

Bibliography

Al-Onaizan, Yaser and Kevin Knight. 2002. Translating named entities using monolingual and bilingual resources. In *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics (ACL)*, pages 400-408, Philadelphia, July.

BCE. 2003. *Britannica Concise Encyclopedia*,
<http://wordpedia.britannica.com/concise/>.

BDC. 1992. *The BDC Chinese-English electronic dictionary* (version 2.0), Behavior Design Corporation, Taiwan.

Bikel, Daniel M., Richard Schwartz, and Ralph M. Weischedel. 1999. An algorithm that learns what's in a name. *Machine Learning*, 34(1/3).

Black, William J, Fabio Rinaldi, and David Mowatt. 1998. FACILE: Description of the NE System Used for MUC-7. In *Proceedings of the 7th Message Understanding Conference (MUC-7)*.

Black, William J. and Vasilakopoulos Argyrios. 2002. Language Independent Named Entity Classification by modified Transformation-based Learning and by Decision Tree Induction. In *Proceedings of the Sixth Conference on Natural Language Learning (CoNLL-2002)*, pages 159-162, Taipei, Taiwan.

Borthwick, Andrew. 1999. A maximum entropy approach to named entity recognition. PhD Dissertation, New York University.

Brown, P. F., Della Pietra S. A., Della Pietra V. J., and Mercer R. L. 1993. The mathematics of statistical machine translation: parameter estimation. *Computational Linguistics*, 19 (2): 263-311.

- Carreras, Xavier, Lluís Màrquez, and Lluís Padró. 2002. Named entity extraction using adaboost. In *Proceedings of the Sixth Conference on Natural Language Learning (CoNLL-2002)*, pages 167-170, Taipei, Taiwan.
- Chang, Jason S., David Yu, and Chun-Jen Lee. 2001. Statistical Translation Model for Phrases. *Computational Linguistics and Chinese Language Processing*, 6(2):43-64.
- Chao, Yuen Ren. 1968. *A Grammar of spoken Chinese*. Berkeley, University of California Press.
- Chen, Hsin-Hsi, Yung-Wei Ding, Shih-Chung Tsai and Guo-Wei Bian. 1998a. Description of the NTU system used for MET2. In *Proceedings of 7th Message Understanding Conference (MUC-7)*.
- Chen, Hsin-Hsi, Sheng-Jie Huang, Yung-Wei Ding, and Shih-Chung Tsai. 1998b. Proper name translation in cross-language information retrieval. In *Proceedings of 17th COLING and 36th ACL*, pages 232-236.
- Chen, Hsin-Hsi, Changhua Yang, and Ying Lin. 2003. Learning formulation and transformation rules for multilingual named entities. In *Proceedings of the ACL 2003 Workshop on Multilingual and Mixed-language Named Entity Recognition*, pages 1-8.
- Chen, Keh-Jiann and Shing-Huan Liu. 1992. Word identification for Mandarin Chinese sentences. In *Proceedings of COLING*, pages 101-107.
- Cheng, Pu-Jen, Jei-Wen Teng, Ruei-Cheng Chen, Jenq-Haur Wang, Wen-Hsiang Lu, and Lee-Feng Chien. 2004. Translating unknown queries with Web corpora for cross-language information retrieval. In *Proceedings of the 27th ACM International Conference on Research and Development in Information Retrieval (SIGIR)*.
- Chien, Lee-Feng, Chun-Liang Chen, Wen-Hsiang Lu, and Yuan-Lu Chang. 1999. Recent results on domain-specific term extraction from online Chinese text resources. In *Proceedings of ROCLING XII*, Hsinchu, Taiwan, pages 203-218.

- Chinchor, Nancy. 1997. MUC-7 Named entity task definition. In *Proceedings of the 7th Message Understanding Conference (MUC-7)*.
- Chua, Tat-Seng and Jimin Liu. 2002. Learning pattern rules for Chinese named-entity extraction. In *Proceedings of the Eighteenth National Conference on Artificial Intelligence (AAAI)*, pages 411-418, Edmonton, Canada.
- Chuang, Thomas C., Geeng Neng You, and Jason S. Chang. 2002. Adaptive bilingual sentence alignment. *Lecture Notes in Artificial Intelligence*, 2499:21-30.
- Cibelli, Jose B., Robert P. Lanza, Michael D. West, and Carol Ezzell. 2002. What Clones? In *Scientific American*, January. (<http://www.sciam.com>)
- CNA. 2003. *Central News Agency*, <http://client.cna.com.tw>.
- Collins, Michael. 2002. Ranking Algorithms for Named-Entity Extraction: Boosting and the Voted Perceptron. In *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics (ACL)*, pages 489-496.
- Dagan, Ido, Kenneth W. Church, and William A. Gale. 1993. Robust bilingual word alignment for machine aided translation. In *Proceedings of the Workshop on Very Large Corpora: Academic and Industrial Perspectives*, pages 1-8, Columbus Ohio.
- Damerau, F. 1964. A technique for computer detection and correction of spelling errors. *Comm. of the ACM*, 7(3): 171-176.
- Dempster, A. P., N. M. Laird, and D. B. Rubin. 1977. Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society*, 39(1):1-38.
- Feng, Donghui, Yajuan Lv, and Ming Zhou. 2004. A new approach for English-Chinese named entity alignment. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP 2004)*, pages 372-379.
- Forney, G. D. 1973. The Viterbi algorithm. *Proceedings of IEEE*, 61:268-278, March.

- Fukumoto, J., M. Shimohata, F. Masui and M. Sasaki. 1998. Oki Electric Industry: Description of the Oki System as Used for MET-2. In *Proceedings of the 7th Message Understanding Conference (MUC-7)*.
- Fung, Pascale and Lo Yuen Yee. 1998. An IR approach for translating new words from nonparallel, comparable texts. In *Proceedings of the 36th Annual Conference of the Association for Computational Linguistics (ACL)*, pages 414-420.
- Hall, Patrick A.V. and Geoff R. Dowling. 1980. Approximate String Matching. *ACM Computing Surveys*, 12:381- 402.
- Huai, Lu. 1989. Handbook of English Name Knowledge, ISBN 7-5012-0144-7/Z.10, 1st edition.
- Huang, Fei, Stephan Vogel, and A. Waibel. 2003. Automatic extraction of named entity translingual equivalence based on multi-feature cost minimization. In *Proceedings of ACL Workshop on Multilingual and Mixed-language NER*, Sapporo, Japan.
- Humphreys K., Gaizauskas R., *et al.* 1998. University of Sheffield: description of the LaSIE-II system as used for MUC-7. In *Proceedings of 7th Message Understanding Conference (MUC-7)*.
- Isozaki, Hideki and Hideto Kazawa. 2002. Efficient Support Vector Classifiers for Named Entity Recognition. In *Proceedings of the 19th International Conference on Computational Linguistics (COLING)*, pages 390-396, Taipei, Taiwan.
- Kang, Byung-Ju and Key-Sun Choi. 2001. Two approaches for the resolution of word mismatch problem caused by English words and foreign words in Korean information retrieval. *International Journal of Computer Processing of Oriental Languages*, 14 (2):109-131.
- Kilgarrieff, Adam and Gregory Grefenstette. 2003. Introduction to the special issue on the Web as corpus. *Computational Linguistics*, 29(3): 333-347.
- Kim, Ji-Hwan and P. C. Woodland. 2000. A rule-based named entity recognition system for speech input. In *Proceedings of the Sixth International Conference on Spoken Language Processing*, Beijing, China.

- Knight, Kevin and Jonathan Graehl. 1998. Machine transliteration. *Computational Linguistics*, 24(4):599-612.
- Kraaij, Wessel, Jian-Yun Nie, and Michel Simard. 2003. Embedding Web-based statistical translation models in cross-language information retrieval. *Computational Linguistics*, 29(3): 381-419.
- Krupka, George R. and Kevin Hausman. 1998. IsoQuest, Inc.: Description of the NetOwlTM Extractor System as Used for MUC-7. In *Proceedings of 7th Message Understanding Conference (MUC-7)*.
- Kumano, Tadashi, Hideki, Kashioka, Hideki Tanaka and Takahiro Fukusima. 2004. Acquiring Bilingual Named Entity Translations from Content-aligned Corpora. In *Proceedings of the First International Joint Conference on Natural Language Processing (IJCNLP-04)*, Hainan Island, China.
- Kupiec, Julian. 1993. An algorithm for finding noun phrase correspondences in bilingual corpora. In *Proceedings of the 40th Annual Conference of the Association for Computational Linguistics (ACL)*, pages 17-22, Columbus, Ohio.
- Larkey, Leah S., Paul Ogilvie, and M. Andrew Price. 2000. Acrophile: an automated acronym extractor and server. In *Proceedings of Fifth ACM Conference on Digital Libraries*.
- Lee, Chun-Jen and Jason S. Chang. 2003. Acquisition of English-Chinese transliterated word pairs from parallel-aligned texts using a statistical machine transliteration model. In *Proceedings of HLT-NAACL 2003 Workshop on Building and Using Parallel Texts: Data Driven Machine Translation and Beyond*, pages 96-103, Edmonton, Canada, May.
- Lee, Chun-Jen, Jason S. Chang and Jyh-Shing Roger Jang. 2003. A statistical approach to Chinese-to-English Back-transliteration. In *Proceedings of the 17th Pacific Asia Conference on Language, Information, and Computation (PACLIC)*, pages 310-318, Singapore, Dec.
- Lee, Chun-Jen, Jason S. Chang and Jyh-Shing Roger Jang. 2004a. Bilingual named-entity pairs extraction from parallel corpora. In *Proceedings of IJCNLP-04*

Workshop on Named Entity Recognition for Natural Language Processing Applications, pages 9-16, Hainan Island, China, Jan.

Lee, Chun-Jen, Jason S. Chang and Thomas C. Chuang. 2004b. Alignment of bilingual named entities in parallel corpora using statistical model. *Lecture Notes in Artificial Intelligence*, 3265:144-153.

Lee, Chun-Jen, Jason S. Chang, and Jyh-Shing Roger Jang. 2006a. Extraction of transliteration pairs from parallel corpora using a statistical transliteration model. *Information Sciences*, 176: 67-90.

Lee, Chun-Jen, Jason S. Chang, and Jyh-Shing Roger Jang. 2006b. Alignment of bilingual named entities in parallel corpora using statistical models and multiple knowledge sources. Accepted and to appear in *ACM transactions on Asian Language Information Processing*.

Lee, Jae Sung and Key-Sun Choi. 1997. A statistical method to generate various foreign word transliterations in multilingual information retrieval system. In *Proceedings of the 2nd International Workshop on Information Retrieval with Asian Languages (IRAL'97)*, pages 123-128, Tsukuba, Japan.

Lin, Wei-Hao and Hsin-Hsi Chen. 2002. Backward transliteration by learning phonetic similarity. In *CoNLL-2002, Sixth Conference on Natural Language Learning*, Taipei, Taiwan.

Lu, Wen-Hsiang, Lee-Feng Chien, and Hsi-Jian Lee. 2004. Anchor text mining for translation of Web queries: a transitive translation approach. *ACM transactions on Information Systems*, 22(2): 242-269.

Macmillan. 2002. Macmillan English Dictionary, ISBN 0-333-96671-6, 1st edition. Macmillan Publishers Limited.

Manning, Christopher D. and Hinrich Schutze. 1999. *Foundations of Statistical Natural Language Processing*, MIT Press; 1st edition.

McNamee, Paul and James Mayfield. 2002. Entity extraction without language specific resources. In *Proceedings of the Sixth Conference on Natural Language Learning (CoNLL-2002)*, pages 183–186, Taipei, Taiwan.

- Melamed, I Dan. 1996. Automatic construction of clean broad coverage translation lexicons. In *Proceedings of the 2nd Conference of the Association for Machine Translation in the Americas (AMTA'96)*, Montreal, Canada.
- Melamed, I. Dan. 1997. A Word-to-Word Model of Translational Equivalence. In *Proceedings of the 35th Annual Conference of the Association for Computational Linguistics (ACL)*, pages 490-497.
- Mikheev, Andrei, Calire Grover, and Marc Moens. 1998. Description of the LTG system used for MUC-7. In *Proceedings of the 7th Message Understanding Conference (MUC-7)*.
- Miller, Scott, Michael Crystal, Heidi Fox, Lance Ramshaw, Richard Schwartz, Rebecca Stone, Ralph Weischedel, and the Annotation Group. 1998. Algorithms that learn to extract information - BBN: description of the SIFT system as used for MUC-7. In *Proceedings of the 7th Message Understanding Conference (MUC-7)*.
- Moore, Robert C. 2003. Learning Translations of Named-Entity Phrases from Parallel Corpora. In *Proceedings of the 10th Conference of the European Chapter of the Association for Computational Linguistics*, pages 259-266, Budapest, Hungary.
- Nie, Jian-Yun, Michel Simard, Pierre Isabelle, and Richard Durand. 1999. Cross-language information retrieval based on parallel texts and automatic mining of parallel texts from the Web. In *Proceedings of the 22nd ACM International Conference on Research and Development in Information Retrieval (SIGIR)*.
- Och, Franz Josef and Hermann Ney. 2003. A systematic comparison of various statistical alignment models. *Computational Linguistics*, 29(1): 19-51.
- Oh, Jong-Hoon and Key-Sun Choi. 2002. An English-Korean transliteration model using pronunciation and contextual rules. In *Proceedings of the 19th International Conference on Computational Linguistics (COLING)*, Taipei, Taiwan.
- Proctor, P., 1988. *Longman English-Chinese Dictionary of Contemporary English*, Longman Group (Far East) Ltd., Hong Kong.
- Rapp, Reinhard. 1999. Automatic identification of word translations from unrelated

- English and German corpora. In *Proceedings of the 37th Annual Conference of the Association for Computational Linguistics (ACL)*, pages 519-525.
- Resnik, Philip and Noah A. Smith. 2003. The Web as a parallel corpus. *Computational Linguistics*, 29(3): 349-380.
- Sang, Erik F. Tjong Kim and Fien De Meulder. 2003. Introduction to the CoNLL-2003 shared task: language-independent named entity recognition. In *Proceedings of the Seventh Conference on Natural Language Learning (CoNLL-2003)*, Edmonton, Canada, pages 142-147.
- Schwartz, Ariel S. and Marti A. Hearst. 2003. A simple algorithm for identifying Abbreviation definitions in biomedical text. In *Proceedings of the Pacific Symposium Biocomputing (PSB)*.
- Sekine, Satoshi, Ralph Grishman, and Hiroyuki Shinnou. 1998. A decision tree method for finding and classifying names in Japanese texts. In *Proceedings of the Sixth Workshop on Very Large Corpora, Montreal, Canada*.
- Sinorama. 2002. *Sinorama Magazine*, <http://www.greatman.com.tw/sinorama.htm>.
- Smadja, Frank Z., Kathleen McKeown, and Vasileios Hatzivassiloglou. 1996. Translating collocations for bilingual lexicons: a statistical approach. *Computational Linguistics*, 22(1):1-38.
- Solorio, Thamar and Aurelio López López. 2004. Learning named entity classifiers using support vector machines. In *Proceedings of the Fifth International Conference on Intelligent Text Processing and Computational Linguistics (CICLing-2004)*, pages 158-167, Seoul, Korea.
- Stalls, Bonnie Glover and Kevin Knight. 1998. Translating names and technical terms in Arabic text. In *Proceedings of the COLING/ACL Workshop on Computational Approaches to Semitic Languages*.
- Sun, Jian, Ming Zhou, and Jianfeng Gao. 2003. A class-based language model approach to Chinese named entity identification. *Computational Linguistics and Chinese Language Processing*, 8(2):1-28.
- Taghva, Kazem and Jeff Gilbreth. 1999. Recognizing acronyms and their definitions.

International Journal on Document Analysis and Recognition, 1:191-198.

Takeuchi, Koichi and Nigel Collier. 2002. Use of Support Vector Machines in Extended Named Entity Recognition. In *Proceedings of the Sixth Conference on Natural Language Learning (CoNLL-2002)*, Taipei, Taiwan.

Tsuji, Keita. 2002. Automatic extraction of translational Japanese-KATAKANA and English word pairs from bilingual corpora. *International Journal of Computer Processing of Oriental Languages*, 15(3):261-279.

Tsukamoto, Koji, Yutaka Mitsuishi, and Manabu Sassano. 2002. Learning with multiple stacking for named entity recognition. In *Proceedings of the Sixth Conference on Natural Language Learning (CoNLL-2002)*, pages 191-194, Taipei, Taiwan.

Wan, Stephen and Cornelia Maria Verspoor. 1998. Automatic English-Chinese name transliteration for development of multilingual resources. In *Proceedings of 17th COLING and 36th ACL*, pages 1352-1356.

Wells, J. C. 2001. *Longman Pronunciation Dictionary (New Edition)*, Addison Wesley Longman, Inc.

Wu, Chien-Cheng and Jason S. Chang. 2004. Bilingual Collocation Extraction Based on Syntactic and Statistical Analyses. *Computational Linguistics and Chinese Language Processing*, 9(1): 1-20.

Wu, Dekai, Grace Ngai, Marine Carpuat, Jeppe Larsen, and Yongsheng Yang. 2002. Boosting for named entity recognition. In *Proceedings of the Sixth Conference on Natural Language Learning (CoNLL-2002)*, pages 195-198, Taipei, Taiwan.

Wu, Dekai and Xuanyin Xia. 1994. Learning an English-Chinese lexicon from a parallel corpus. In *Proceedings of the First Conference of the Association for Machine Translation in the Americas*, pages 206-213.

Yang, Christopher C. and Kar Wing Li. 2003. Automatic construction of English/Chinese parallel corpora. *Journal of the American Society for Information Science and Technology*, 54(8): 730-742.

- Yeates, Stuart. 1999. Automatic extraction of acronyms from text. In *Proceedings of the Third New Zealand Computer Science Research Students' Conference*, pages 117-124.
- Yeates, Stuart, David Bainbridge, and Ian H. Witten. 2000. Using compression to identify acronyms in text. In *Proceedings of the Data Compression Conference*.
- Zhang, Ying and Phil Vines. 2004. Using the Web for automated translation extraction in cross-language information retrieval. In *Proceedings of the 27th ACM International Conference on Research and Development in Information Retrieval (SIGIR)*.
- Zhou, GuoDong and Jian Su. 2002. Named Entity Recognition using an HMM-based Chunk Tagger. In *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics (ACL)*, pages 473-480, Philadelphia.

