

















20. K. Papineni, S. Roukos, T. Ward, and W.-Jing Zhu. Bleu: a method for automatic evaluation of machine translation. In *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics*, pages 311–318, Philadelphia, PA, USA, 2002.
21. P. Koehn. Statistical significance tests for machine translation evaluation. In *Proceedings of EMNLP 2004*, Barcelona, Spain, July 2004.
22. M. Bisani and H. Ney. Bootstrap estimates for confidence intervals in asr performance evaluation. In *IEEE International Conference on Acoustics, Speech, and Signal Processing*, volume 1, pages 409–412, Montréal, Canada, May 2004.
23. A. Stolcke, Y. Konig, and M. Weintraub. Explicit word error ization in N-best list rescoring. In *Proc. Eurospeech '97*, pages 163–166, Rhodes, Greece, 1997.
24. J. Fiscus. A post-processing system to yield reduced word error rates: Recogniser output voting error reduction (rover). In *Proceedings 1997 IEEE Workshop on Automatic Speech Recognition and Understanding*, pages 347–352, Santa Barbara, CA, USA, 1997.
25. F. Huang and K. Papineni. Hierarchical system combination for machine translation. In *Proceedings of the 2007 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning (EMNLP-CoNLL)*, pages 277–286, Prague, Czech Republic, June 2007.