

**IWSLT – Honolulu, HW – October 2008**

# **Improvements in DP Beam Search for Phrase-based SMT**

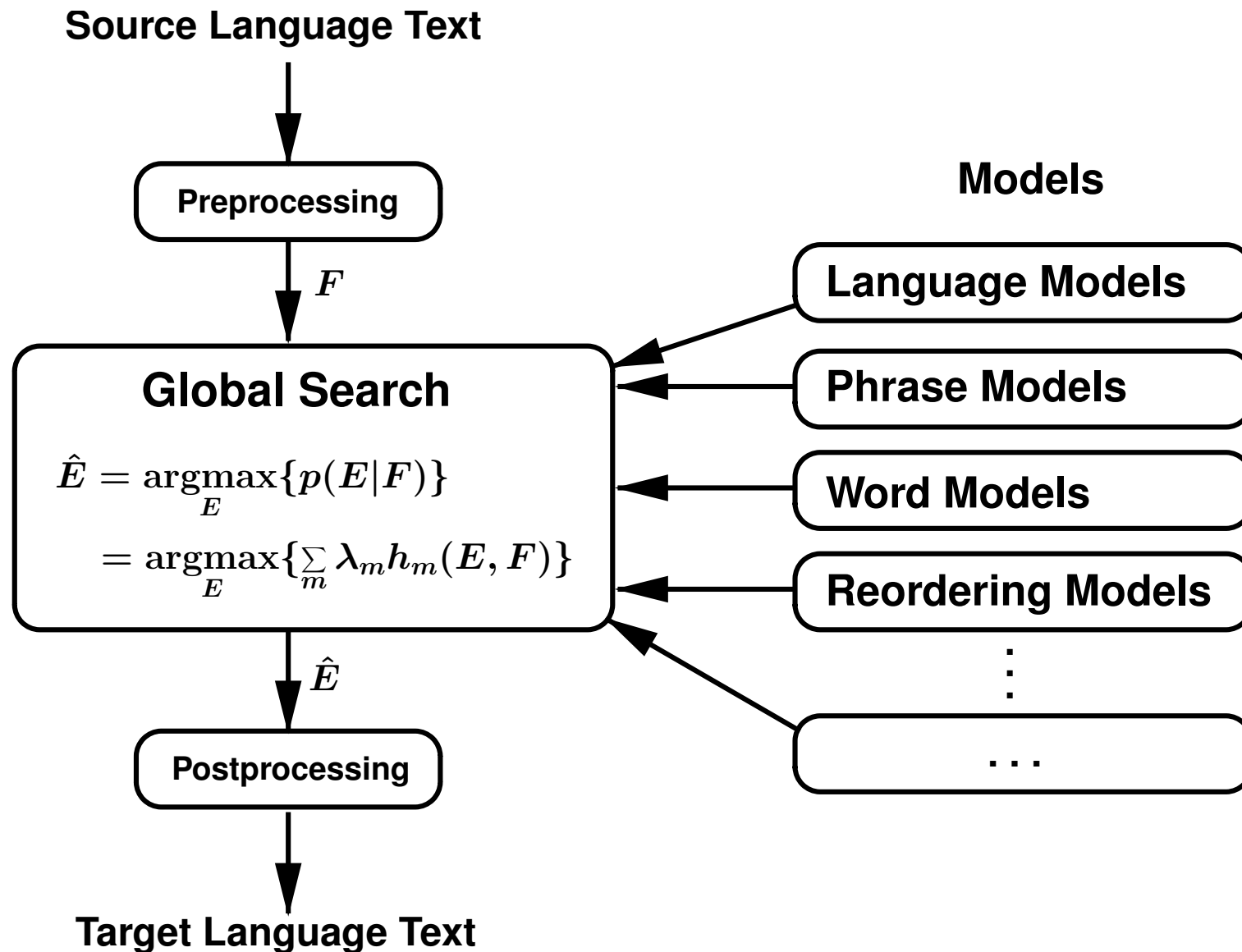
**Richard Zens and Hermann Ney**

**Human Language Technology and Pattern Recognition  
Lehrstuhl für Informatik 6  
Computer Science Department  
RWTH Aachen University, Germany**

- 1. Introduction & related work**
- 2. Search for phrase-based MT**
- 3. Experimental results**
- 4. Summary & conclusions**

- **clear & precise description of phrase-based search**
  - **analysis of important aspects**
    - **rest score estimation**
    - **lexical vs. coverage hypotheses**
    - **beam search including cube pruning**
- on a large data task**

- **based on**
  - [Zens & Och<sup>+</sup> 02]: phrase-based model
  - [Och 02]: rest score estimation (for AT)
  - [Tillmann & Ney 03]: search for SWB models
- **other related work:**
  - Pharaoh [Koehn 03], Moses [Koehn & Hoang<sup>+</sup> 07]
  - many others, e.g. [Tillmann 06], [Moore & Quirk 07], ...



			■
		■	
■			
	■		

## interdependencies:

- find phrase boundaries
- reordering in target language
- find most 'plausible' sentence

## constraints:

- no gaps
- no overlaps

- **goal:** 
$$\operatorname{argmax}_E \left\{ \max_S \sum_{m=1}^M \lambda_m h_m(E, S; F) \right\}$$

with target sentence  $E$ , segmentation  $S$ , source sentence  $F$ , models  $h(\cdot)$ , weights  $\lambda$

- **models:**

- **within phrase models:**

- phrase lexica, word lexica, word penalty, phrase penalty

- $n$ -gram backing-off language model

- distortion penalty

# Search Space

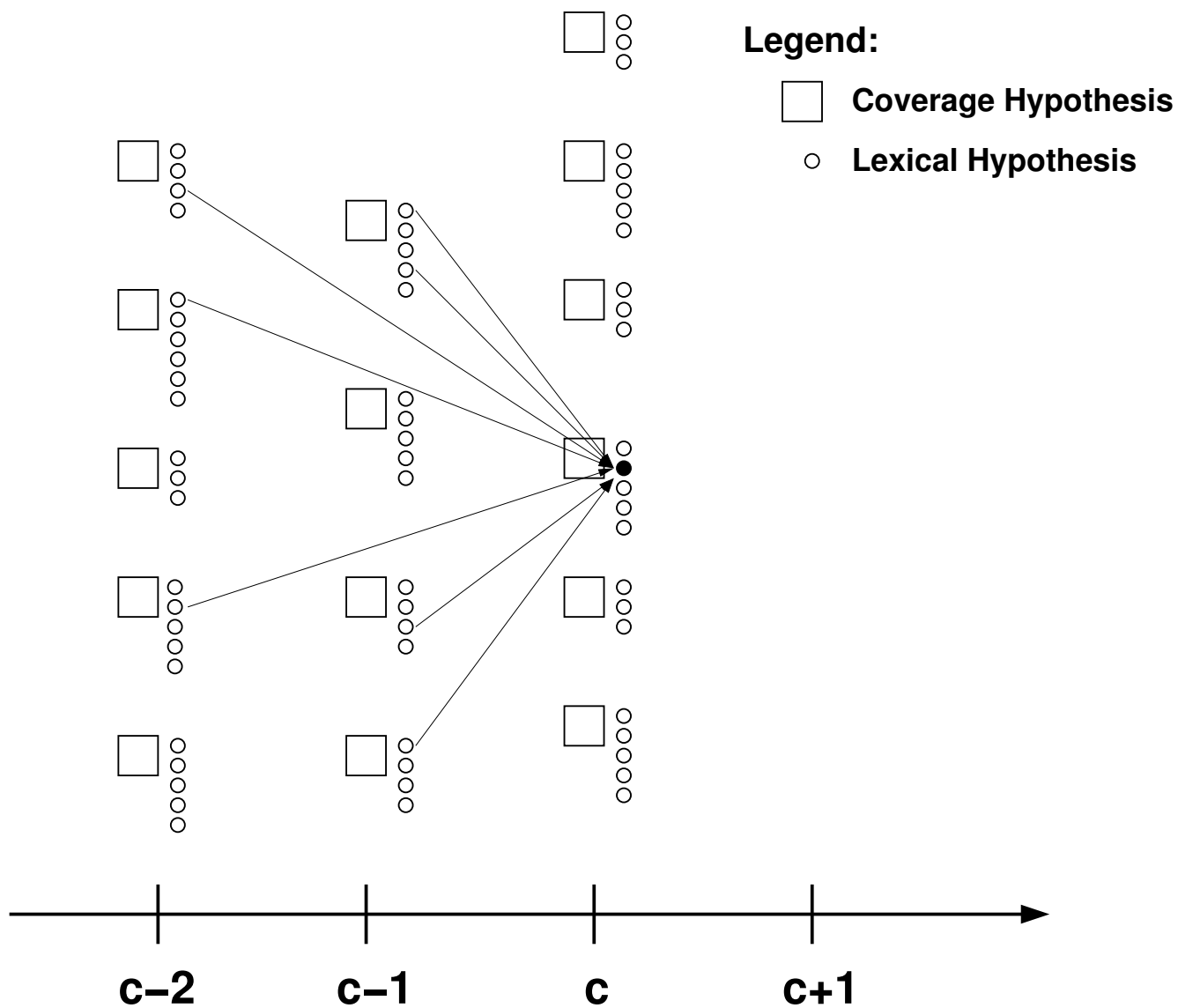
- **source sentence**  $F = f_1, \dots, f_J$
- **states**  $(C, \tilde{e}, j)$ 
  - **coverage**  $C \subseteq \{1, \dots, J\}$ : translated input positions
  - **LM history**  $\tilde{e}$  to predict the next target word
  - **source position**  $j$  for the distortion model
- **edges**  $(\tilde{e}, j, j')$ 
  - **generate target phrase**  $\tilde{e}$
  - **which covers the source sentence words**  $f_j, \dots, f_{j'}$
- **expanding**  $(C, \tilde{e}, j)$  with  $(\tilde{e}', j'', j')$  results in state

$$(C \cup \{j'', \dots, j'\}, \tilde{e} \oplus \tilde{e}', j')$$



- (partial) hypothesis: path to state  $(C, \tilde{e}, j)$
- for each cardinality  $c = |C|$ :  
we have a list of *coverage* hypotheses  $C$
- for each coverage  $C$ :  
we have a list of *lexical* hypotheses  $(\tilde{e}, j)$
- beam search: limit the list sizes

# Search Illustration

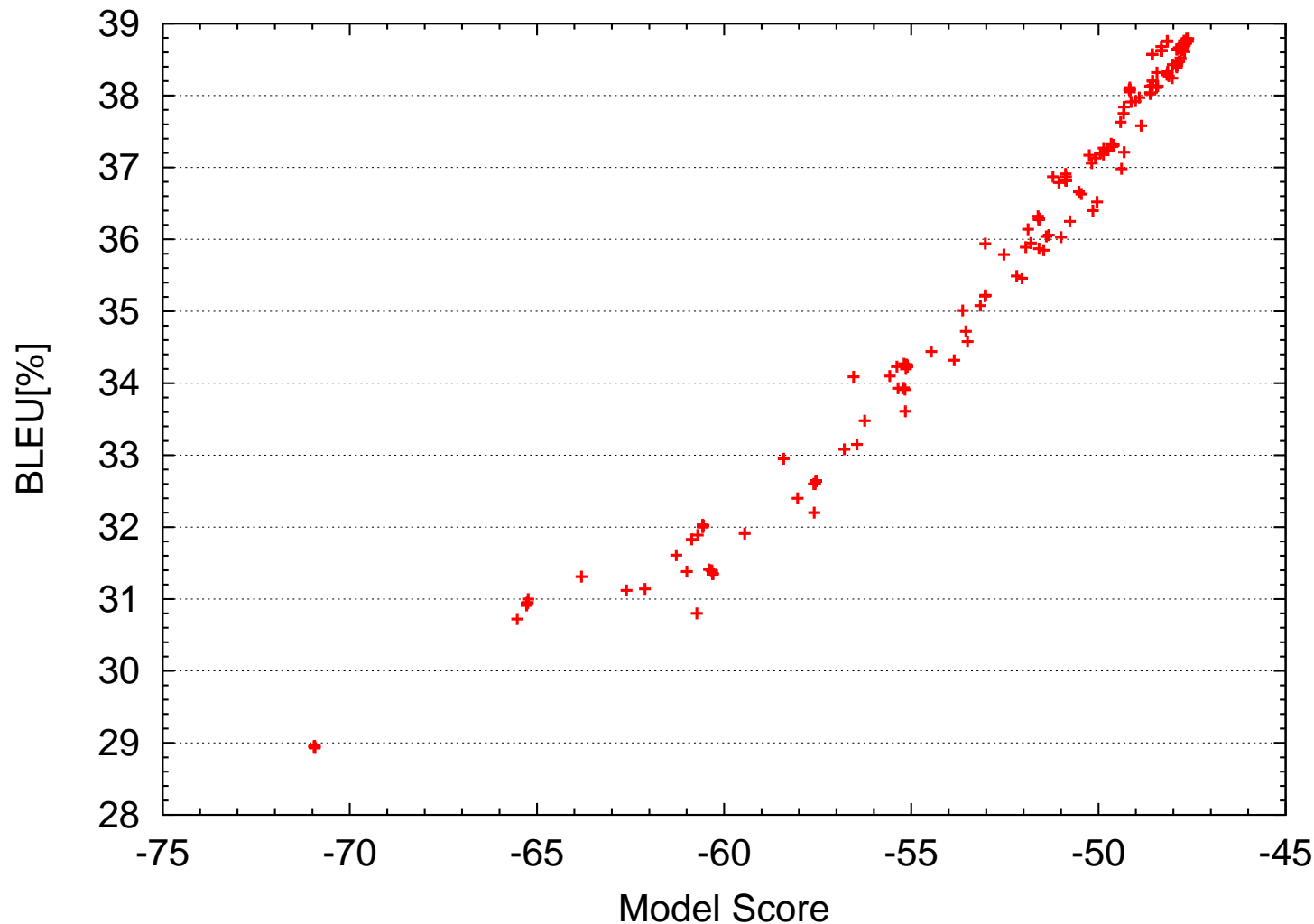


- **DP beam search**
  - generate hypotheses with increasing cardinality by expanding hypotheses with lower cardinality
  - recombine hypotheses with same state
  - expand only promising hypotheses
- share computations between expansions, e.g. check for overlap, rest score computation, ...
- early pruning  
stop expansion as soon as possible
- expand most promising candidates first

- **estimated score of hypothesis completion (inspired by  $A^*$ )**
- **previous work:**
  - [Och 02, Och & Ney 04]  
TM & LM per source position, distortion
  - [Koehn 03]  
TM & LM per source sequence, no distortion
- **here: comparison of**
  - computation per position and per sequence
  - models: TM only; TM & LM; TM, LM & distortion

- **NIST Chinese-English large data task**
- **TM:**  
training data: 8 M sentence pairs, 250 M words  
phrase-based, word-based lexica, word / phrase penalty
- **LM:**  
4-gram, trained on 650 M words, SRILM [Stolcke 02]
- **reordering:**  
distortion penalty, reordering window: 10  
lexicalized reordering model [Zens & Ney 06]
- **evaluation:**  
case-insensitive Bleu score (mt-eval) on NIST 2002 test set

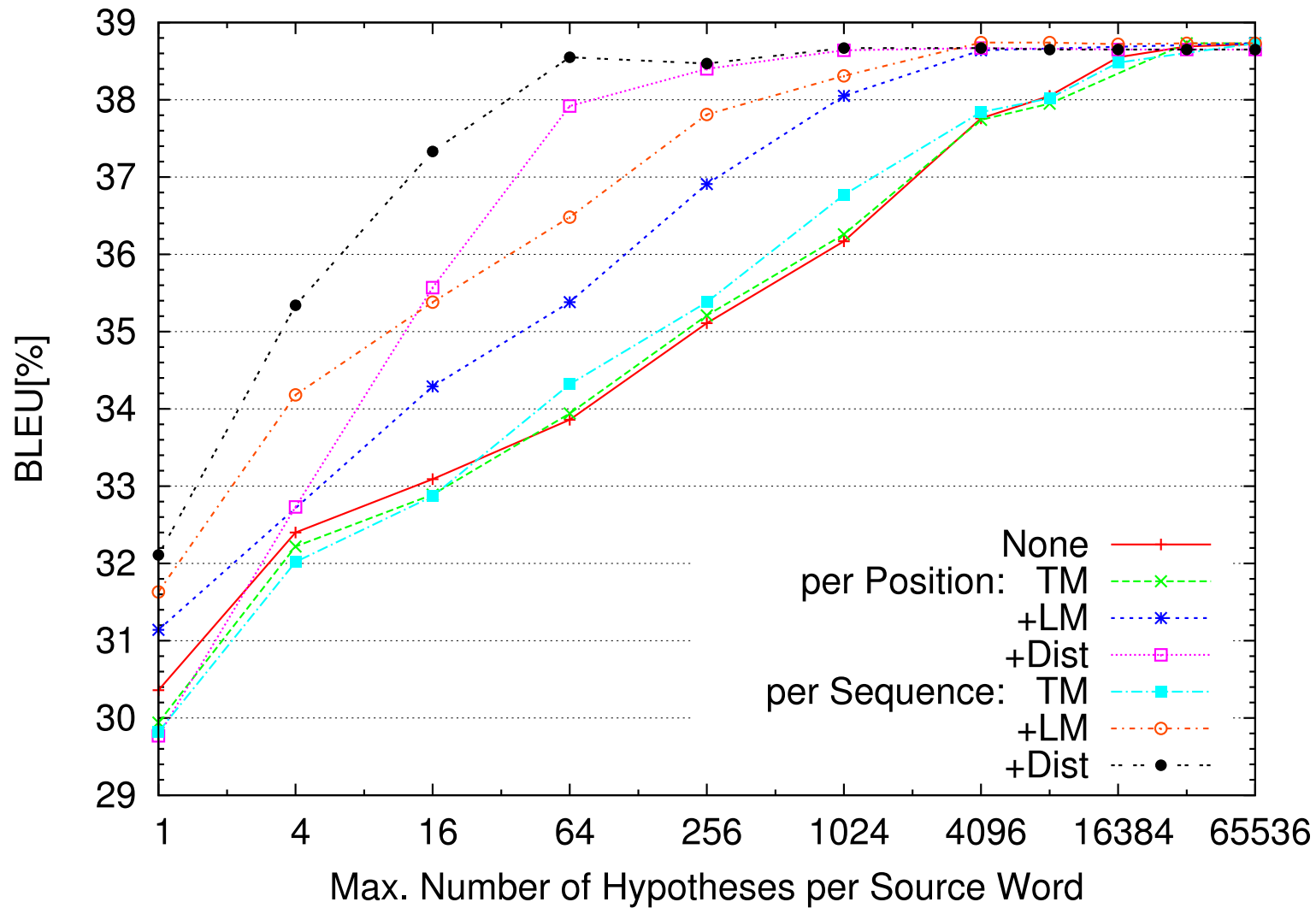
# Effect of Search Errors



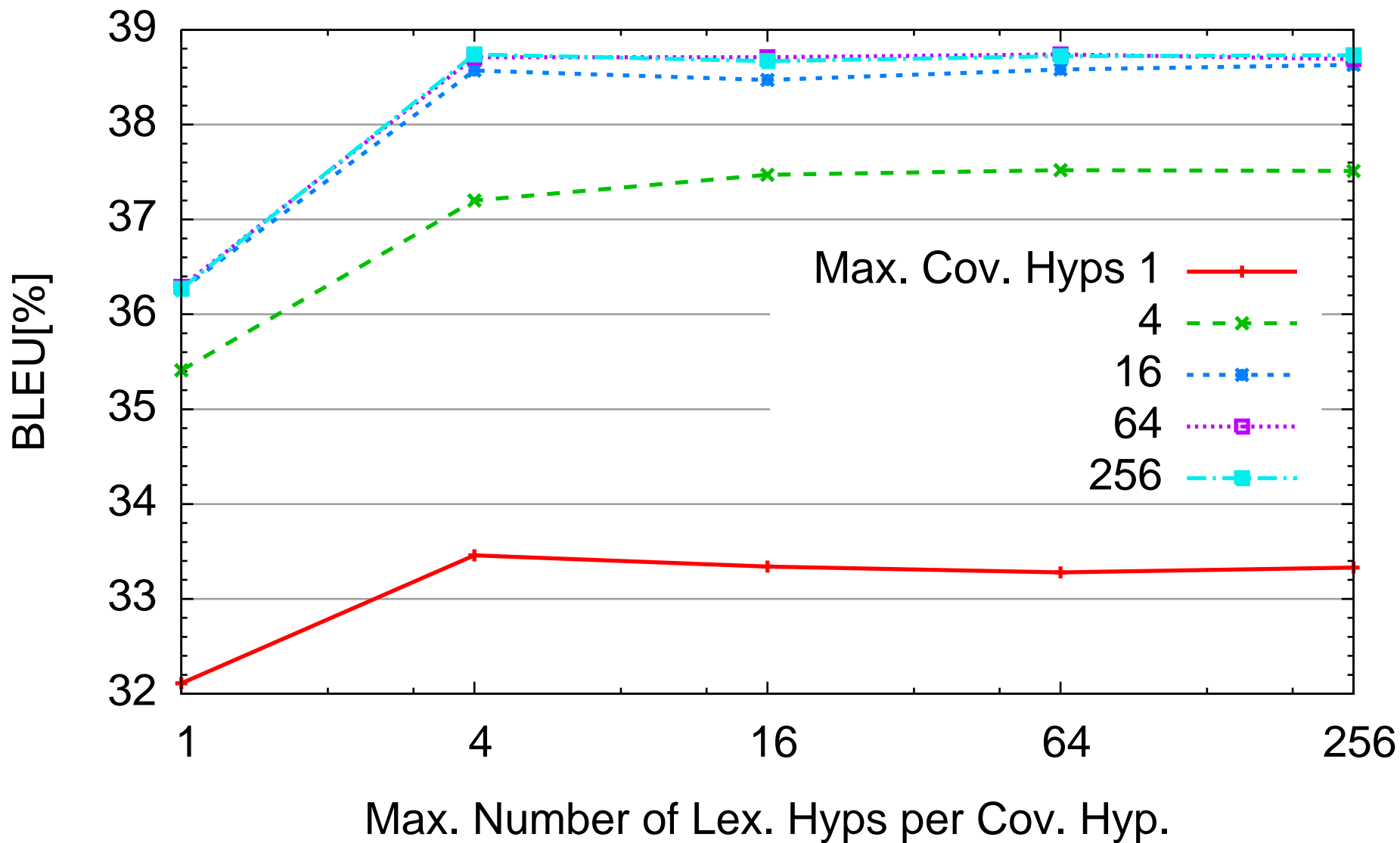
**Translate test set with various pruning parameters settings.**

**Model score averaged over whole test set (878 sentences).**

# Rest Score Estimation

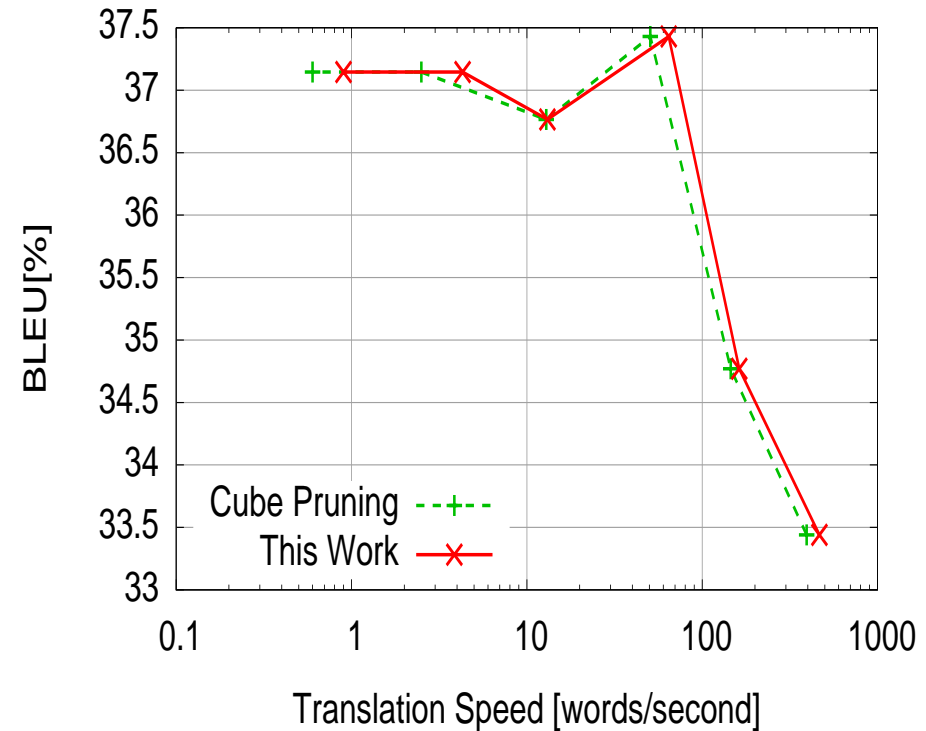
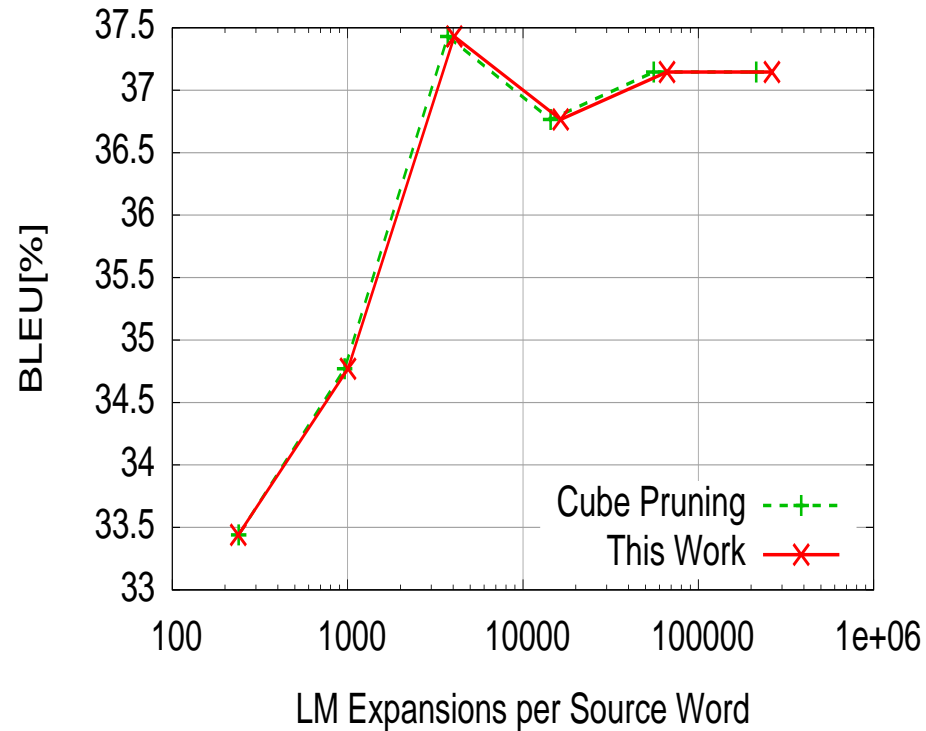


# Lexical vs. Coverage Hypotheses





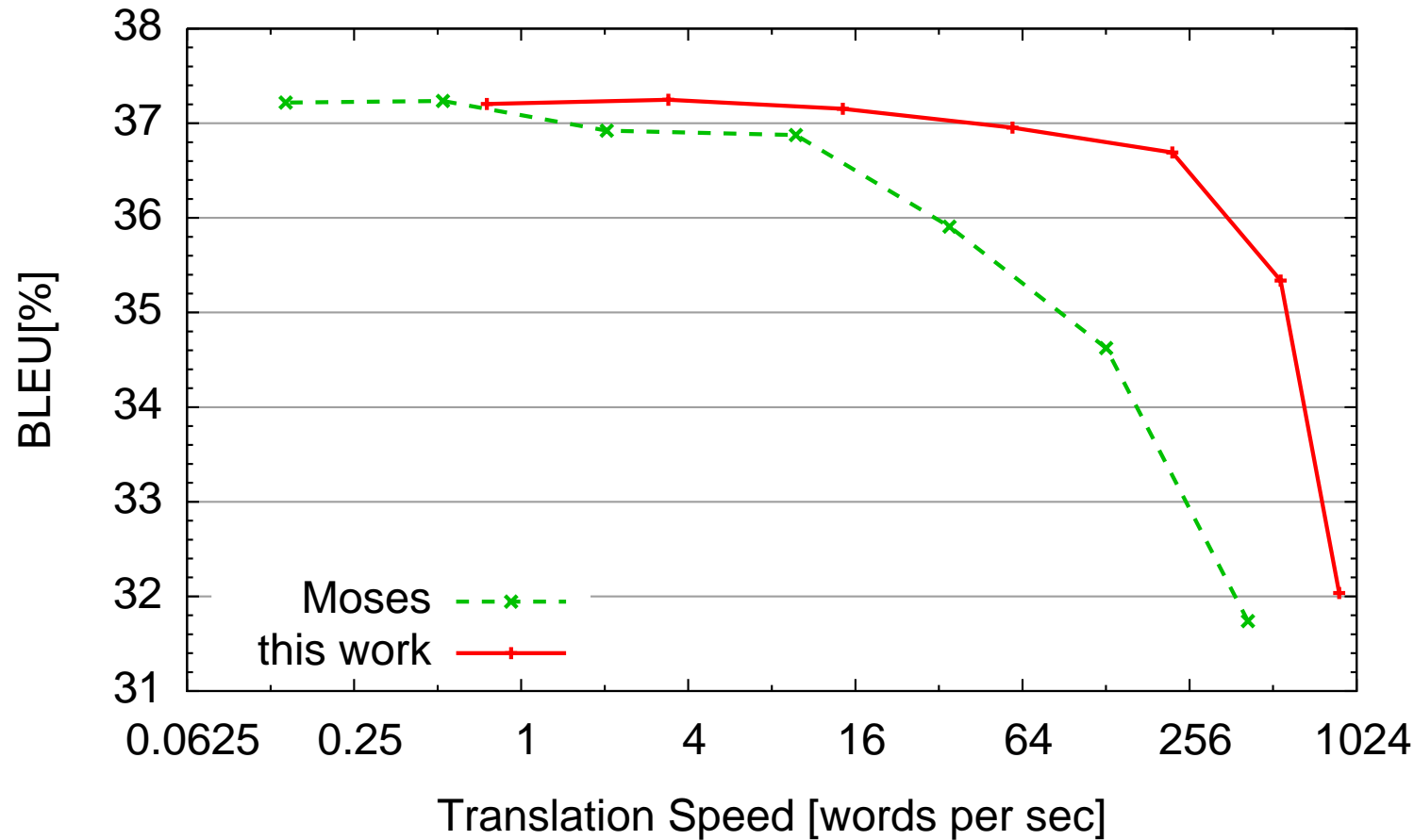
# Effect of Cube Pruning



**Numbers averaged over whole test set; vary beam sizes.**

**Lexicalized reordering not used, just distortion penalty.**

# Comparison with Moses



**Same TM, LM, etc.; vary beam setting**

**Lexicalized reordering not used, just distortion penalty.**

- **Summary**

- detailed problem description
- efficient solution
- in-depth analysis

- **Conclusions**

- search important for good translation quality
- rest score estimation allows for small beam sizes
- distinction lexical vs. coverage hypothesis important
- additional cube pruning not necessary
- significantly faster than Moses

**THANK YOU FOR YOUR ATTENTION!**

# References

- [Bellman 57] R. Bellman: *Dynamic Programming*. Princeton University Press, Princeton, NJ, 1957.
- [Chiang 05] D. Chiang: A Hierarchical Phrase-Based Model for Statistical Machine Translation. Proc. *43rd Annual Meeting of the Assoc. for Computational Linguistics (ACL)*, pp. 263–270, Ann Arbor, Michigan, June 2005.
- [Huang & Chiang 07] L. Huang, D. Chiang: Forest Rescoring: Faster Decoding with Integrated Language Models. Proc. *45th Annual Meeting of the Assoc. for Computational Linguistics (ACL)*, Prague, Czech Republic, June 2007.
- [Jelinek 98] F. Jelinek: *Statistical Methods for Speech Recognition*. MIT Press, Cambridge, MA, 1998.
- [Knight 99] K. Knight: Decoding Complexity in Word-Replacement Translation Models. *Computational Linguistics*, Vol. 25, No. 4, pp. 607–615, December 1999.

- [Koehn 03] P. Koehn: *Noun Phrase Translation*. Ph.D. thesis, University of Southern California, 2003.
- [Koehn & Hoang<sup>+</sup> 07] P. Koehn, H. Hoang, A. Birch, C. Callison-Burch, M. Federico, N. Bertoldi, B. Cowan, W. Shen, C. Moran, R. Zens, C. Dyer, O. Bojar, A. Constantine, E. Herbst: Moses: Open Source Toolkit for Statistical Machine Translation. Proc. *45th Annual Meeting of the Assoc. for Computational Linguistics (ACL): Poster Session*, pp. 177–180, Prague, Czech Republic, June 2007.
- [Koehn & Och<sup>+</sup> 03] P. Koehn, F.J. Och, D. Marcu: Statistical Phrase-Based Translation. Proc. *Human Language Technology Conf. / North American Chapter of the Assoc. for Computational Linguistics Annual Meeting (HLT-NAACL)*, pp. 127–133, Edmonton, Canada, May/June 2003.
- [Moore & Quirk 07] R.C. Moore, C. Quirk: Faster Beam-Search Decoding for Phrasal Statistical Machine Translation. Proc. *MT Summit XI*, Copenhagen, Denmark, September 2007.
- [Och 02] F.J. Och: *Statistical Machine Translation: From Single-Word Models to Alignment Templates*. Ph.D. thesis, Lehrstuhl für Informatik 6, Computer Science Department, RWTH Aachen University, Aachen, Germany, October 2002.

- [Och & Ney 04] F.J. Och, H. Ney: The Alignment Template Approach to Statistical Machine Translation. *Computational Linguistics*, Vol. 30, No. 4, pp. 417–449, December 2004.
- [Och & Tillmann<sup>+</sup> 99] F.J. Och, C. Tillmann, H. Ney: Improved Alignment Models for Statistical Machine Translation. Proc. *Joint SIGDAT Conf. on Empirical Methods in Natural Language Processing and Very Large Corpora (EMNLP)*, pp. 20–28, College Park, MD, June 1999.
- [Stolcke 02] A. Stolcke: SRILM – An Extensible Language Modeling Toolkit. Proc. *Int. Conf. on Spoken Language Processing (ICSLP)*, Vol. 2, pp. 901–904, Denver, CO, September 2002.
- [Tillmann 06] C. Tillmann: Efficient Dynamic Programming Search Algorithms for Phrase-Based SMT. Proc. *Workshop on Computationally Hard Problems and Joint Inference in Speech and Language Processing*, pp. 9–16, New York City, NY, June 2006.
- [Tillmann & Ney 03] C. Tillmann, H. Ney: Word Reordering and a Dynamic Programming Beam Search Algorithm for Statistical Machine Translation. *Computational Linguistics*, Vol. 29, No. 1, pp. 97–133, March 2003.

- [Zens 08] R. Zens: *Phrase-based Statistical Machine Translation: Models, Search, Training*. Ph.D. thesis, Lehrstuhl für Informatik 6, Computer Science Department, RWTH Aachen University, Aachen, Germany, February 2008.
- [Zens & Ney 06] R. Zens, H. Ney: Discriminative Reordering Models for Statistical Machine Translation. Proc. *Human Language Technology Conf. / North American Chapter of the Assoc. for Computational Linguistics Annual Meeting (HLT-NAACL): Workshop on Statistical Machine Translation*, pp. 55–63, New York City, NY, June 2006.
- [Zens & Och<sup>+</sup> 02] R. Zens, F.J. Och, H. Ney: Phrase-Based Statistical Machine Translation. Proc. M. Jarke, J. Koehler, G. Lakemeyer, editors, *25th German Conf. on Artificial Intelligence (KI2002)*, Vol. 2479 of *Lecture Notes in Artificial Intelligence (LNAI)*, pp. 18–32, Aachen, Germany, September 2002. Springer Verlag.