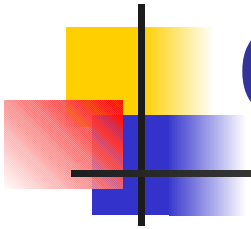
NUDT Machine Translation System for IWSLT2007

Presenter: Boxing Chen

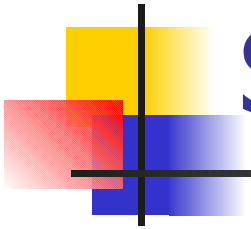
Authors: Wen-Han Chao & Zhou-Jun Li

National University of Defense Technology, China



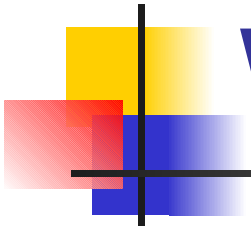
Outline

- System Description
  - Word Alignment
  - Phrase Pair Extracting
  - Example-Based Decoder
- Experiments
- Future Works



System Description

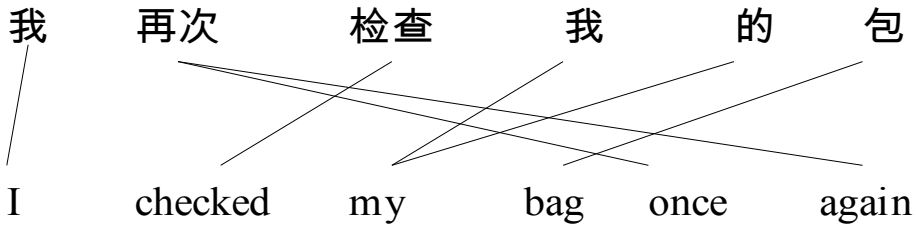
- Three Main Components
  - Word Alignment
  - Phrase Pair Extraction
  - Example-based Decoder



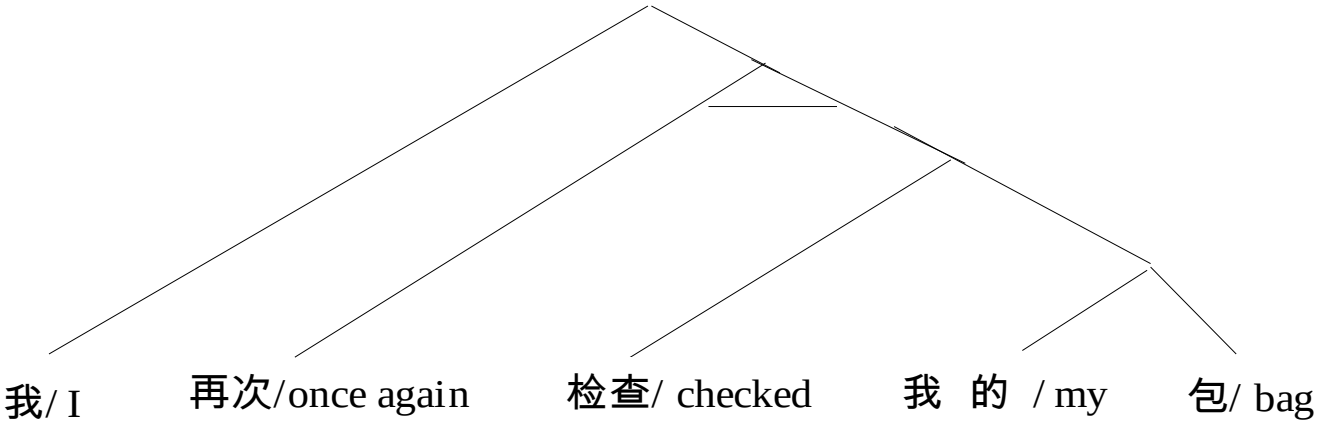
Word Alignment - (1)

- DP algorithm and log-linear model to compute the word alignment
  - Feature functions
    - ITG constraint
    - Word co-occurrence probability
    - Distortion model
- Word alignment
  - a binary branching tree
  - satisfy ITG constraint
  - many-to-many alignment

Word Alignment - (2)



(a) A valid word alignment example

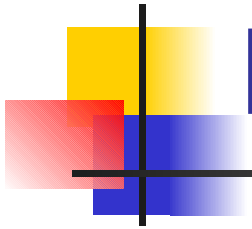


(b) An ITG tree for the word alignment (a)

Phrase Pair Extraction

- Extracting phrase pairs
  - each constituent sub-tree is a phrase pair (block)
- Building models
  - Direct and inverse word- and phrase-based TM
  - Reordering model over all block pairs

Chinese	English
我	i
再次	once again
检查	checked
我 的	my
包	Bag
我的 包	my bag
检查 我的 包	checked my bag
再次 检查 我的 包	checked my bag once again
我 再次 检查 我的 包	i checked my bag once again



Decoding

- Log-linear model used to score the partial hypotheses
- Baseline: CKY style decoder
- Example-based decoder
  - Examples retrieval
  - Generation: matching and merging

Example-based Decoder - (1)

- Examples retrieval
  - Collect source blocks from input
  - Collect (top N) phrase pairs from the TM for each source block
  - If an example (sentence-pair) can be matched by at least one phrase-pair, it is a valid example.
  - Keep top M examples for a specific phrase-pair

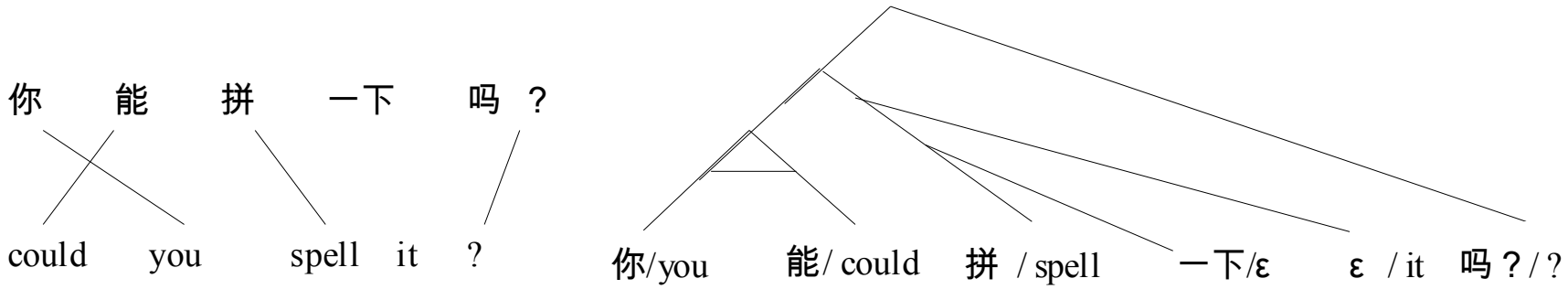
Example-based Decoder - (2)

- Matching
  - Matching the input with each example
    - get a translation template
    - the translation template forms a new ITG tree.
  - Decoding each un-translated string (child input) iteratively
    - get a set of child translation templates.

Example-based Decoder - (3)

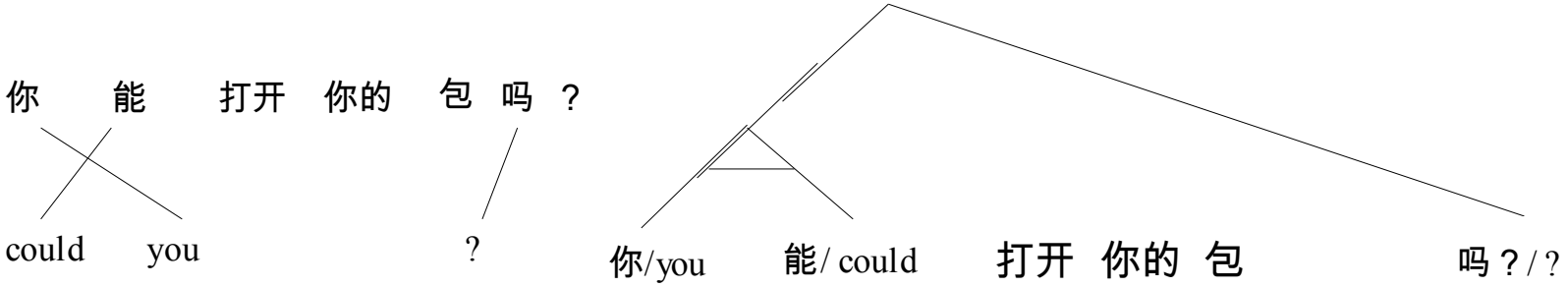
你 能 打开 你的 包 吗 ？

(a) Input



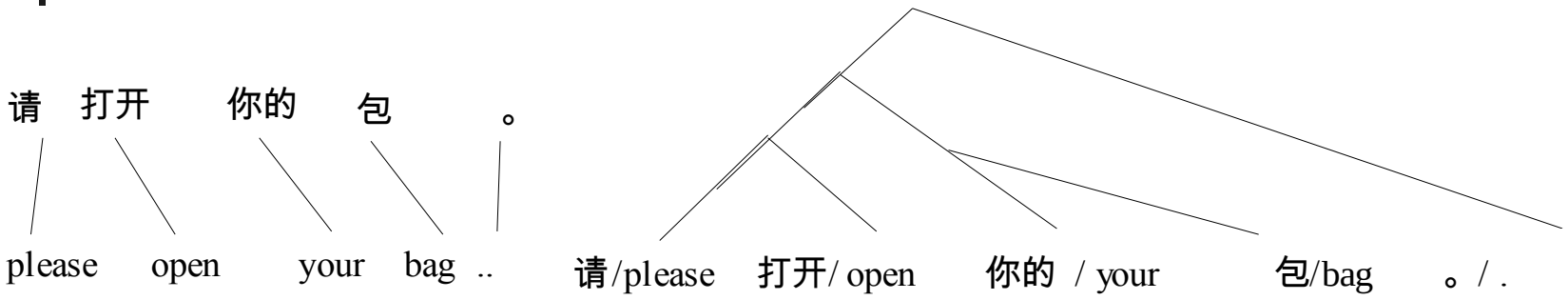
(b) Example A

Example-based Decoder - (4)

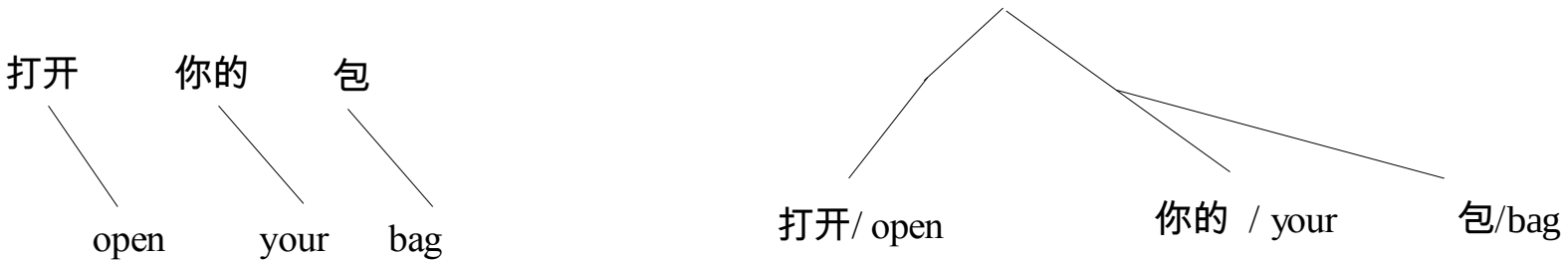


(c) Translation Tempate after match input with Example A

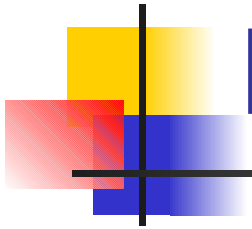
Example-based Decoder - (5)



(d) Example B



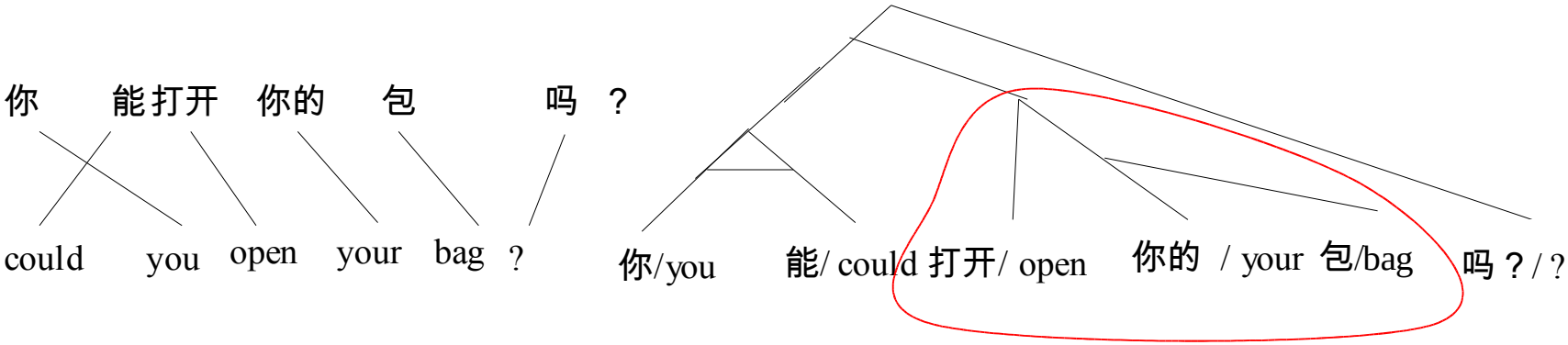
(e) Translation Tempate after match the child input with Example B



Example-based Decoder - (6)

- Merging
  - merging the child templates with the parent templates.
    - consider the child template as a sub-tree
    - merge it with the parent ITG tree.

Example-based Decoder - (7)



(f) Final translation after merged (c) and (e)

Example-based Decoder - (8)

- beam search
  - scoring function for each translation template is:

$$f(temp) = \log P(E_{trans} \mid C_{trans}) + \log H(C_{untrans})$$

Experiments-(1)

- We take part in the C-E task
  - Training corpus
    - 39,953 sentence pairs
  - development set
    - IWSLT07_devset3_*
    - 506 Chinese sentences and 16 references
  - test set
    - IWSLT07_devset2_*
    - 500 Chinese sentences and 16 references
  - official test set
    - 489 Chinese sentences

Experiments-(2)

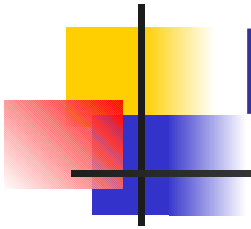
- Preprocessing
  - tokenization
  - lowercasing
  - Stemming by using a morphological dictionary

		Chinese	English (stemmed)
Train. Data	Sentences	39,963	
	Words	351,060	377,890
	Vocabulary	11,302	7,610
Dev. Set	Sentences	506	
	Words	3,826	
Test Set	Sentences	489	
	Words	3,189	

Experiments-(3)

- Baseline
  - CKY decoder
- Results for the stemmed corpus

Decoder	Bleu		
	Dev set	Test Set	Official Test set
CKY-Decoder	0.3614	0.3156	0.2741
EB-Decoder	0.3958	0.3334	0.3012



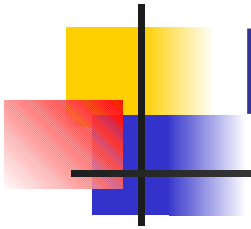
Experiments-(4)

- Post-Processing
  - Morphological changes
    - using a morphological dictionary and 3-gram LM
  - Case sensitive outputs: two simple rules
    - Uppercasing the first word of a sentence
    - Changing the word “i” to “I”

Experiments-(5)

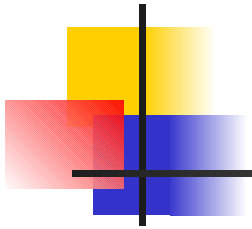
- Reasons for the result
  - the data becomes more sparse
  - the post-processing is too simple

Decoder	Bleu		
	Dev set	Test set	Official Test set
CKY-Decoder	0.1843	0.1769	0.1758
EB-Decoder	0.2024	0.1966	0.1934



Future Works

- Better examples retrieval algorithm.
- More feature functions in log-linear model.
- Improving morphological generation.



- Thank Dr. Boxing Chen for the presentation and the useful suggestions
- Thank you for your attention.
- Any Questions?
  - cwh2k@163.com

Main References

- Philipp Koehn, Franz Josef Och and Daniel Marcu: "Statistical Phrase-Based Translation". In NAACL/HLT 2003, pages 127-133(2003)
- David Chiang: "A Hierarchical Phrase-Based Model for Statistical Machine Translation". In Proc. of ACL 2005, pages 263–270 (2005)
- Franz Joseph Och and Hermann Ney: "Discriminative training and maximum entropy models for statistical machine translation". In Proceedings of the 40th Annual Meeting of the ACL, pp. 295–302(2002)
- Dekai Wu: "Stochastic Inversion Transduction Grammars and Bilingual Parsing of Parallel Corpora". Computational Linguistics, 23(3):374(1997)
- Franz Joseph Och and Hermann Ney: "A Systematic Comparison of Various Statistical Alignment Models". Computational Linguistics, 29(1):19–52, March(2003)
- Wen-Han Chao and Zhou-Jun Li: "Incorporating Constituent Structure Constraint into Discriminative Word Alignment", MT Summit XI, Copenhagen, Denmark, September 10-14, 2007, accepted. (2007)
- R. Moore. "A discriminative framework for bilingual word alignment". In Proceedings of HLT-EMNLP, pages 81–88, Vancouver, Canada, October. (2005)
- I. Dan Melamed. "Models of Translational Equivalence among Words". Computational Linguistics, 26(2): 221–249. (2000)
- Taro Watanabe and Eiichiro Sumita: "Example-based Decoding for Statistical Machine Translation". In Machine Translation Summit IX pp. 410-417 (2003)
- Kishore Papineni, Salim Roukos, Todd Ward and Wei-Jing Zhu: "BLEU: a Method for Automatic Evaluation of Machine Translation". In Proceedings of the 40th Annual Meeting of the Association fo Computational Linguistics(ACL), Philadelphia, July 2002, pp. 311-318(2002)
- A. Stolcke, "SRILM – An extensible language modeling toolkit," in Proceedings of the International Conference on Spoken Language Processing, Denver, Colorado, 2002, pp. 901–904