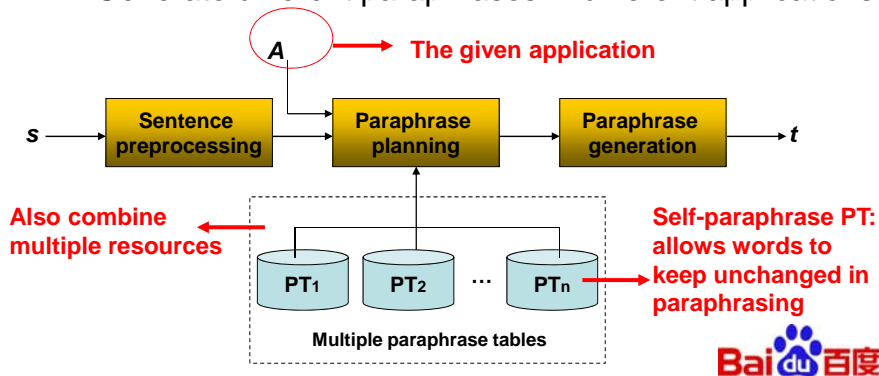
In PG, multiple resources need to be combined

In MT, automatic evaluation metrics (e.g., BLEU) are available

In PG, automatic evaluation metrics are not available

## Application-driven Statistical Paraphrase Generation

- Zhao et al., 2009
  - Propose a statistical model for paraphrase generation
  - Generate different paraphrases in different applications



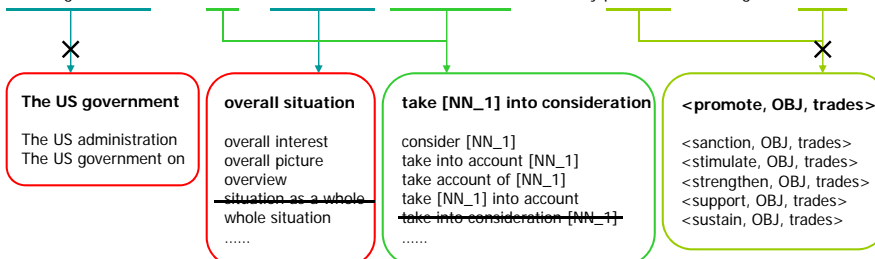
## Application-driven Statistical Paraphrase Generation (cont')

- Paraphrase planning
  - When an application  $A$  is given, only the paraphrase pairs that can achieve  $A$  are kept

### Example:

Paraphrase application: sentence compression

The US government should take the overall situation into consideration and actively promote bilateral high-tech trades.



## Application-driven Statistical Paraphrase Generation (cont')

- Model:
  - Log-linear model

$$p(\mathbf{t} | \mathbf{s}) = \sum_{k=1}^K (\lambda_k \sum_{k_i} \log \phi_k(\bar{s}_{k_i}, \bar{t}_{k_i})) \longrightarrow \text{Paraphrase model}$$
$$+ \lambda_{lm} \sum_{j=1}^J \log p(t_j | t_{j-2} t_{j-1}) \longrightarrow \text{Language model}$$
$$+ \lambda_{um} \sum_{i=1}^I \mu(\bar{s}_i, \bar{t}_i) \longrightarrow \text{Usability model (defined for each application)}$$



## References

- Lin. 1998. Automatic Retrieval and Clustering of Similar Words.
- Quirk et al. 2004. Monolingual Machine Translation for Paraphrase Generation.
- Finch et al. 2004. Paraphrasing as Machine Translation.
- Zhao et al. 2008. Combining Multiple Resources to Improve SMT-based Paraphrasing Model.
- Zhao et al. 2009. Application-driven Statistical Paraphrase Generation.



## Outline

- Part 2

- Paraphrase Generation

- Rule based Method
    - Thesaurus based Method
    - NLG based Method
    - MT based Method
    - **Pivot based Method**

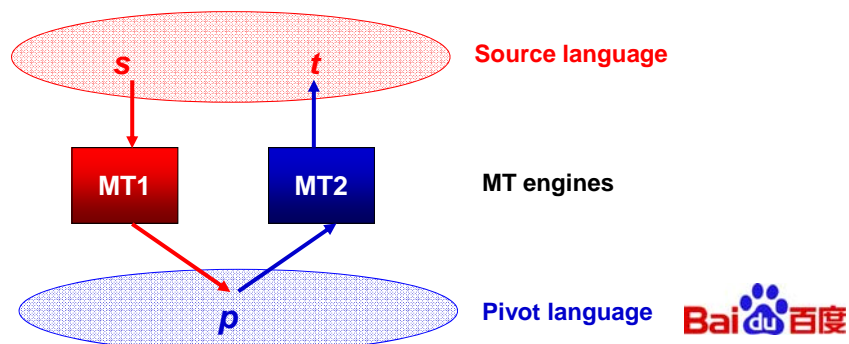
- Applications of Paraphrases
    - Evaluation of Paraphrases
    - Conclusions and Future work



## Overview

- Basic idea

- We can generate a paraphrase  $t$  for a sentence  $s$  by translating  $s$  into a foreign language, and then translating it back into the source language.



## Overview (cont')

- Example:

English What toxins are most **hazardous** to **expectant mothers**?

Italian Che tossine sono più pericolose alle donne incinte?

English What toxins are more **dangerous** to **pregnant women**?

- Single-pivot

- Using a single pivot language

- Multi-pivot

- Using multiple pivot languages



## Pivot based Methods

- Duboue and Chu-Carroll, 2006

- Applied in QA systems

- Paraphrase the input questions so as to improve the coverage in answer extraction

- Pivot languages

- 11

- MT engines

- 2: Babelfish (**B**) and Google MT (**G**)
- 4 combinations: B+B, B+G, G+G, G+B



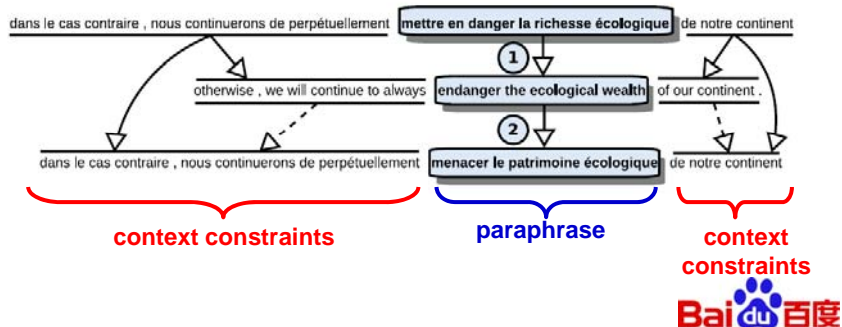
## Pivot based Methods (cont')

- Duboue and Chu-Carroll, 2006 (cont')
  - Given a list of automatically generated paraphrases, we need to select a **best** one.
    - For QA, we need to select the paraphrase that can find the answer more easily than the original question.

Features for paraphrase selection (in a classification framework)	
SUM IDF	The sum of the IDF scores for all terms in the original question and the paraphrase. (prefer paraphrases with more informative terms)
Lengths	Number of query terms for each of the paraphrase and the original question. (prefer shorter paraphrases)
Cosine Distance	The distance between the vectors of both questions, IDF-weighted. (filter paraphrases that diverge too much from the original)
Answer Types	Whether answer types, as predicted by the question analyzer, are the same or overlap. (the answer type should be the same)

## Pivot based Methods (cont')

- Max, 2009
  - Paraphrasing sub-sentential fragments
    - Allows the exploitation of context during both source-pivot translation and pivot-source back-translation



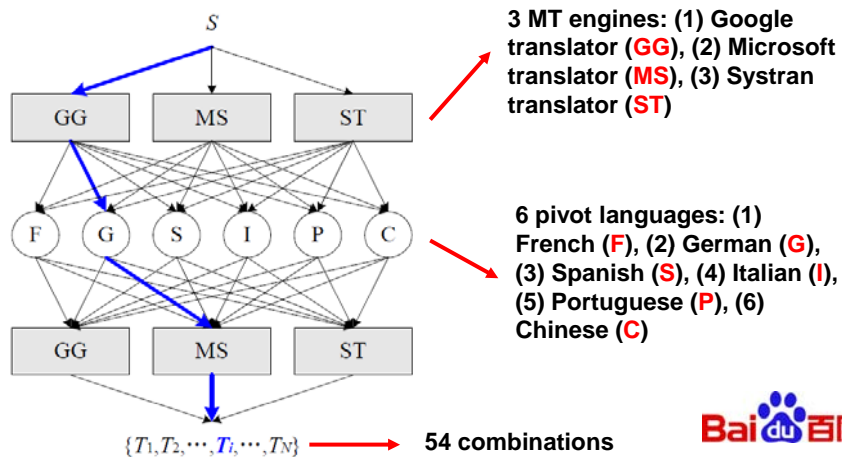
## Pivot based Methods (cont')

- Max, 2009 (cont')
  - Application
    - Text revision
  - Pivot language
    - English
      - Paraphrases are acquired for French sub-sentences
  - MT engine
    - Source context aware SMT (Stroppa et al., 2007)



## Pivot based Methods (cont')

- Zhao et al., 2010



## Pivot based Methods (cont')

- Zhao et al., 2010 (cont')
  - Produce a high-quality paraphrase using the list of candidates

Source	he said there will be major cuts in the salaries of high-level civil servants
(GG, G, MS)	he said there <b>are significant</b> cuts in the salaries of high-level <b>officials</b>
(GG, F, GG)	he said there will be <b>significant</b> cuts in the salaries of <b>top civil level</b>
(GG, P, GG)	he said there will be <b>big</b> cuts in salaries of <b>high-level civil</b>
(MS, C, MS)	he said <b>that</b> there will be <b>a major senior civil service pay cut</b>
(MS, S, GG)	he said there will be <b>significant</b> cuts in the salaries of <b>senior officials</b>
(MS, F, ST)	he said there will be <b>great</b> cuts in the <b>wages</b> of the <b>high level civils servant</b>
(ST, G, GG)	he said <b>that</b> there <b>are</b> major cuts in the salaries of <b>senior government officials</b>
.....	.....

■ Good paraphrases    ■ Bad paraphrases



## Pivot based Methods (cont')

- Zhao et al., 2010 (cont')
  - Two techniques for producing high-quality paraphrases using the candidates
    - Selection-based technique
      - Select a best paraphrase from the 54 candidates based on Minimum Bayes Risk (MBR)
    - Decoding-based technique
      - Train a MT model using the 54 candidates, and generates a new paraphrase with it



## References

- Duboue and Chu-Carroll. 2006. Answering the Question You Wish They Had Asked: The Impact of Paraphrasing for Question Answering.
- Stroppa et al. 2007. Exploiting Source Similarity for SMT using Context-informed Features.
- Max. 2009. Sub-sentential Paraphrasing by Contextual Pivot Translation.
- Zhao et al. 2010. Leveraging Multiple MT Engines for Paraphrase Generation.



## Outline

- **Part 2**
  - Paraphrase Generation
  - Applications of Paraphrases**
    - Paraphrasing for MT
    - Other Applications
  - Evaluation of Paraphrases
  - Conclusions and Future work



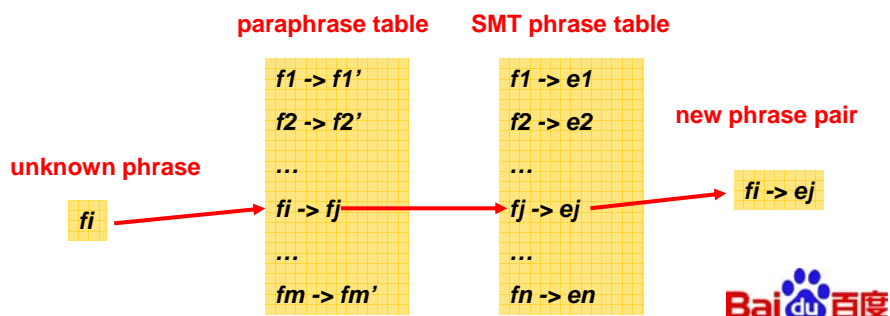
## Paraphrasing for MT

- Applications:
  - Translate unknown terms (phrases)
  - Expand training data
  - Rewrite input sentences
  - Improve automatic evaluation
  - Tune parameters



## Translate Unknown Terms (Phrases)

- Basic idea:
  - In SMT, when encountering an unknown source term (phrase), we can substitute a paraphrase for it and then proceed using the translation of that paraphrase



## Translate Unknown Terms (Phrases) (cont')

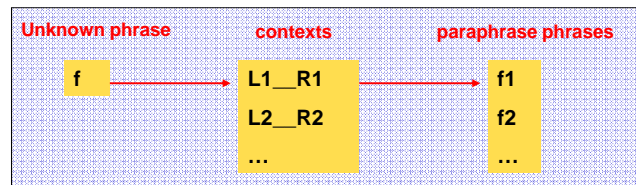
- Callison-Burch et al., 2006
  - Paraphrases are extracted from bilingual parallel corpora using the pivot approach
  - New phrase pairs generated through paraphrasing are incorporated into the phrase table
    - The paraphrase probability is added as a new feature function:

**paraphrase probability** ← 
$$h(e, f_1) = \begin{cases} p(f_2 | f_1) & \text{If phrase table entry } (e, f_1) \\ & \text{is generated from } (e, f_2) \\ 1 & \text{Otherwise} \end{cases}$$



## Translate Unknown Terms (Phrases) (cont')

- Marton et al., 2009
  - Paraphrases are extracted from monolingual corpora, based on distributional hypothesis



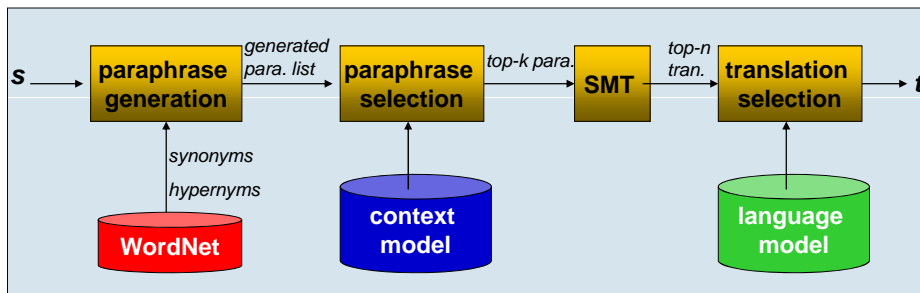
- Combine the new phrase pairs in the phrase table

**Context similarity** ← 
$$h(e, f_1) = \begin{cases} psim(DP_{f_1}, DP_{f_2}) & \text{If phrase table entry } (e, f_1) \\ & \text{is generated from } (e, f_2) \\ 1 & \text{Otherwise} \end{cases}$$



## Translate Unknown Terms (Phrases) (cont')

- Mirkin et al., 2009
  - Use not only paraphrases but also entailment rules
    - From WordNet
      - Paraphrases: *synonyms* in WordNet
      - Entailment rules: *hypernyms* in WordNet



## Translate Unknown Terms (Phrases) (cont')

- Onishi et al., 2010
  - Using paraphrase lattices for SMT
    - **Step-1:** Paraphrase the input sentence, and generate a paraphrase lattice
      - Paraphrases are extracted from bilingual parallel corpora based on the pivot approach
    - **Step-2:** Give the paraphrase lattice as the input to the lattice decoder

## Translate Unknown Terms (Phrases) (cont')

- Effectiveness

- When the training data of SMT is small

- Effective 😊

- Problem of unknown terms is more serious when the training data is small

- When the training data of SMT is large

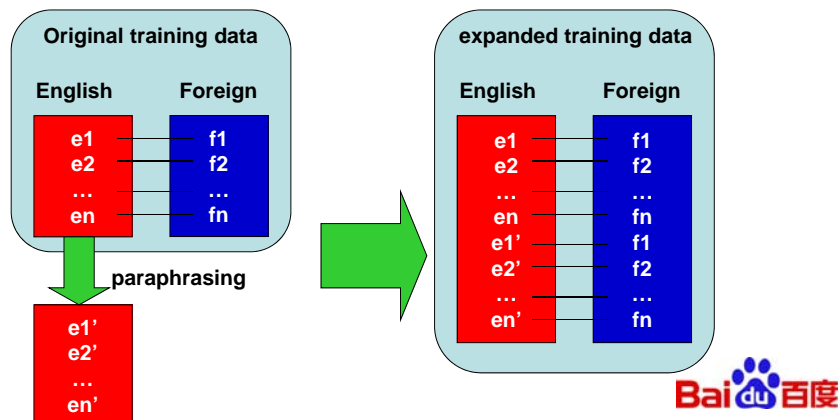
- Ineffective ☹️

- Unknown terms can be covered by adding more training data



## Expand Training Data

- Enlarge training data via paraphrasing the source-side sentences in the parallel corpus



## Rewrite Input Sentences

- Paraphrase the sentence to be translated, so as to make it more translatable
  - Yamamoto, 2002; Zhang and Yamamoto, 2002
    - Rule-based Paraphraser for simplifying the source sentences
  - Shimohata et al., 2004
    - Shorten long sentences and sentences with redundant information in a speech translation system



## Improve Automatic Evaluation

- Automatic evaluation of MT
  - Based on counting the overlaps between the references and machine outputs
    - E.g., BLEU, NIST...
  - Only computing the surface similarity is limited
    - A meaning may be expressed in a way that is not included in the references
  - Human references are expensive to produce
  - **Solution: paraphrase the references so as to include as many correct expressions as possible!**



## Improve Automatic Evaluation (cont')

- Kauchak and Barzilay, 2006

- Find a paraphrase of the reference that is closer in wording to the system output
  - Extract candidates from WordNet synonyms

System output

It is **hard** to believe that such tremendous changes have taken **place** for those people and **lands** that I have never stopped missing while living abroad.

Reference

For someone born here but has been sentimentally attached to a foreign country far from **home**, it is **difficult** to believe this kind of changes.

Correct Wrong

- Filter the invalid substitution given the context
  - Binary classification
    - Features: context n-grams and local collocations



## Improve Automatic Evaluation (cont')

- Zhou et al., 2006

- ParaEval: Compute the similarity of reference and system output using paraphrases
  - Paraphrases are learned from bilingual parallel corpora with a pivot approach
- Two-tier matching strategy for SMT evaluation
  - First tier: paraphrase match
  - Second tier: unigram match for words not matched by paraphrases



## Tune Parameters

- Madnani et al. 2007
  - Similar to the studies using paraphrases to improve automatic evaluation of MT
  - Parameter tuning in SMT also needs references
    - Parameter estimation of SMT:
      - optimize BLEU on a development set
  - Expand the references automatically via paraphrasing
    - Paraphrase generation
      - Paraphrase resources are acquired based on a pivot approach
      - Recast paraphrase generation as a monolingual MT problem and decode with a typical SMT decoder



## References

- Translate unknown terms (phrases)
  - Callison-Burch et al. 2006. Improved Statistical Machine Translation Using Paraphrases.
  - Marton et al. 2009. Improved Statistical Machine Translation Using Monolingually-Derived Paraphrases.
  - Mirkin et al. 2009. Source-Language Entailment Modeling for Translating Unknown Terms.
  - Onishi et al. 2010. Paraphrase Lattice for Statistical Machine Translation.
- Expand training data
  - Nakov. 2008. Improved Statistical Machine Translation Using Monolingual Paraphrases.
  - Bond et al. 2008. Improving Statistical Machine Translation by Paraphrasing the Training Data.



## References (cont')

- Rewrite input sentences
  - Yamamoto. 2002. Machine Translation by Interaction between Paraphraser and Transfer.
  - Zhang and Yamamoto. 2002. Paraphrasing of Chinese Utterances.
  - Shimohata et al. 2004. Building a Paraphrase Corpus for Speech Translation.
- Improve automatic evaluation
  - Kauchak and Barzilay. 2006. Paraphrasing for Automatic Evaluation.
  - Zhou et al. 2006. Re-evaluating Machine Translation Results with Paraphrase Support.
- Tune parameters
  - Madnani et al. 2007. Using Paraphrases for Parameter Tuning in Statistical Machine Translation.



## Outline

- **Part 2**
  - Paraphrase Generation
  - **Applications of Paraphrases**
    - Paraphrasing for MT
    - **Other Applications**
  - Evaluation of Paraphrases
  - Conclusions and Future work



## Paraphrasing for QA

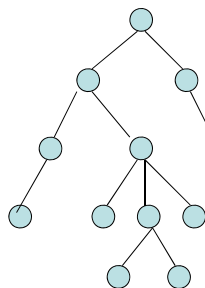
- Goal:
  - Alleviate the problem of *word mismatch* between questions and answers
- Two directions:
  - Paraphrase questions
    - Rewrite a question into a group of paraphrases, so as to improve the coverage in answer extraction
  - Paraphrase answer extraction patterns
    - Generate answer extraction patterns as many as possible



## Paraphrasing for QA

- Ravichandran and Hovy, 2002.
  - Mining paraphrase patterns from the web
    - Using hand-crafted seeds (e.g., (*Mozart*, 1756) for BIRTHDAY)
    - Mining patterns containing the seeds

Question taxonomy



BIRTHDAY	
1.00	<NAME> (<ANSWER> -)
0.85	<NAME> was born on <ANSWER>
0.60	<NAME> was born in <ANSWER>
0.59	<NAME> was born <ANSWER>
0.53	<ANSWER> <NAME> was born
0.50	- <NAME> (<ANSWER> -)
0.36	<NAME> (<ANSWER> -)

Given seed (Mozart, 1756)



## Paraphrasing for Summarization

- Improve automatic evaluation of summaries
  - Zhou et al., 2006
  - Similar to the automatic evaluation of MT
    - Measure the similarity between references and system outputs using paraphrase match as well as exact match
- Improve sentence clustering
  - Barzilay et al., 1999
  - Considering paraphrase match when Computing sentence similarity



## Other Applications

- Paraphrasing for NLG
  - Text revision and transformation
    - Dras, 1997
      - Text transformation in order to meet external constraints, such as length and readability
- Paraphrasing for IR
  - Query rewriting
    - Zukerman and Raskutti. 2002.
      - Paraphrase user queries with WordNet synonyms



## Other Applications (cont')

- Writing style transformation
  - Kaji et al., 2004
    - Paraphrasing predicates from written language to spoken language
- Text simplification
  - Carroll et al. 1999
    - Simplifying texts for language-impaired readers or non-native speakers
- Identify plagiarism
  - Uzuner et al. 2005
    - Using paraphrases to better identify plagiarism



## References

- Paraphrasing for QA
  - Ravichandran and Hovy. 2002. Learning Surface Text Patterns for a Question Answering System.
  - Duboue and Chu-Carroll. 2006. Answering the Question You Wish They Had Asked: The Impact of Paraphrasing for Question Answering.
- Paraphrasing for summarization
  - Barzilay et al. 1999. Information Fusion in the Context of Multi-Document Summarization.
  - Zhou et al. 2006. ParaEval: Using Paraphrases to Evaluate Summaries Automatically.
- Paraphrasing for NLG
  - Dras. 1997. Reluctant Paraphrase: Textual Restructuring under an Optimisation Model.



## References (cont')

- Paraphrasing for IR
  - Zukerman and Raskutti. 2002. Lexical Query Paraphrasing for Document Retrieval.
- Writing style transformation
  - Kaji et al. 2004. Paraphrasing Predicates from Written Language to Spoken Language Using the Web.
- Text simplification
  - Carroll et al. 1999. Simplifying Text for Language-Impaired Readers.
- Identify plagiarism
  - Uzuner et al. 2005. Using Syntactic Information to Identify Plagiarism.



## Outline

- **Part 2**
  - Paraphrase Generation
  - Applications of Paraphrases
    - Paraphrasing for MT
    - Other Applications
  - **Evaluation of Paraphrases**
  - Conclusions and Future work



## Evaluation of Paraphrases

- No widely accepted evaluation criteria☹
  - **Problem-1:** Researchers define various evaluation methods in their studies
    - Difficult to make a direct comparison among different works
  - **Problem-2:** Human evaluation is commonly used
    - Human evaluation is rather subjective
    - Difficult to replicate



## Evaluation of Paraphrase Identification

- Human evaluation
- Automatic evaluation
  - Brockett and Dolan, 2005
  - **A**lignment **E**rror **R**ate (AER)
    - AER is indicative of how far the corpus is from providing a solution under a standard SMT tool

$$AER = \frac{|A \cap P| + |A \cap S|}{|A + S|}$$

Automatic alignment

POSSIBLE + SURE alignment in the gold standard

SURE alignment in the gold standard



## Evaluation of Lexical Substitution

- Automatic evaluation
  - McCarthy and Navigli, 2007
  - Construction of gold standard data
    - Five annotators, who are native speakers
    - For each test word, each annotator provides up to three substitutes
  - Evaluation:
    - Precision and Recall



## Evaluation of Paraphrase Phrases

- Human evaluation
  - Ask judges:
    - Whether paraphrases were approximately conceptual equivalent
    - Whether the paraphrases were roughly interchangeable given the genre
    - Whether the substitutions preserved the meaning and remained grammatical
    - .....
  - The criteria above are vaguely defined and not easy to reproduce



## Evaluation of Paraphrase Phrases (cont')

- Automatic evaluation
  - Callison-Burch et al., 2008
  - Data:
    - Parallel sentences, in which paraphrases are annotated through manual alignment (gold standard)
  - Two fashions of evaluation
    - Calculate how well an automatic paraphrasing technique can **align** the paraphrases in a sentence pair
    - Calculate the **lower-bound precision** and **relative recall** of a paraphrasing technique (which extracts paraphrases from other resources)



## Evaluation of Paraphrase Phrases (cont')

- Alignment precision and recall
- Lower-bound precision and relative recall

System alignment      Manual alignment      Paraphrase acquired with a method MET      Paraphrase in the gold standard set

$$Align_{Prec} = \frac{\sum_{\langle e_1, e_2 \rangle \in C} |PP(e_1, e_2, S) \cap PP(e_1, e_2, M)|}{\sum_{\langle e_1, e_2 \rangle \in C} |PP(e_1, e_2, S)|}$$

$$LB - Precision = \frac{\sum_{\langle s, G \rangle \in C} \sum_{p \in s} |para_{MET}(p, s) \cap para_{REF}(p_1, s, G)|}{\sum_{\langle s, G \rangle \in C} \sum_{p \in s} |para_{MET}(p, s)|}$$

$$Align_{Recall} = \frac{\sum_{\langle e_1, e_2 \rangle \in C} |PP(e_1, e_2, S) \cap PP(e_1, e_2, M)|}{\sum_{\langle e_1, e_2 \rangle \in C} |PP(e_1, e_2, M)|}$$

$$Rel - Recall = \frac{\sum_{\langle s, G \rangle \in C} \sum_{p \in s} |para_{MET}(p, s) \cap para_{REF}(p_1, s, G)|}{\sum_{\langle s, G \rangle \in C} \sum_{p \in s} |para_{REF}(p_1, s, G)|}$$

## Evaluation of Paraphrase Patterns

- Human evaluation
  - Paraphrase patterns cannot be evaluated without context information
    - E.g., *X acquire Y, X buy Y*
      - Correct or not? It depends on what fill in slots *X* and *Y*
    - Common view:
      - A pair of paraphrase patterns is considered correct if the judge could think of contexts under which it holds
    - Problem:
      - Different judges may think of totally distinct contexts, thus the agreement among the judges could be low



## Evaluation of Paraphrase Patterns (cont')

- Szpektor et al., 2007
  - Evaluate paraphrase patterns (and entailment rules) with instances rather than directly evaluate patterns
    - Judges are presented not only with a pair of patterns, but also a sample of sentences that match its left-hand side
    - Judges assess whether two patterns are paraphrases under each specific example
    - A pair of paraphrase patterns is considered as correct only when the percentage of correct examples is high enough



## Evaluation of Paraphrase Generation

- Human evaluation
  - Similar to human evaluation of SMT
  - Criteria (Zhao et al., 2009, 2010)
    - **Adequacy**: If the meaning of the source sentence is preserved in the paraphrase?
    - **Fluency**: if the generated paraphrase is well-formed?
    - **Usability** (Zhao et al., 2009): If the paraphrase meets the requirement of the given application?
    - **Paraphrase rate** (Zhao et al., 2009): How different the paraphrase is from the source sentence?



## Evaluation of Paraphrase Generation (cont')

- Three scales for adequacy, fluency, and usability (Zhao et al., 2009)

Adequacy	1	The meaning is evidently changed.
	2	The meaning is generally preserved.
	3	The meaning is completely preserved.
Fluency	1	The paraphrase t is incomprehensible.
	2	t is comprehensible.
	3	t is a flawless sentence.
Usability	1	t is opposite to the application purpose.
	2	t does not achieve the application.
	3	t achieves the application.

- Five scales for adequacy and fluency (Zhao et al., 2010)



## Evaluation of Paraphrase Generation (cont')

- Paraphrase rate (Zhao et al., 2010):

- PR-1: based on word overlap rate

$$PR1(T) = 1 - \frac{OL(S,T)}{L(S)}$$

→ Word overlap rate  
→ Number of words in the source sen.

- PR-2: based on edit distance

$$PR2(T) = \frac{ED(S,T)}{L(S)}$$

→ Edit distance



## Evaluation of Paraphrase Generation (cont')

- Two questions:

- Q1: Why not adopt automatic MT methods here, e.g., BLEU, NIST, TER...?

- Reason-1: It is much more difficult to construct human references in paraphrase generation than MT
- Reason-2: Paraphrases that change less will get larger scores in criteria like BLEU

- Q2: How to combine the evaluation of *paraphrase quality* and *paraphrase rate*?

- They seem to be incompatible



## Evaluation within Applications

- Evaluate the role of a paraphrasing module within a certain application system
  - E.g., in MT, examine whether a paraphrasing module helps to alleviate the unknown term problem
  - E.g., in QA, whether paraphrasing the answer patterns can improve the coverage of answer extraction
- Problems:
  - Whether the result can hold for a different application?
  - How to evaluate the role of the paraphrase module independently (not influenced by other modules)?



## References

- Brockett and Dolan. 2005. Support Vector Machines for Paraphrase Identification.
- Szpektor et al. 2007. Instance-based Evaluation of Entailment Rule Acquisition.
- McCarthy and Navigli. 2007. SemEval-2007 Task 10: English Lexical Substitution Task.
- Callison-Burch et al. 2008. ParaMetric: An Automatic Evaluation Metric for Paraphrasing.
- Zhao et al. 2009. Application-driven Statistical Paraphrase Generation.
- Zhao et al. 2010. Leveraging Multiple MT Engines for Paraphrase Generation.



## Outline

- **Part 2**

- Paraphrase Generation
- Applications of Paraphrases
  - Paraphrasing for MT
  - Other Applications
- Evaluation of Paraphrases
- **Conclusions and Future work**



## Conclusions and Future Work

- **Conclusions**

- Paraphrasing is important in various research areas
- Many different kinds of corpora and data resources have been investigated for paraphrase extraction
- Paraphrase generation is a task similar to MT, but not the same
- Paraphrase evaluation is problematic. Automatic evaluation methods are in need



## Conclusions and Future Work (cont')

- Future work
  - Paraphrase extraction
    - Improve the quality of the extracted paraphrases
  - Paraphrase generation
    - Application-driven paraphrase generation
  - Paraphrase application
    - Apply paraphrasing techniques in commercial NLP systems, rather than merely in labs
  - Paraphrase evaluation
    - Come up with evaluation methods that can be widely accepted



Thanks!

QA

