

References

Antony P.J. (2013). *Machine Translation Approaches and Survey for Indian Languages. Computational Linguistics and Chinese Language Processing*. Vol. 18, No. 1, March 2013, pp. 47-78.

Antony P.J. and Soman, K.P. (2011). *Parts Of Speech Tagging for Indian Languages: A Literature Survey*. International Journal of Computer Applications (0975 – 8887).Volume 34– No.8, November 2011.

Aires, J., Lopes, G. P., and Silva, J. F. (2008). Efficient multi-word expressions extractor using suffix arrays and related structures. In Proceeding of the 2nd ACM Workshop on Improving Non English Web Searching (Napa Valley, California, USA, October 30 - 30, 2008). iNEWS '08. ACM, New York, NY, 1-8.

Barman, A. K., Sarmah, J., Sarma, S. K. (2013). *Automated Identification of Assameses and Bodo Multiword Expressions*. Proceeding of 2013 International conference on Advances in computing, communication and Informatics (ICACCI).

Bharathi, A., Prashanth R. Mannem (2007), “Introduction to the Shallow Parsing Contest for South Asian Languages”, Language Technologies Research Center, International Institute of Information Technology, Hyderabad, India 500032.

Baldwin, T. and F. Bond (2002). Multiword expressions: Some problems for Japanese NLP. In Proceedings of the 8th Annual Meeting of the Association for Natural Language Processing (Japan), Keihanna, Japan, pp. 379–382.

Baldwin, T., J. Beavers, L. Van Der Beek, F. Bond, D. Flickinger, and I. A. Sag (2006). In search of a systematic treatment of determinerless PPs. In P. Saint-Dizier (Ed.), *Syntax and Semantics of Prepositions*.

Baldwin, Timothy, Colin Bannard, Takaaki Tanaka, and Dominic Widdows. (2003a). An empirical model of multiword expression decomposability. In *Proceedings of the*

ACL2003 Workshop on Multiword Expressions: analysis, acquisition and treatment, pp. 89–96, Sapporo, Japan.

Baldwin, Timothy, Kim S. N. (2010) Multiword expressions. In: Indurkha N, Damerau FJ (eds) *Handbook on Natural Language Processing*, 2nd edition, CRC Press, Boca Raton, USA. pp-267-292.

Bannard, C., Baldwin, T., & Lascarides, A. (2003). *A statistical approach to the semantics of verb-particles*. In proceedings of the ACL 2003 workshop on Multiword expressions, (pp. 65–72), Morristown, NJ, USA. Association for Computational Linguistics.

Bauer, L. (1983). *English Word-formation*. Cambridge, UK: Cambridge University Press.

Benson, M. (1990). Collocations and general-purpose dictionaries. *International Journal of Lexicography* 3 (1), 23–35.

Bharati A., Chaitanya V. and Sangal R. (1995). *Natural Language Processing: A Paninian Perspective*. Prentice Hall India.

Biber D, Johansson S, Leech G, Conrad S, Finegan E (1999) *Longman grammar of spoken and written English, 1st edn*. Pearson Education, Harlow, 1204p.

Bond, F. (2005). *Translating the Untranslatable: A solution to the Problem of Generating English Determiners*. CSLI Studies in Computational Linguistics. CSLI Publications.

Bourigault, D., 1992. Surface Grammatical Analysis for the Extraction of Terminological Noun Phrases. Proceedings of 14th International Conference on Computational Linguistics, Nantes, France, pp. 977-981.

Brants T. (2000). TnT – A statistical part-of-speech tagger. In Proceedings of the 6th Applied NLP Conference. 224-231.

Brill, E. and Resnik, P. (1994). *A rule-based approach to prepositional phrase attachment disambiguation*. In COLING-94, Kyoto, pp. 1198-1204.

C. Ramisch, *Multiword Expressions Acquisition*, Theory and Applications of Natural Language Processing, Springer International Publishing Switzerland 2015, DOI 10.1007/978-3-319-09207-2,

Chafe, Wallace L. 1968. Idiomaticity as an anomaly in the Chomskyan paradigm. *Foundations of Language* . pp. 109–127.

Chakraborty, T.(2010). *Multiword Expressions*, Jadavpur University, Kolkata

Chander, I. (1998). Automated postediting of documents. Ph. D. thesis, University of Southern California.

Chelliah S.L. (1997). A Grammar of Meithei. Mouton de Gruyter, Berlin.

Choueka, Y. (1988). Lookin for needles in a haystack or locating interesting collocational expressions in large textual databases. In Proceedings of RIAO, pp. 43–38.

Church, K. W. & Hanks, P. (1990). *Word association norms, mutual information, and Lexicography*. *Computational Linguistics*, 16(1), 22–29.

Church, K.W., & Hanks, P. (1989). Word association norms, mutual information and lexicography, Proceedings of the 27th Annual Meeting of the Association for Computational Linguistics, University of British Columbia, Vancouver, Canada, 26-29 June 1989, pp. 76-83.

Church, Kenneth W., and Patrick Hanks. (1989). Word association norms, mutual information and lexicography. In *Proceedings of the 27th Annual Meeting of the Association of Computational Linguistics (ACL-1989)*, 76–83, Vancouver, Canada.

Cruse, Alan D. (1986). *Lexical Semantics*. Cambridge, UK: Cambridge University Press.

Cutting D., Kupiec J, Pederson J. and Sibun P. (1992). A practical partof- speech tagger. In Proceedings of the 3rd Conference on Applied NLP. 133-140.

Dagan, Ido, Church, Kenneth W. and Gale, William A. (1993). Robust Bilingual Word Alignment for Machine-Aided Translation. In Proceedings, Workshop on Very

Large Corpora: Academic and Industrial Perspectives, Columbus, Ohio, 1-8. Association for Computational Linguistics.

Dash, N., S., Prayukt, B., B., O. (2013) An International Journal on Linguistics and Language Technology, Vol. 1, No. 1, Pp. 53-96

Daille, B., (1995) Study and Implementation of Combined Techniques for Automatic Extraction of Terminology. The Balancing Act Combining Symbolic and Statistical Approaches to Language, MIT Press.

Dandapat, S., & Sarkar, S. (2006). *Part-of-Speech Tagging for Bengali with Hidden Markov Model*. NLP/ML workshop on Part of speech tagging and Chunking for Indian language.

De Rose S. J. (1988). Grammatical category disambiguation by statistical optimization. Computational Linguistics, 14:31-39.

Debasri Chakrabarti (2011), "Layered Parts of Speech Tagging for Bangla", Language in India www.languageinindia.com, May 2011 Special Volume: Problems of Parsing in Indian Languages.

Dermatas E. and George K. (1995). Automatic stochastic tagging of natural language texts. Computational Linguistics, 21(2): 137-163.

Dias, G., Gylloré, S., Lopes, G. (1999). Language Independent Automatic Acquisition of Rigid Multiword Units from Unrestricted Text corpora. In Proceedings of the TALN'99, p. 333-338. Corse, July 12-17.

Dias, G., S. Guillor, J-C. Bassano, and J.G. Pereira Lopes. (2000). Combining linguistics with statistics for multiword term extraction: A fruitful association; In Proc. of Recherche d'Informations Assistée par Ordinateur.

Dias, Gaël Harry. (2003). Multiword Unit Hybrid Extraction. In *proceedings of the First Association for Computational Linguistics, Workshop on Multiword Expressions: Analysis, Acquisition and Treatment*, pp. 41-48.

Dunning, T., (1993). Accurate Methods for the Statistics of Surprise and Coincidence. In Association for Computational Linguistics, 19[1].

Ekbal, A., Mandal, S., & Bandyopadhyay, S. (2007). *POS tagging using HMM and rule based chunking*. Workshop on Shallow Parsing for South Asian Languages.

Evert, S. (2004). The Statistics of Word Cooccurrences: Word Pairs and Collocations. Ph. D. thesis, University of Stuttgart.

Evert, Stephen. (2004). *The Statistics of Word Co occurrences: Word Pairs and Collocations*. University of Stuttgart dissertation.

Fillmore, Charles, Paul Kay, and Mary C. O'Connor. (1988). Regularity and idiomaticity in grammatical constructions. *Language* 64.501–538.

Firth JR (1957) Papers in Linguistics 1934-1951. Oxford University Press, Oxford, p.233

Frantzi, K, S. Ananiadou, and H. Mima. 2000. Automatic Recognition of Multi-word term: the C-value/NC-value Method. *International Journal on Digital Libraries*, 3[2]:115–130, August.

GC, Seretan V (eds) Proceedings of the MT summit 2013 workshop on multi-word units in machine translation and translation technology (MUMTTT 2013), Nice, pp 11–11

Gibbs, R. W. (1980). Spilling the beans on understanding and memory for idioms in conversation. *Memory and Cognition* 8 (2), 149–156.

Green, B and Rubin, G (1971). Automated Grammatical Tagging of English. Department of Linguistics, Brown University.

Ghose, A., R. (2013). *Memory based learner for POS tagging*.

Harris Z. (1962). *String analysis of the language structure*. Mutton and Co., The Hauge.

Haspelmath, M. (1997). *From Space to Time in The World's Languages*. Munich, Germany: Lincorn Europa.

Heeman P.A. and Allen J.F. (1997). Incorporating POS tagging into language modeling. In *Proceedings of the eight conference on European chapter of the Association for Computational Linguistics*, Madrid, Spain. 230-237.

Huo , W. (2012), *Automatic Multi-word Term Extraction and its Application to Web-page Summarization*,page-43, Guelph, Ontario, Canada.

Jacquemin, Christian. (1999). Syntagmatic and paradigmatic representations of term variation. In *Proceedings of the 37rd Annual Meeting of the Association for Computational Linguistics*,pages 341.348. College Park, MD, USA, 20-26 June 1999.

Jurafsky D. and Martin J.H. (2009). *Speech and Language Processing (2nd ed.)*”, Pearson Education (Singapore).

Justeson, J. S., Katz, S. M., (1995). Technical Terminology: Some Linguistic Properties and an Algorithm for Identification in Text, *Natural Language Engineering*. 1(1):9-27.

Karthik Kumar G, Sudheer K and Avinesh Pvs. (2006). Comparative Study of Various Machine Learning Methods for Telegu Part of speech tagging. In *proceedings of the NLP AI Machine Learning 2006 Competition*.

Kastovsky, D. (1982). *Wortbildung und Semantik*. Dusseldorf: Bagel/Francke.

Katz, Jerrold J., and Paul M. Postal. (2004). Semantic interpretation of idioms and sentences containing them. In *Quarterly Progress Report (70)*, *MIT Research Laboratory of Electronics*, 275–282. MIT Press.

Klein S. and Simmons R.F (1963). A computational approach to grammatical coding of English words. *Journal of the Association for Computing Machinery*, Vol. 10, pp. 334-47.

Kumar, D., Josan G., S. (2010), “Part of Speech Taggers for Morphologically Rich Indian Languages: A Survey”, *International Journal of Computer. Applications* (0975

– 8887) Volume6–No.5, September, 2010, [www.ijcaonline.org/ volume6/number5 /pxc3871409 .pdf.](http://www.ijcaonline.org/volume6/number5/pxc3871409.pdf)

Kumar, E. (2011) '*Natural Language Processing*' ISBN:978-93-80578-77-4.

L. Dolamic (2010), "Comparative Study of Indexing and Search Strategies for the Hindi, Marathi and Bengali Language", *ACM Transactions on Asian Language Information Processing (TALIP)*, Volume 9, Issue 3.

Leech, G., and Wilson A. (2008). Recommendations for the Morphosyntactic Annotation of Corpora." March 1996. EAGLES Report. EAG-TCWG- MAC/R.

Leech, Geoffrey and Eyes, E. (1993) "Syntactic annotation: linguistic aspects of grammatical tagging and skeleton parsing". In, Black, Ezra,; Roger Garside, and Geoffrey Leech (Eds.) *Statistically-driven Computer Grammars of English: the IBM/Lancaster Approach*. Amsterdam: Rodopi. Pp. 36-61.

Leech, Geoffrey and Roger Garside (1982) "Grammatical tagging of the LOB Corpus: general survey". In, Johansson, Stig and Knut Hofland (Eds.) *Computer Corpora in English Language Research*. Bergen: NAVF. Pp. 110-117.

Li, Wentian. (1992). Random Texts Exhibit Zipf's-Law-Like Word Frequency Distribution. *IEEE Transactions on Information Theory* 38 (6): 1842–1845

Liberman, M. and R. Sproat (1992). The stress and structure of modified noun phrases in English. In I. A. Sag and A. Szabolcsi (Eds.), *Lexical Matters – CSLI Lecture Notes No. 24*. Stanford, USA: CSLI Publications.

Lin, Dekang. 1998b. Extracting collocations from text corpora. In *Proceedings of the 1st Workshop on Computational Terminology*, Montreal, Canada.

M.Minia. (2012), *Literature Surver on Multi-Lingual Multiword Expressions*, IIT Bombay

Mahmud, A., Khan, M. *Syntactic Part of Speech Tagging Guidelines for Bangla Text*. Center for Research on Bangla Language Processing (CRBLP), BRAC University, Dhaka, Bangladesh.

Manju K, Soumya S. and Idicula S. M. (2009). Development of a POS tagger for Malayalam – An Experience. International Conference on Advances in Recent Technologies in Communication and Computing, 2009.

Manning, C. & Schutze, H. (1999). Foundations of Statistical Natural Language Processing, chapter 5: Collocations. MIT Press.

McKeown, K. R. and D. R. Radev (2000). Collocations. In R. Dale, H. Moisl, and H. Somers (Eds.), *A Handbook of Natural Language Processing*, Chapter 15. Marcel Dekker.

Mimmelmann, N. P. (1998). Regularity in irregularity: Article use in adpositional phrases. *Linguistic Typology* 2, 315–353.

Mona Parakh, Rajesha N. and Ramya M (2011), “Sentence Boundary Disambiguation in Kannada Texts”, *Language in India* www.languageinindia.com 11:5 May 2011 Special Volume: Problems of Parsing in Indian Languages, Pages 17-19.

Nunberg, Geoffrey, Ivan A. Sag, and Tom Wasow. (1994). Idioms. *Language*. Pp. 491–538.

Pecine, P. (2005), *An extension empirical study of collocation extraction methods*. ACL standard research Workshop.

Pammi, S., C., Prahallad, K. (2007). POS Tagging and Chunking using Decision Forests, Workshop on shallow parsing in South Asian languages, 2007. shiva.iiit.ac.in/SPSAL2007/proceedings.php.

Pawley, Andrew, and Frances Hodgetts Syder (1983). Two puzzles for linguistic theory: nativelike selection and nativelike fluency.

Pearce, D. (2001). Synonymy in collocation extraction. In Proceedings of the NAACL 2001 Workshop on WordNet and Other Lexical Resources: Applications, Extensions and Customizations, Pittsburgh, USA, pp. 41–46.

Piao, S., P. Rayson, D. Archer, A. Wilson, and T. McEnery (2003). Extracting multiword expressions with a semantic tagger. In Proceedings of the ACL2003 Workshop on Multiword Expressions: analysis, acquisition and treatment, Sapporo, Japan, pp. 49–56.

Ray, J. (1998). *The Architecture of the Language Faculty*, Cambridge, MA, MIT Press.

Ray, J. (1997). *The Architecture of the Language Faculty*. Cambridge, USA: MIT Press.

Quirk, R., S. Greenbaum, G. Leech, and J. Svartvik (1985). *A Comprehensive Grammar of the English Language*. London, UK: Longman.

Riehemann, Susanne, (2001). *A Constructional Approach to Idioms and Word Formation*.

Ross, H. (1995). Defective noun phrases. In In Papers of the 31st Regional Meeting of the Chicago Linguistics Society, Chicago, Illinois, USA, pp. 398–440.

Sag I, Baldwin T, Bond F, Copestake A, Flickinger D. (2002) Multiword Expressions: a pain in the neck for NLP. In Proceedings of 3rd International conference on intelligent text processing and computational Linguistics (CICLing-2002), Mexico-city. Lecture notes in computer science, vol 2276/2010. Springer, pp1-15

Sag, I. A., T. Baldwin, F. Bond, A. Copestake, and D. Flickinger (2002). Multiword expressions: A pain in the neck for NLP. In Proceedings of the 3rd International Conference on Intelligent Text Processing and Computational Linguistics (CICLing-2002), Mexico City, Mexico, pp.1–15. Springer. Stanford University dissertation.

Schmid, H., (1994), ‘*Probabilistic part-of-speech tagging using decision trees*, ‘Proceeding of International Conference on New Methods in Language Processing.

Seretan, V. (2013). *On translating syntactically-flexible expressions*. In: Mitkov R, Monti J, Pastor

Seretan, V. (2011). *Syntax-based collocation extraction, text, speech and language technology*, vol 44, 1st edn. Springer, Dordrecht, p.212

Silva, J.F, G.Dias, S.Guilloré, J.G.P. Lopes. (1999). Using Lo-calMaxs Algorithm for the Extraction of Contiguous and Non-contiguous Multiword Lexical Units. In P. Barahona, editor, *Progress in Artificial Intelligence: 9th Portuguese Conference on AI, EPIA'99, Évora Portugal September 1999*, Proceedings. LNAI series, Springer-Verlag, Vol. 1695, p. 113-132.

Siddiqui, T., Tiwary, U.S.(2008) *Language Processing and Information Retrieval*, ISBN-10-0-19-569232-2.

Stvan, L. S. (1998). *The Semantics and Pragmatics of Bare Singular Noun Phrases*. Ph. D. thesis, Northwestern University.

Tanaka, T. and T. Baldwin (2003). Noun-noun compound machine translation a feasibility study on shallow processing. In *Proceedings of the ACL 2003 Workshop on Multiword Expressions: Analysis, Acquisition and Treatment*, Sapporo, Japan, pp. 17–24.

Toutanova, Krisahna, Dan Klein, Christopher D. Manning, and Yoram Singer, (2003), *Feature-rich part of Speech tagging with a cyclic dependency network*, *Proceedings of HLT NAACL* (available at <http://acl.ldc.upenn.edu/N/No3-1033.pdf>), pp.252-59.

Venkatapathy, S. and A. Joshi (2006). Using information about multi-word expressions for the word-alignment task. In *Proceedings of the COLING/ ACL 2006 Workshop on Multiword Expressions: Identifying and Exploiting Underlying Properties*, Sydney, Australia, pp. 53–60.

Villavicencio, Aline, Timothy Baldwin, and Benjamin Waldron. (2004). A multilingual database of idioms. In *Proceedings of the 4th International Conference on Language Resources and Evaluation (LREC-2004)*, 1127–1130, Lisbon, Portugal. www.tagoreweb.in (Rabindranath Tagore essay)

Xu, R., Q. Lu, S. Li (2006). The design and construction of a Chinese collocation bank. In Proceedings of the 5th International Conference on Language Resources and Evaluation, Genoa, Italy.