

# Bibliography

- Abeillé, Anne. 1995. The flexibility of French idioms: A representation with Lexical Tree Adjoining Grammar. In M. Everaert, E. van der Linden, A. Schen & R. Schreuder (eds.), *Idioms: Structural and psychological perspectives*, chap. 1, 15–42. Hillsdale: Lawrence Erlbaum Associates.
- Adar, E. 2004. SaRAD: A simple and robust abbreviation dictionary. *BioInformatics* 20. 527–533.
- Alegria, Iñaki, Olatz Ansa, Xabier Artola, Nerea Ezeiza, Koldo Gojenola & Ruben Urizar. 2004. Representation and Treatment of Multiword Expressions in Basque. In Takaaki Tanaka, Aline Villavicencio, Francis Bond & Anna Korhonen (eds.), *Second ACL Workshop on Multiword Expressions: Integrating Processing*, 48–55. Barcelona, Spain: Association for Computational Linguistics.
- Angelov, Krasimir. 2011. *The mechanics of the Grammatical Framework*. Chalmers University of Technology dissertation.
- Angelov, Krasimir. 2015. Orthography engineering in Grammatical Framework. In *Proceedings of the Grammar Engineering Across Frameworks (GEAF) Workshop, 53rd Annual Meeting of the ACL and 7th IJCNLP*, 33–40. Beijing, China.
- Angelov, Krasimir. 2019. Multiword expressions in multilingual applications within the Grammatical Framework. In Yannick Parmentier & Jakub Waszczuk (eds.), *Representation and parsing of multiword expressions: Current trends*, 129–148. Berlin: Language Science Press.
- Angelov, Krasimir, Björn Bringert & Aarne Ranta. 2014. Speech-Enabled hybrid multilingual translation for mobile devices. In *Proceedings of the Demonstrations at the 14th Conference of the European Chapter of the Association for Computational Linguistics*, 41–44. Gothenburg, Sweden: Association for Computational Linguistics. <http://aclweb.org/anthology/E14-2011>.
- Angelov, Krasimir & Aarne Ranta. 2010. Implementing controlled languages in GF. In *Proceedings of the Workshop on Controlled Natural Language (CNL'09)*, 82–101. Marettimo Island, Italy: Springer-Verlag.
- Apidianaki, Marianna, Prokopis Prokopidis & Haris Papageorgiou. 2018. Combining cross-lingual and syntactic evidence for Greek multiword expression identification. *Bulletin of Scientific Terminology and Neologisms of the Academy*

- of Athens, Special issue on MWEs in Greek and other languages: From theory to implementation **no volume. no pages.**
- Arun, Abhishek & Frank Keller. 2005. Lexicalization in crosslinguistic probabilistic parsing: The case of French. In *Proceedings of the 43rd Annual Meeting of the Association for Computational Linguistics (ACL'05)*, 306–313. Ann Arbor, Michigan: Association for Computational Linguistics. <http://aclweb.org/anthology/P05-1038>.
- Attia, Mohammed A. 2006. Accommodating Multiword Expressions in an Arabic LFG Grammar. In Tapio Salakoski, Filip Ginter, Tapio Pahikkala & Tampo Pyysalo (eds.), *Advances in natural language processing: 5th international Conference, FinTAL, Proceedings* (Lecture Notes in Computer Science 4139), 87–98. Berlin, Heidelberg: Springer.
- Bäckström, Linnéa, Benjamin Lyngfelt & Emma Sköldbberg. 2014. Towards interlingual constructicography: On correspondence between construction resources for English and Swedish. *Constructions and Frames* 6. 9–33.
- Bagga, Amit & Breck Baldwin. 1998. Algorithms for scoring coreference chains. *Proceedings of the First International Conference on Language Resources and Evaluation, Workshop on Linguistic Coreference* **no volume**. 563–566.
- Baldwin, Timothy & Su Nam Kim. 2010. Multiword expressions. In Nitin Indurkha & Fred J. Damerau (eds.), *Handbook of natural language processing*, Second edition, chap. 12, 267–292. Boca Raton, FL, USA: Chapman & Hall/CRC.
- Bansal, Mohit & Dan Klein. 2011. Web-scale features for full-scale parsing. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies (ACL-HLT'11)*, 693–702. Portland, Oregon. <http://www.aclweb.org/anthology/P11-1070>.
- Barbu, Ana Maria. 2004. Simple linguistic methods for improving a word alignment algorithm. In *Proceedings of the 7th International Conference on the Statistical Analysis of Textual Data (JADT 2004)*, 88–98. Louvain-la-Neuve, Belgium.
- Bargmann, Sascha & Manfred Sailer. 2015. *Syntactic flexibility of non-decomposable idioms*. Poster at the 4th General PARSEME Meeting, Valletta, Malta.
- Bargmann, Sascha & Manfred Sailer. 2018. The syntactic flexibility of semantically non-decomposable idioms. In Manfred Sailer & Stella Markantonatou (eds.), *Mutliword expressions: Insights from a multi-lingual perspective*, 1–40. Language Science Press. DOI:10.5281/zenodo.1182587
- Baroni, Marco, Silvia Bernardini, Adriano Ferraresi & Eros Zanchetta. 2009. The WaCky wide web: A collection of very large linguistically processed web-crawled corpora. *Language Resources and Evaluation* 43(3). 209–226. <http://www.springerlink.com/index/10.1007/s10579-009-9081-4>.

- Barz, Irmhild. 1996. Komposition und Kollokation. In Clemens Knobloch & Burkhard Schaefer (eds.), *Nomination – fachsprachlich und gemeinsprachlich*, 127–146. **no address**: Springer.
- Bejček, Eduard, Jarmila Panevová, Jan Popelka, Pavel Straňák, Magda Ševčíková, Jan Štěpánek & Zdeněk Žabokrtský. 2012. Prague Dependency Treebank 2.5—a revisited version of PDT 2.0. In *Proc. Of COLING 2012*, 231–246. Bombay, India.
- Besaçon, R., G. De Chalendar, O. Ferret, F. Gara, M. Laib, O. Mesnard & N. Semmar. 2010. LIMA: A multilingual framework of linguistic analysis and linguistic resources development and evaluation. In *Proceedings of the Seventh International Conference on Language Resources and Evaluation (LREC 2010)*, 3697–3704.
- Bilgram, Thomas & Britt Keson. 1998. The construction of a tagged Danish corpus. In *Proceedings of the 11th Nordic Conference of Computational Linguistics (NODALIDA 1998)*, 129–139. Copenhagen, Denmark: Center for Sprogteknologi, University of Copenhagen, Denmark. <http://www.aclweb.org/anthology/W98-1614>.
- Birch, Alexandra, Miles Osborne & Philipp Koehn. 2007. CCG supertags in factored statistical machine translation. In *Proceedings of the Second Workshop on Statistical Machine Translation*, 9–16.
- Black, E., S. Abney, S. Flickenger, C. Gdaniec, C. Grishman, P. Harrison, D. Hindle, R. Ingria, F. Jelinek, J. Klavans, M. Liberman, M. Marcus, S. Roukos, B. Santorini & T. Strzalkowski. 1991. Procedure for quantitatively comparing the syntactic coverage of English grammars. In *Proceedings of the Workshop on Speech and Natural Language*, 306–311. Pacific Grove, California: Association for Computational Linguistics. DOI:10.3115/112405.112467
- Blank, Ingeborg. 2000. Terminology extraction from parallel technical texts. In **no editor** (ed.), *Parallel text processing*, 237–252. **no address**: Springer.
- Blunsom, Phil & Timothy Baldwin. 2006. Multilingual Deep Lexical Acquisition for HPSGs via Supertagging. In *Proceedings of the 2006 Conference on Empirical Methods in Natural Language Processing (EMNLP 2006)*, 164–171. Sydney.
- Bohnet, Bernd, Joakim Nivre, Igor Boguslavsky, Richard Farkas, Filip Ginter & Jan Hajic. 2013. Joint Morphological and Syntactic Analysis for Richly Inflected Languages. *Transactions of the Association for Computational Linguistics (TACL)* 1. 415–428.
- Bouamor, Dhouha, Nasredine Semmar & Pierre Zweigenbaum. 2012a. A study in using English-Arabic MultiWord Expressions for Statistical Machine Translation. In *Proceedings of the Fourth International Conference on Arabic Language Processing (CITALA-2012)*, 71–76. Rabat, Morocco.

## Bibliography

- Bouamor, Dhrouha, Nasredine Semmar & Pierre Zweigenbaum. 2012b. Automatic construction of a multiword expressions bilingual lexicon: A statistical machine translation evaluation perspective. In *Proceedings of the 3rd Workshop on Cognitive Aspects of the Lexicon (CogALex-III)*, 95–108. Mumbai, India.
- Bouamor, Dhrouha, Nasredine Semmar & Pierre Zweigenbaum. 2012c. Identifying bilingual Multi-Word Expressions for Statistical Machine Translation. In *Proceedings of the Eighth International Conference on Language Resources and Evaluation (LREC 2012)*, 674–679. Istanbul, Turkey.
- Boulaknadel, Siham, Béatrice Daille & Driss Aboutajdine. 2008. A multiterm extraction program for Arabic language. In *Proceedings of the International Conference on Language Resources and Evaluation (LREC 2008)*, 1485–1488. Marrakech, Morocco.
- Bresnan, Joan. 2001. *Lexical-Functional syntax*. Malden, MA: Blackwell.
- Brun, Caroline. 1998. Terminology finite-state preprocessing for computational LFG. In *COLING 1998 volume 1: The 17th international Conference on Computational Linguistics*, 196–200. Montreal, QC.
- Buchholz, Sabine & Erwin Marsi. 2006. CoNLL-X Shared Task on Multilingual Dependency Parsing. In *Proceedings of the Tenth Conference on Computational Natural Language Learning (CoNLL-X 2006)*, 149–164. New York, NY.
- Burger, H. 2010. Phraseologie: Eine Einführung am Beispiel des Deutschen [Phraseology: An Introduction for German]. *Grundlagen der Germanistik* 36(4).  
no pages.
- Butt, Miriam, Helge Dyvik, Tracy Holloway King, Hiroshi Masuichi & Christian Rohrer. 2002. The Parallel Grammar Project. In John Carroll, Nelleke Oostdijk & Richard Sutcliffe (eds.), *Proceedings of the Workshop on Grammar Engineering and Evaluation at the 19th International Conference on Computational Linguistics (COLING)*, 1–7. Taipei, Taiwan: Association for Computational Linguistics.
- Cafferkey, Conor, Deirdre Hogan & Josef van Genabith. 2007. Multi-word units in treebank-based probabilistic parsing and generation. In *Proceedings of the International Conference on Recent Advances in Natural Language Processing (RANLP 2007)*. Borovets, Bulgaria.
- Calzolari, Nicoletta, Charles J Fillmore, Ralph Grishman, Nancy Ide, Alessandro Lenci, Catherine MacLeod & Antonio Zampolli. 2002. Towards Best Practice for Multiword Expressions in Computational Lexicons. In *Proceedings of the Third International Conference on Language Resources and Evaluation (LREC 2002)*, 1934–1940. Las Palmas, Canary Islands, Spain.
- Calzolari, Nicoletta, Alessandro Lenci, Francesca Bertagna & Antonio Zampolli. 2002. Broadening the scope of the EAGLES/ISLE lexical standardization initiative. In *COLING '02: Proceedings of the 3rd workshop on Asian language resources*

- and international standardization, 1–8. Morristown, NJ, USA: Association for Computational Linguistics.
- Candito, Marie & Matthieu Constant. 2014. Strategies for contiguous multiword expression analysis and dependency parsing. In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, 743–753. Baltimore, Maryland: Association for Computational Linguistics. <http://aclweb.org/anthology/P14-1070>.
- Candito, Marie & Benoît Crabbé. 2009. Improving generative statistical parsing with semi-supervised word clustering. In *Proc. Of the 11th international Conference on Parsing Technologies (IWPT'09)*, 138–141. Paris, France: Association for Computational Linguistics. <http://www.aclweb.org/anthology/W09-3821>.
- Candito, Marie, Joakim Nivre, Pascal Denis & Enrique Henestroza Anguiano. 2010. Benchmarking of Statistical Dependency Parsers for French. In *Proceedings of COLING 2010, 23rd International Conference on Computational Linguistics, Posters Volume*, 108–116. Beijing, China.
- Candito, Marie & Djamé Seddah. 2010. Parsing Word Clusters. In *Proceedings of the NAACL HLT 2010 First Workshop on Statistical Parsing of Morphologically-Rich Languages*, 76–84. Los Angeles, California.
- Carpuat, Marine & Mona Diab. 2010. Task-based evaluation of multiword expressions: A pilot study in statistical machine translation. In *Human Language Technologies: The 2010 annual Conference of the North American Chapter of the Association for Computational Linguistics (NAACL/HLT 2010)*, 242–245.
- Carroll, John, Guido Minnen & Edward Briscoe. 2003. Parser evaluation: Using a grammatical relation annotation scheme. In Anne Abeillé (ed.), *Treebanks: Building and using parsed corpora*, 299–316. Dordrecht: Kluwer.
- Chafe, Wallace L. 1968. Idiomaticity as an anomaly in the Chomskyan paradigm. *Foundations of language* 4(2). 109–127.
- Chang, Jeffrey T., Hinrich Schütze & Russ B. Altman. 2002. Creating an online dictionary of abbreviations from MEDLINE. *Journal of the American Medical Informatics Association* 9(6). 612–620.
- Charniak, Eugene. 2000. A maximum-entropy-inspired parser. In *Proceedings of the 1st North American chapter of the Association for Computational Linguistics conference*, 132–139. Seattle, Washington.
- Chomsky, Noam. 1957. *Syntactic Structures*. The Hague: Mouton.
- Chomsky, Noam. 1977. On Wh-movement. In P. W. Culicover, T. Wasw & A. Akmajian (eds.), *Formal syntax*, no pages. San Francisco, London: Academic Press.

- Choueka, Yaacov. 1988. Looking for needles in a haystack or locating interesting collocational expressions in large textual databases. In *Proceedings of the 2nd International Conference on Computer-Assisted Information Retrieval (Recherche d'Information Assistée par Ordinateur) - RIAO'88*, 609–624. Cambridge, MA, USA.
- Christodouloupoulos, Christos. 2008. *Creating a natural logic inference system with Combinatory Categorical Grammar*. University of Edinburgh MA thesis.
- Clark, Stephen & James R Curran. 2007. Wide-Coverage efficient statistical parsing with CCG and Log-linear models. *Computational Linguistics* 4(33). 493–552.
- Clark, Stephen & Julia Hockenmaier. 2002. Evaluating a wide-coverage CCG parser. In *Proceedings of the LREC 2002 Beyond PARSEVAL workshop*, 60–66.
- Collins, Michael. 1996. A new statistical parser based on bigram lexical dependencies. In *Proceedings of the 34th annual meeting of the Association for Computational Linguistics (ACL 1996)*, 184–191. Santa Cruz, California.
- Collins, Michael. 1999. *Head-Driven statistical models for natural language parsing*. Philadelphia, PA: University of Pennsylvania Ph.D. thesis.
- Collins, Michael & Yoram Singer. 1999. Unsupervised models for named entity classification. In *Proceedings of the Joint SIGDAT Conference on Empirical Methods in Natural Language Processing and Very Large Corpora (EMNLP-VLC)*, 100–110. College Park, MD: University of Maryland.
- Constable, James & James Curran. 2009. Integrating Verb-Particle constructions into CCG parsing. In *Proceedings of the Australasian Language Technology Association Workshop 2009*, 114–118. Sydney, Australia. <http://aclweb.org/anthology/U09-1017>.
- Constant, Mathieu, Gülşen Eryiğit, Carlos Ramisch, Mike Rosner & Gerold Schneider. 2019. Statistical MWE-aware parsing. In Yannick Parmentier & Jakub Waszczuk (eds.), *Representation and parsing of multiword expressions: Current trends*, 149–184. Berlin: Language Science Press.
- Constant, Matthieu, Marie Candito & Djamé Seddah. 2013. The LIGM-Alpage architecture for the SPMRL 2013 shared task: Multiword expression analysis and dependency parsing. In *Proceedings of the Fourth Workshop on Statistical Parsing of Morphologically-Rich Languages*, 46–52. Seattle, Washington, USA: Association for Computational Linguistics. <http://aclweb.org/anthology/W13-4905>.
- Constant, Matthieu, Joseph Le Roux & Anthony Sigogne. 2013. Combining compound recognition and PCFG-LA parsing with word lattices and conditional random fields. *ACM Transaction on Speech and Language Processing (TSLP), Special Issue on MWEs* 10(3). no pages.

- Constant, Matthieu, Joseph Le Roux & Nadi Tomeh. 2016. Deep lexical segmentation and syntactic parsing in the Easy-First dependency framework. In *Proceedings of the 15th Annual Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL HLT 2016)*, 1095–1101. San Diego, California.
- Constant, Matthieu & Joakim Nivre. 2016. A Transition-Based system for joint lexical and syntactic analysis. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics*, vol. 1: *Long papers*, 161–171. Berlin, Germany: Association for Computational Linguistics. <http://www.aclweb.org/anthology/P16-1016>.
- Constant, Matthieu, Anthony Sigogne & Patrick Watrin. 2012. Discriminative strategies to integrate multiword expression recognition and parsing. In *Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics: Long Papers – Volume 1 (ACL '12)*, 204–212. Jeju Island, Korea: Association for Computational Linguistics. <http://dl.acm.org/citation.cfm?id=2390524.2390554>.
- Constant, Matthieu, Isabelle Tellier, Denys Duchier, Yoann Dupont, Anthony Sigogne & Sylvie Billot. 2011. Intégrer des connaissances linguistiques dans un CRF: Application à l'apprentissage d'un segmenteur-étiqueteur du français. In *TALN*, vol. 1, 321.
- Copestake, Ann. 2002. *Implementing Typed Feature Structure Grammars*. Stanford: CSLI Publications.
- Copestake, Ann, Fabre Lambeau, Aline Villavicencio, Francis Bond, Timothy Baldwin, Ivan A. Sag & Dan Flickinger. 2002. Multiword expressions: Linguistic precision and reusability. In *Proceedings of the Third International Conference on Language Resources and Evaluation (LREC 2002)*, 1941–1947. Las Palmas, Canary Islands, Spain.
- Crabbé, Benoît, Denys Duchier, Claire Gardent, Joseph Le Roux & Yannick Parmentier. 2013. XMG: eXtensible MetaGrammar. *Computational Linguistics* 39(3). 591–629. DOI:10.1162/COLI\_a\_00144
- Curran, James, Stephen Clark & Johan Bos. 2007. Linguistically motivated Large-Scale NLP with C&C and boxer. In *Proceedings of the 45th Annual Meeting of the Association for Computational Linguistics, Companion Volume, Proceedings of the Demo and Poster Sessions*, 33–36. Prague, Czech Republic: Association for Computational Linguistics.
- Daelemans, Walter & Koenraad De Smedt. 1994. Default inheritance in an object-oriented representation of linguistic categories. *International Journal of Human-Computer Studies* 41(1). 149–177.



- Daelemans, Walter & Erik-Jan van der Linden. 1992. *Evaluation of lexical representation formalisms*. ITK Research Memo. Tilburg: Institute for Language Technology & Artificial Intelligence, Tilburg University.
- Dagan, Ido & Ken Church. 1994. Termight: Identifying and translating technical terminology. In *Proceedings of the Fourth Conference on Applied Natural Language Processing (ANLP'94)*, 34–40.
- Daille, Béatrice. 2001. Extraction de collocation à partir de textes. In *Actes de la 8ème conférence sur le traitement automatique des langues naturelles (TALN)*, 3–8. Tours, France: Association pour le Traitement Automatique des Langues.
- Daille, Béatrice, Eric Gaussier & Jean-Marc Langé. 1994. Towards automatic extraction of monolingual and bilingual terminology. In *Proceedings of the 15th International Conference on Computational linguistics (COLING 1994)*, 515–521.
- Dalrymple, Mary. 2001. *Lexical functional grammar* (Syntax and Semantics 34). San Diego, CA: Academic Press.
- Dalrymple, Mary, Ronald M. Kaplan & Tracy Holloway King. 2004. Linguistic generalizations over descriptions. In Miriam Butt & Tracy H. King (eds.), *Proceedings of the LFG'04 Conference*, 200–208. University of Canterbury, Christchurch, New Zealand: CSLI Publications.
- de Lhoneux, Miryam, Omri Abend & Mark Steedman. 2019. Investigating the effect of automatic MWE recognition on CCG parsing. In Yannick Parmentier & Jakub Waszczuk (eds.), *Representation and parsing of multiword expressions: Current trends*, 185–217. Berlin: Language Science Press.
- DeNero, John & Dan Klein. 2008. The complexity of phrase alignment problems. In *Proceedings of the 46th Annual Meeting of the Association for Computational Linguistics on Human Language Technologies: Short Papers*, 25–28.
- Denis, Pascal & Benoît Sagot. 2012. Coupling an Annotated Corpus and a Lexicon for State-of-the-art POS Tagging. *Language Resources and Evaluation* 46(4). 721–736. DOI:10.1007/s10579-012-9193-0
- Deoskar, Tejaswini, Christos Christodoulopoulos, Alexandra Birch & Mark Steedman. 2014. Generalizing a strongly lexicalized parser using unlabeled data. In Gosse Bouma & Yannick Parmentier (eds.), *Proceedings of the 14th Conference of the European Chapter of the Association for Computational Linguistics (EACL'2014)*, 126–134. Gothenburg.
- Détrez, Grégoire & Aarne Ranta. 2012. Smart paradigms and the predictability and complexity of inflectional morphology. In *Proceedings of the 13th Conference of the European Chapter of the Association for Computational Linguistics (EACL 2012)*, 645–653. Avignon, France: Association for Computational Linguistics. <http://aclweb.org/anthology/E12-1066>.



- Doron, Edit. 2003. Agency and voice: The semantics of the Semitic templates. *Natural Language Semantics* 11(1). 1–67.
- Doron, Edit. 2008. The contribution of the template to verb meaning. In Galia Hataav (ed.), *Modern linguistics of Hebrew*, 57–88. Jerusalem: Magnes Press. In Hebrew.
- Downey, Doug, Matthew Broadhead & Oren Etzioni. 2007. Locating complex named entities in web text. In *Proceedings of the 20th International Joint Conference on Artificial Intelligence (IJCAI'07)*, 2733–2739. Hyderabad, India: Morgan Kaufmann Publishers Inc. <http://dl.acm.org/citation.cfm?id=1625275.1625715>.
- Dunning, Ted. 1993. Accurate Methods for the Statistics of Surprise and Coincidence. *Computational Linguistics* 19(1). 61–74. <http://dl.acm.org/citation.cfm?id=972450.972454>.
- Durrett, Greg & Dan Klein. 2015. Neural CRF parsing. In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, 302–312. Beijing.
- Dyer, Chris, Miguel Ballesteros, Wang Ling, Austin Matthews & Noah A. Smith. 2015. Transition-based dependency parsing with stack long short-term memory. In *Proc. Of ACL 2015*, 334–343. Beijing.
- Dyer, Chris, Victor Chahuneau & Noah A. Smith. 2013. A Simple, Fast, and Effective Reparameterization of IBM Model 2. In *Proceedings of the 2013 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL-HLT)*, 644–648.
- Dyvik, Helge. 2000. Nødvendige noder i norsk: Grunntrekk i en leksikalsk-funksjonell beskrivelse av norsk syntaks [Necessary nodes in Norwegian: Basic properties of a lexical-functional description of Norwegian syntax]. In Øivin Andersen, Kjersti Fløttum & Torodd Kinn (eds.), *Menneske, språk og felleskap*, 25–45. Oslo: Novus forlag.
- Dyvik, Helge, Gyri Smørdal Losnegaard & Victoria Rosén. 2019. Multiword expressions in an LFG grammar for Norwegian. In Yannick Parmentier & Jakub Waszczuk (eds.), *Representation and parsing of multiword expressions: Current trends*, 69–107. Berlin: Language Science Press.
- Dyvik, Helge, Paul Meurer, Victoria Rosén, Koenraad De Smedt, Petter Haugereid, Gyri Smørdal Losnegaard, Gunn Inger Lyse & Martha Thunes. 2016. NorGramBank: A ‘Deep’ Treebank for Norwegian. In Nicoletta Calzolari, Khalid Choukri, Thierry Declerck, Marko Grobelnik, Bente Maegaard, Joseph Mariani, Asunción Moreno, Jan Odijk & Stelios Piperidis (eds.), *Proceedings of the tenth international conference on language resources and evaluation (LREC*

- 2016), 3555–3562. Portorož, Slovenia. <http://www.lrec-conf.org/proceedings/lrec2016/summaries/943.html>.
- Ehrmann, Maud, Guillaume Jacquet & Ralf Steinberger. 2015. JRC-Names: Multilingual entity name variants and titles as Linked Data. *Semantic Web* **no volume**(Preprint). 1–13.
- Ehrmann, Maud, Leo D. Rocca, Ralf Steinberger & Hristo Tanev. 2013. Acronym recognition and processing in 22 languages. In *Proceedings of the 9th Conference Recent Advances in Natural Language Processing (RANLP)*, 237–244. Hissar, Bulgaria.
- Eisner, Jason M. 1996. Three new probabilistic models for dependency parsing: An exploration. In *Proceedings of the 16th conference on Computational linguistics (ACL 1996)*, 340–345. Santa Cruz, California.
- Enache, Ramona, Inari Listenmaa & Prasanth Kolachina. 2014. Handling non-compositionality in multilingual CNLs. In *Controlled natural language. CNL 2014* (Lecture Notes in Computer Science 8625), 147–154. Cham: Springer.
- Erman, Britt & Beatrice Warren. 2000. The idiom principle and the open choice principle. *Text* 20(1). 29–62.
- Ernst, Thomas. 1981. Grist for the linguistic mill: Idioms and ‘extras’ adjectives. *Journal of Linguistic Research* 1(3). 51–68.
- Eryiğit, Gülşen, Tugay İlbay & Ozan Arkan Can. 2011. Multiword expressions in statistical dependency parsing. In *Proceedings of the Second Workshop on Statistical Parsing of Morphologically Rich Languages*, 45–55. Dublin, Ireland.
- Evans, Roger & Gerald Gazdar. 1996. DATR: A language for lexical knowledge representation. *Computational Linguistics* 22(2). 167–216.
- Fazly, Afsaneh, Paul Cook & Suzanne Stevenson. 2009. Unsupervised type and token identification of idiomatic expressions. *Computational Linguistics* 35(1). 61–103.
- Fellbaum, Christiane (ed.). 1998. *WordNet: An electronic lexical database* (Language, Speech and Communication). **no address**: MIT Press.
- Finkel, Jenny Rose & Christopher D. Manning. 2009. Joint Parsing and Named Entity Recognition. In *Proceedings of Human Language Technologies: The 2009 Annual Conference of the North American Chapter of the Association for Computational Linguistics*, 326–334. Boulder, Colorado: Association for Computational Linguistics. <http://www.aclweb.org/anthology/N/N09/N09-1037>.
- Finlayson, Mark Alan & Nidhi Kulkarni. 2011. Detecting multi-word expressions improves word sense disambiguation. In *Proceedings of the Workshop on Multiword Expressions: From Parsing and Generation to the Real World* (MWE ’11), 20–24. Stroudsburg, PA, USA: Association for Computational Linguistics.

- Foufi, Vasiliki, Luka Nerima & Eric Wehrli. 2019. Multilingual parsing and MWE detection. In Yannick Parmentier & Jakub Waszczuk (eds.), *Representation and parsing of multiword expressions: Current trends*, 219–240. Berlin: Language Science Press.
- Fox, Edward A. & Joseph A. Shaw. 1994. Combination of multiple searches. *NIST Special Publication no volume*. 243–243.
- Francopoulo, Gil, Monte George, Nicoletta Calzolari, Monica Monachini, Nuria Bel, Mandy Pet & Claudia Soria. 2006. Lexical Markup Framework (LMF). In *Proceedings of the International Conference on Language Resources and Evaluation (LREC 2006)*, 233–236.
- Fraser, Alexander & Daniel Marcu. 2007. Measuring word alignment quality for statistical machine translation. *Computational Linguistics* 33(3). 293–303.
- Gao, Zhao-Ming. 1998. Automatic Acquisition of a High-Precision Translation Lexicon from Parallel Chinese-English Corpora. *Language* 248. 254.
- Gaussier, Éric & François Yvon. 2011. *Modèles statistiques pour l'accès à l'information textuelle*. no address: Lavoisier.
- Gibbs, Raymond W. & Nandini P. Nayak. 1989. Psycholinguistic studies on the syntactic behavior of idioms. *Cognitive psychology* 21(1). 100–138.
- Gibbs, Raymond W., Nandini P. Nayak, John L. Bolton & Melissa E. Keppel. 1989. Speakers' assumptions about the lexical flexibility of idioms. *Memory & Cognition* 17(1). 58–68.
- Głowińska, Katarzyna & Adam Przepiórkowski. 2010. The Design of Syntactic Annotation Levels in the National Corpus of Polish. In *Proc. Of the seventh international Conference on Language Resources and Evaluation (LREC 2010)*. Valletta, Malta.
- Green, Spence, Marie-Catherine de Marneffe, John Bauer & Christopher D. Manning. 2011. Multiword expression identification with tree substitution grammars: A Parsing tour de force with French. In *Proc. Of EMNLP 2011*, 725–735. Edinburgh.
- Green, Spence, Marie-Catherine de Marneffe & Christopher D. Manning. 2013. Parsing models for identifying multiword expressions. *Computational Linguistics* 39(1). 195–227.
- Grégoire, Nicole. 2010a. DuELME: A Dutch electronic lexicon of multiword expressions. *Language Resources and Evaluation* 44(1–2). 23–39.
- Grégoire, Nicole. 2007a. Design and implementation of a lexicon of Dutch Multiword Expressions. In *Proceedings of the Workshop on A Broader Perspective on Multiword Expressions*, 17–24. Prague, Czech Republic: Association for Computational Linguistics. <http://www.aclweb.org/anthology/W/W07/W07-1103>.

## Bibliography

- Grégoire, Nicole. 2007b. *MWE Lexicon for Dutch: Encoding protocol*. <http://duelme.clarin.inl.nl/documentation.php>.
- Grégoire, Nicole. 2007c. *MWE Lexicon for Dutch: Overview of pattern descriptions*. <http://duelme.clarin.inl.nl/documentation.php>.
- Grégoire, Nicole. 2010b. *DUELME LMF XML data: User documentation*. <http://duelme.clarin.inl.nl/documentation.php>.
- Gross, Maurice. 1984. Lexicon-grammar and the syntactic analysis of French. In *Proceedings of the 10th International Conference on Computational Linguistics and 22nd Annual Meeting of the Association for Computational Linguistics*, 275–282. Stanford, California, USA: Association for Computational Linguistics.
- Gross, Maurice. 1986. Lexicon-grammar: The Representation of Compound Words. In *Proceedings of the 11th Conference on Computational Linguistics (COLING '86)*, 1–6. Bonn, Germany: Association for Computational Linguistics.
- Gross, Maurice. 1988a. Les limites de la phrase figée. *Langage* 90. 7–23.
- Gross, Maurice. 1988b. Sur les phrases figées complexes du français. *Langue française* 77. 47–70.
- Gross, Maurice. 1994. Constructing Lexicon-Grammars. In B. T.S. Atkins & A. Zampolli (eds.), *Computational approaches to the lexicon*, 213–263. Oxford: Oxford University Press.
- Grover, Claire. 2008. *LT-TTT2: Example pipelines documentation*. Tech. rep. Edinburgh Language Technology Group.
- Gruzitis, Normunds, Dana Dannells, Benjamin Lyngfelt & Aarne Ranta. 2015. Formalising the Swedish constructicon in Grammatical Framework. In *Proceedings of the Grammar Engineering Across Frameworks (GEAF) 2015 Workshop*, 49–56. Beijing, China: Association for Computational Linguistics. <http://aclweb.org/anthology/W15-3307>.
- Habert, Benoît & Christian Jacquemin. 1995. Construction nominales à contraintes fortes et grammaires d'unification. *Linguisticae Investigationes* 19(2). 401–427.
- Hahn, Udo, Philipp Daumke, Stefan Schulz & Kornél Markó. 2005. Cross-language mining for acronyms and their completions from the web. *Proceedings of the 8th international conference on Discovery Science (DS'05)* 9. 113–123.
- Hedström, Björn, Matilda Horppu & David Michaëlsson. 2016. *Parlira: An interactive phrasebook for Android devices*. Tech. rep. Chalmers University of Technology.
- Heid, Ulrich. 1994. On Ways Words Work Together – Topics in Lexical Combinatorics. In Willy Martin et al. (ed.), *Proceedings of the VIth Euralex International Congress (EURALEX'94)*, 226–257. Amsterdam. Eingeladener Hauptvortrag.

- Henderson, James, Paola Merlo, Ivan Titov & Gabriele Musillo. 2013. Multilingual Joint Parsing of Syntactic and Semantic Dependencies with a Latent Variable Model. *Computational Linguistics* 39(4). 949–998.
- Hockenmaier, Julia & Mark Steedman. 2002. Generative models for statistical parsing with Combinatory Categorical Grammar. In *Proceedings of the 40th Annual Meeting on Association for Computational Linguistics (ACL '02)*, 335–342. Stroudsburg, PA, USA: Association for Computational Linguistics.
- Hockenmaier, Julia & Mark Steedman. 2007. CCGbank: A corpus of CCG derivations and dependency structures extracted from the Penn Treebank. *Computational Linguistics* 33(3). 355–396.
- Hockenmaier, Julie. 2003. *Data and models for statistical parsing with Combinatory Categorical Grammar*. University of Edinburgh dissertation.
- Hoffart, Johannes, Fabian M Suchanek, Klaus Berberich & Gerhard Weikum. 2013. YAGO2: A spatially and temporally enhanced knowledge base from Wikipedia. In *Proceedings of the Twenty-Third international joint conference on Artificial Intelligence*, 3161–3165.
- Hoffmann, Thomas & Graeme Trousdale. 2013. Construction grammar: Introduction. In Thomas Hoffmann & Graeme Trousdale (eds.), *The Oxford handbook of Construction Grammar* (Oxford Handbooks in Linguistics), 1–14. **no address**: Oxford University Press.
- Hogan, Deirdre, Conor Cafferkey, Aoife Cahill & Josef Van Genabith. 2007. Exploiting multi-word units in history-based probabilistic generation. In *Proceedings of the Joint Conference on Empirical Methods in Natural Language Processing (EMNLP) and Computational Natural Language Learning (CoNLL)*, 267–276. Prague, Czech Republic.
- Jackendoff, Ray. 1997. *The architecture of the language faculty*. Cambridge, MA, USA: MIT Press.
- Jacquet, Guillaume, Maud Ehrmann, Jakub Piskorski, Hristo Tanev & Ralf Steinberger. 2019. Cross-lingual linking of multi-word entities and language-dependent learning of multi-word entity patterns. In Yannick Parmentier & Jakub Waszczuk (eds.), *Representation and parsing of multiword expressions: Current trends*, 271–300. Berlin: Language Science Press.
- Jacquet, Guillaume, Maud Ehrmann & Ralf Steinberger. 2014. Clustering of Multi-Word Named Entity variants: Multilingual Evaluation. In *Proceedings of the 9th Language Resources and Evaluation Conference (LREC 2014)*, 2548–2553. Reykjavik, Iceland.
- Jacquet, Guillaume, Maud Ehrmann, Ralf Steinberger & Jaakko Väyrynen. 2016. Cross-lingual Linking of Multi-word Entities and their corresponding Acro-

- nyms. In *Proceedings of the 10th Language Resources and Evaluation Conference (LREC 2016)*, 528–535. Portorož, Slovenia.
- James, J. Pustejovsky, José Castano, Brent Cochran, Maciej Kotecki & Michael Morrell. 2001. Automatic extraction of acronym-meaning pairs from MEDLINE databases. *Studies in health technology and informatics* 1. 371–375.
- Kaplan, Ronald M. & Joan Bresnan. 1982. Lexical-Functional Grammar: A Formal System for Grammatical Representation. In Joan Bresnan (ed.), *The mental representation of grammatical relations*, 173–281. Cambridge, MA: The MIT Press.
- Kato, Akihiko, Hiroyuki Shindo & Yuji Matsumoto. 2016. Construction of an English Dependency Corpus incorporating Compound Function Words. In *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC 2016)*. Portorož, Slovenia: European Language Resources Association (ELRA).
- Kay, Paul & Ivan A. Sag. 2012. *A lexical theory of phrasal idioms*. Stanford, CA. <http://www1.icsi.berkeley.edu/~kay/idiom-pdflatex.11-13-15.pdf>.
- Kilbury, James, Petra Naerger & Ingrid Renz. 1991. DATR as a lexical component for PATR. In *Proceedings of the fifth conference of the European Chapter of the Association for Computational Linguistics (EACL-91)*, 137–142.
- Kilgarrieff, Adam, Pavel Rychly, Pavel Smrz & David Tugwell. 2004. The Sketch engine. In *Proceedings of EURALEX*, 105–116. Lorient, France.
- Kim, Su Nam. 2008. *Statistical Modeling of Multiword Expressions*. University of Melbourne dissertation.
- Koehn, Philipp & Hieu Hoang. 2007. Factored Translation Models. In *Proceedings of the 2007 joint conference on empirical methods in Natural Language Processing and Computational Natural Language Learning (EMNLP-CoNLL)*, 868–876.
- Koehn, Philipp, Hieu Hoang, Alexandra Birch, Chris Callison-Burch, Marcello Federico, Nicola Bertoldi, Brooke Cowan, Wade Shen, Christine Moran, Richard Zens, Chris Dyer, Ondrej Bojar, Alexandra Constantin & Eva Herbst. 2007. Moses: Open source toolkit for statistical machine translation. In *Proceedings of the 45th annual meeting of the ACL on interactive poster and demonstration sessions*, 177–180. Prague, Czech Republic.
- Koehn, Philipp, Franz J. Och & Daniel Marcu. 2003. Statistical phrase-based translation. In *Proceedings of the 2003 Conference of the North American Chapter of the Association for Computational Linguistics on Human Language Technology*, vol. 1, 48–54.
- Kokkinakis, Dimitrios & Dana Dannélls. 2006. Recognizing acronyms and their definitions in Swedish medical texts. In *Proceedings of the 5th Conference on Language Resources and Evaluation (LREC)*, 1971–1974. Genoa, Italy.

- Kompara, Mojca. 2010. Automatic recognition of abbreviations and abbreviations' expansions in multilingual electronic texts. In Chris Cummins, Chi-Hé Elder, Thomas Godard, Morgan Macleod, Elaine Schmidt & George Walkden (eds.), *Proceedings of the Sixth Cambridge Postgraduate Conference in Language Research (CAMLing)*, 82–91.
- Koo, Terry, Xavier Carreras & Michael Collins. 2008. Simple Semi-supervised Dependency Parsing. In *Proceedings of ACL-08: HLT*, 595–603. Columbus, Ohio.
- Korkontzelos, Ioannis & Suresh Manandhar. 2010. Can recognising multiword expressions improve shallow parsing? In *Proc. Of the 11th annual Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies NAACL/HLT 2010*, 636–644. Los Angeles, California.
- Krishnamurthy, Jayant & Tom M Mitchell. 2012. Weakly Supervised Training of Semantic Parsers. In *Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning (EMNLP/CoNLL)*, 754–765.
- Kupiec, Julian. 1993. An algorithm for finding noun phrase correspondences in bilingual corpora. In *Proceedings of the 31st annual meeting on Association for Computational Linguistics*, 17–22.
- Le Roux, Joseph, Antoine Rozenknop & Matthieu Constant. 2014. Syntactic parsing and compound recognition via dual decomposition: Application to French. In *Proc. Of COLING 2014*. Dublin, Ireland.
- Lehmann, Hans Martin & Gerold Schneider. 2011. A large-scale investigation of verb-attached prepositional phrases. In S. Hoffmann, P. Rayson & G. Leech (eds.), *Studies in variation, contacts and change in English, Volume 6: Methodological and historical dimensions of corpus linguistics*, no pages. Helsinki: Varieng.
- Lewis, Mike & Mark Steedman. 2013. Combining distributional and logical semantics. *Transactions of the Association for Computational Linguistics* 1. 179–192.
- Lichte, Timm & Simon Petitjean. 2015. Implementing semantic frames as typed feature structures with XMG. *Journal of Language Modelling* 3(1). 185–228.
- Lichte, Timm, Simon Petitjean, Agata Savary & Jakub Waszczuk. 2019. Lexical encoding formats for multi-word expressions: The challenge of “irregular” regularities. In Yannick Parmentier & Jakub Waszczuk (eds.), *Representation and parsing of multiword expressions: Current trends*, 1–33. Berlin: Language Science Press.
- Lønning, Jan Tore, Stephan Oepen, Dorothee Beermann, Lars Hellan, John Carroll, Helge Dyvik, Dan Flickinger, Janne Bondi Johannessen, Paul Meurer, Torbjørn Nordgård, Victoria Rosén & Erik Velldal. 2004. LOGON: A Norwegian



- MT Effort. In *Proceedings of the Workshop in Recent Advances in Scandinavian Machine Translation*. Uppsala, Sweden.
- Losnegaard, Gyri Smørdal, Gunn Inger Lyse, Martha Thunes, Victoria Rosén, Koenraad De Smedt, Helge Dyvik & Paul Meurer. 2012. What We have learned from Sofie: Extending lexical and grammatical coverage in an LFG parsebank. In Jan Hajič, Koenraad De Smedt, Marko Tadić & António Branco (eds.), *META-RESEARCH Workshop on Advanced Treebanking at LREC'2012*, 69–76. Istanbul, Turkey.
- Losnegaard, Gyri Smørdal & Harald Ulland. 2015. *Syntactic properties of Norwegian verbal MWEs*. Poster session B, PARSEME 4th General Meeting, Valletta, Malta.
- Lyngfelt, Benjamin, Lars Borin, Markus Forsberg, Julia Prentice, Rudolf Rydstedt, Emma Sköldberg & Sofia Tingsell. 2012. Adding a constructicon to the Swedish resource network of Språkbanken. In Jeremy Jancsary (ed.), *Proceedings of KONVENS 2012*, 452–461. ÖGAI. LexSem 2012 workshop.
- Marchand, Morgane & Nasredine Semmar. 2011. A hybrid Multi-Word terms alignment approach using word co-occurrence with a Bilingual lexicon. In *Proceedings of the fifth Language and Technology Conference: Human Language Technologies as a Challenge for Computer Science and Linguistics*, 430–434. Poznań, Poland.
- Markantonatou, Stella & Niki Samaridi. 2018. Revisiting the grammatical function “object” (OBJ and OBJ<sub>0</sub>). In Manfred Sailer & Stella Markantonatou (eds.), *Mutliword expressions: Insights from a multi-lingual perspective*, 187–213. **no address**: Language Science Press. DOI:10.5281/zenodo.1182599
- Markantonatou, Stella, Niki Samaridi & Panagiotis Minos. 2019. Issues in parsing MWEs in an LFG/XLE framework. In Yannick Parmentier & Jakub Waszczuk (eds.), *Representation and parsing of multiword expressions: Current trends*, 109–127. Berlin: Language Science Press.
- Martens, Scott & Vincent Vandeghinste. 2010. An efficient, generic approach to extracting Multi-Word expressions from dependency trees. In *Proceedings of the Workshop on Multiword Expressions: From Theory to Applications (MWE 2010)*, 84–87. Beijing, China: Association for Computational Linguistics.
- Maxwell, John & Ronald M. Kaplan. 1993. The interface between phrasal and functional constraints. *Computational Linguistics* 19(4). 571–589.
- McConnel, Stephen. 1997. *PC-PATR Reference Manual*. Summer Institute of Linguistics. Dallas, TX. <http://www.sil.org/pcpatr/manual/pcpatr.html>. Version 0.99b5.

- McDonald, Ryan, Kevin Lerman & Fernando Pereira. 2006. Multilingual Dependency Analysis with a Two-Stage Discriminative Parser. In *Proceedings of the Tenth Conference on Computational Natural Language Learning (CoNLL-X)*, 216–220. New York City: Association for Computational Linguistics. <http://www.aclweb.org/anthology/W/W06/W06-2932>.
- McNamee, Paul & Hoa T. Dang. 2009. Overview of the TAC 2009 Knowledge Base Population Track. In *Text analysis Conference (TAC)*, vol. 17, 111–113.
- Mel'čuk, Igor, Nadia Arbatchewsky-Jumarie, Louise Dagenais, Léo Elnitsky, Lidiya Iordanskaja, Marie-Noëlle Lefebvre & Suzanne Mantha. 1988. *Dictionnaire explicatif et combinatoire du français contemporain: Recherches lexico-sémantiques* (Recherches lexico-sémantiques II). no address: Presses de l'Univ. de Montréal.
- Melnik, Nurit. 2016. *The development of the VOS construction*. Paper presented at the “Language contact, language continuity and change, and the emergence of Modern Hebrew” workshop, Hebrew University, Jerusalem.
- Meurer, Paul. 2012. INESS-Search: A search system for LFG (and other) treebanks. In Miriam Butt & Tracy Holloway King (eds.), *Proceedings of the LFG'12 Conference* (LFG Online Proceedings), 404–421. Stanford, CA: CSLI Publications.
- Miller, George Armitage. 1956. The magical number seven, Plus or minus two: Some limits on our capacity for processing information. *Psychological Review* 63. 81–97.
- Minos, Panagiotis, Stella Markantonatou, Georgios Zakis & Elpiniki Margariti. 2016. *Generating LFG/XLE MWE entries from IDION (a theory neutral lexical DB)*. Struga, fYROM. <http://typo.uni-konstanz.de/parseme/index.php/2-general/156-selected-posters-struga-7-8-april-2016>. 6th PARSEME general meeting.
- Mirroshandel, Seyed Abolghasem & Alexis Nasr. 2016. Integrating selectional constraints and subcategorization frames in a dependency parser. *Computational Linguistics* 42(1). 55–90.
- Mirroshandel, Seyed Abolghasem, Alexis Nasr & Joseph Le Roux. 2012. Semi-supervised dependency parsing using lexical affinities. In *Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, 777–785. Jeju Island, Korea: Association for Computational Linguistics. <http://aclweb.org/anthology/P12-1082>.
- Mirroshandel, Seyed Abolghasem, Alexis Nasr & Benoît Sagot. 2013. Enforcing subcategorization constraints in a parser using sub-parses recombining. In *Proceedings of the 2013 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, 239–247. Atlanta, Georgia: Association for Computational Linguistics.

- Moon, Rosamund. 1998. *Fixed Expressions and idioms in English: A Corpus-Based approach*. Oxford: Oxford University Press.
- Müller, Stefan & Stephen M. Wechsler. 2014. Lexical approaches to argument structure. *Theoretical Linguistics* 40(1–2). 1–76.
- Nadeau, David & Peter D. Turney. 2005. A supervised learning approach to acronym identification. In Balázs Kégl & Guy Lapalme (eds.), *Proceedings of the Canadian Conference on Artificial Intelligence, Lecture Notes in Computer Science (LNCS), Advances in Artificial Intelligence*, 319–329. Berlin, Heidelberg: Springer.
- Nagy T., István & Veronika Vincze. 2014. VPCTagger: Detecting verb-particle constructions with syntax-based methods. In *Proceedings of the EACL 2014 Workshop on MWEs*, 17–25. Gothenburg.
- Nakagawa, Tetsuji. 2007. Multilingual dependency parsing using global features. In *Proceedings of the CoNLL Shared Task Session of EMNLP-CoNLL 2007*, 952–956. Prague, Czech Republic: Association for Computational Linguistics. <http://www.aclweb.org/anthology/D07-1100>.
- Nasr, Alexis, Frederic Bechet, Jean-Francois Rey, Benoit Favre & Joseph Le Roux. 2011. MACAON: An NLP tool suite for processing word lattices. In *Proceedings of ACL 2011 Demonstrations*. Portland, Oregon.
- Nasr, Alexis, Carlos Ramisch, José Deulofeu & André Valli. 2015. Joint dependency parsing and multiword expression tokenisation. In *53rd annual meeting of the Association for Computational Linguistics*, 1116–1126. Beijing, China.
- Navigli, Roberto & Simone Paolo Ponzetto. 2012. BabelNet: The automatic construction, evaluation and application of a wide-coverage multilingual semantic network. *Artificial Intelligence* 193. 217–250. DOI:10.1016/j.artint.2012.07.001
- Nerima, Luka, Eric Wehrli & Violeta Seretan. 2010. A Recursive Treatment of Collocations. In *Proceedings of the Seventh International Conference on Language Resources and Evaluation (LREC'10)*, 634–638. Valletta, Malta.
- Nilsson, Jens, Sebastian Riedel & Deniz Yuret. 2007. The CoNLL 2007 shared task on dependency parsing. In *Proceedings of EMNLP/CoNLL 2007 CoNLL shared tasks session*, 915–932.
- Nitoń, Bartłomiej, Tomasz Bartosiak & Elżbieta Hajnicz. 2016. Accessing and elaborating *Walenty* – a valence dictionary of Polish – via Internet browser. In *Proceedings of the 10th edition of the Language Resources and Evaluation Conference (LREC)*, 1352–1359. Portorož, Slovenia.
- Nivre, Joakim. 2014. *Transition-based parsing with multiword expressions*. Athens.
- Nivre, Joakim, Marie-Catherine de Marneffe, Filip Ginter, Yoav Goldberg, Jan Hajič, Christopher D. Manning, Ryan McDonald, Slav Petrov, Sampo Pyysalo, Natalia Silveira, Reut Tsarfaty & Dan Zeman. 2016. Universal Dependencies

- v1: A Multilingual Treebank Collection. In *Proceedings of the 10th International Conference on Language Resources and Evaluation (LREC 2016)*. Portorož, Slovenia.
- Nivre, Joakim, Johan Hall & Jens Nilsson. 2004. Memory-based dependency parsing. In *Proceedings of CoNLL 2004*, 49–56. Boston, Massachusetts.
- Nivre, Joakim, Johan Hall, Jens Nilsson, Atanas Chanev, Gülsen Eryigit, Sandra Kübler, Svetoslav Marinov & Erwin Marsi. 2007. MaltParser: A language-independent system for data-driven dependency parsing. *Natural Language Engineering* 13(02). 95–135.
- Nivre, Joakim & Jens Nilsson. 2004a. Multiword units in syntactic parsing. In *Workshop on Methodologies and Evaluation of Multiword Units in Real-World Applications*, 39–46.
- Nivre, Joakim & Jens Nilsson. 2004b. Multiword units in syntactic parsing. *Proceedings of Methodologies and Evaluation of Multiword Units in Real-World Applications (MEMURA)* no volume. no pages.
- Nordgård, Torbjørn. 2000. Nordkompleks: A Norwegian computational lexicon. In *COMLEX 2000 Workshop on Computational Lexicography and Multimedia Dictionaries*, 89–92. Patras, Greece: University of Patras.
- Nothman, Joel, Nicky Ringland, Will Radford, Tara Murphy & James R. Curran. 2013. Learning Multilingual Named Entity Recognition from Wikipedia. *Artificial Intelligence* 194. 151–175. DOI:10.1016/j.artint.2012.03.006
- Nunberg, Geoffrey, Ivan A. Sag & Thomas Wasow. 1994. Idioms. *Language* 70(3). 491–538.
- Och, F. J. & H. Ney. 2000. Improved statistical alignment models. In *Proceedings of the 38th Annual Meeting on Association for Computational Linguistics (ACL 2000)*, 440–447.
- Odijk, Jan. 2013. DUELME: Dutch electronic lexicon of multiword expressions. In Gil Francopoulo (ed.), *LMF: Lexical Markup Framework*, chap. 9, 133–144. no address: Wiley-ISTE.
- Okazaki, Naoaki & Sophia Ananiadou. 2006. Building an abbreviation dictionary using a term recognition approach. *Bioinformatics* 22. 3089–3095.
- Okazaki, Naoaki, Sophia Ananiadou & Jun'ichi Tsujii. 2010. Building a high-quality sense inventory for improved abbreviation disambiguation. *Bioinformatics* 26. 1246–1253.
- Okita, Tsuyoshi, Alfredo Maldonado Guerra, Yvette Graham & Andy Way. 2010. Multi-word expression-sensitive word alignment. In *4th international Workshop on Cross Lingual Information Access at COLING 2010*, 26–34. Beijing, China.

## Bibliography

- Padó, Sebastian. 2006. *User's guide to sigf: Significance testing by approximate randomisation*.
- Papageorgiou, Haris, Prokopis Prokopidis, Voula Giouli & Stelios Piperidis. 2000. A unified POS tagging architecture and its application to Greek. In M. Gavriliadou, G. Carayannis, S. Markantonatou, S. Piperidis & G. Stainhauer (eds.), *Proceedings of the 2nd Language Resources and Evaluation Conference (LREC 2000)*, 1455–1462. Athens, Greece.
- Papineni, Kishore, Salim Roukos, Todd Ward & Wei-Jing Zhu. 2002. BLEU: A method for automatic evaluation of machine translation. In *Proceedings of the 40th annual meeting on Association for Computational Linguistics (ACL 2002)*, 311–318.
- Pawley, Andrew & Frances Hodgetts Syder. 1983. Two Puzzles for Linguistic Theory: Native-like selection and native-like fluency. In J. C. Richards & R. W. Schmidt (eds.), *Language and communication*, 191–226. London: Longman.
- Pereira, Fernando C. N. & Stuart M. Shieber. 1984. The semantics of grammar formalisms seen as computer languages. In *Proceedings of the 10th international conference on computational linguistics and the 22nd annual meeting of the association for computational linguistics*, 123–129. Stanford, CA.
- Petitjean, Simon, Denys Duchier & Yannick Parmentier. 2016. XMG2: Describing Description Languages. In Maxime Amblard, Philippe de Groote, Sylvain Pogo-dalla & Christian Retoré (eds.), *Logical Aspects of Computational Linguistics. Celebrating 20 Years of LACL (1996–2016): 9th International Conference, LACL 2016, Nancy, France, December 5-7, 2016, Proceedings*, 255–272. Berlin, Heidelberg: Springer.
- Petrov, Slav, Dipanjan Das & Ryan McDonald. 2012. A universal Part-of-Speech tagset. In Nicoletta Calzolari, Khalid Choukri, Thierry Declerck, Mehmet Uğur Doğan, Bente Maegaard, Joseph Mariani, Asuncion Moreno, Jan Odijk & Stelios Piperidis (eds.), *Proceedings of the Eighth International Language Resources and Evaluation Conference, LREC 2012*, 2089–2096. Istanbul, Turkey. <http://www.lrec-conf.org/proceedings/lrec2012/summaries/274.html>.
- Piskorski, Jakub, Martin Atkinson & Jenya Belyaeva. 2011. Exploring the usefulness of cross-lingual information fusion for refining real-time news Event Extraction: A preliminary study. In *Proceedings of the conference Recent Advances in Natural Language Processing*, 210–217. Hissar, Bulgaria.
- Pollard, Carl & Ivan A. Sag. 1987. *Information-based syntax and semantics* (CSLI Lecture Notes 13). Stanford: CSLI.
- Prins, Robbert. 2005. *Finite-State pre-processing for natural language analysis*. Behavioral & Cognitive Neurosciences (BCN) research school, University of Groningen dissertation.

- Przepiórkowski, Adam, Jan Hajič, Elżbieta Hajnicz & Zdeňka Urešová. 2016. Phraseology in two Slavic Valency Dictionaries: Limitations and Perspectives. *International Journal of Lexicography* 29. no pages.
- Przepiórkowski, Adam, Elżbieta Hajnicz, Agnieszka Patejuk & Marcin Woliński. 2014. Extended phraseological information in a valence dictionary for NLP applications. In *Proceedings of the Workshop on Lexical and Grammatical Resources for Language Processing (LG-LP 2014)*, 83–91. Dublin, Ireland.
- Ramisch, Carlos, Aline Villavicencio & Valia Kordoni. 2013. Introduction to the special issue on multiword expressions: From theory to practice and use. *ACM Transactions on Speech and Language Processing (TSLP)* 10(2). 3.
- Ramshaw, Lance A. & Mitchell P. Marcus. 1995. Text chunking using transformation-based learning. In *Proceedings of the 3rd ACL Workshop on Very Large Corpora*, 82–94.
- Ranta, Aarne. 2009. The GF Resource Grammar Library. *Linguistic Issues in Language Technology* no volume. no pages.
- Ranta, Aarne. 2011. *Grammatical Framework: Programming with multilingual grammars*. Stanford: CSLI Publications.
- Ranta, Aarne, Krasimir Angelov & Thomas Hallgren. 2010. Tools for multilingual grammar-based translation on the web. In *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics (ACL 2010) System Demonstrations*, 66–71. Uppsala, Sweden.
- Ranta, Aarne, Christina Unger & Daniel Vidal Hussey. 2015. Grammar engineering for a customer: A case study with five languages. In *Proceedings of the Grammar Engineering Across Frameworks (GEAF) 2015 Workshop*, 1–8. Beijing, China: Association for Computational Linguistics. <http://aclweb.org/anthology/W15-3301>.
- Rao, D., P. McNamee & M. Dredze. 2013. Entity linking: Finding Extracted Entities in a knowledge base. In no editor (ed.), *Multi-source, multilingual information extraction and summarization*, 93–115. no address: Springer.
- Ren, Z., Y. Lü, J. Cao, Q. Liu & Y. Huang. 2009. Improving statistical machine translation using domain bilingual multiword expressions. In *Proceedings of the Workshop on Multiword Expressions: Identification, Interpretation, Disambiguation and Applications*, 47–54.
- Riehemann, Susanne. 2001. *A constructional approach to idioms and word formation*. Stanford University dissertation.
- Riloff, Ellen. 1996. Automatically generating extraction patterns from untagged text. In *Proceedings of Thirteenth National Conference on Artificial Intelligence (AAAI-96)*, 1044–1049. The AAAI Press/MIT Press.

- Ronan, Patricia & Gerold Schneider. 2015. Determining light verb constructions in contemporary British and Irish English. *International Journal of Corpus Linguistics* 20(3). 326–354.
- Rosén, Victoria, Koenraad De Smedt, Paul Meurer & Helge Dyvik. 2012. An open infrastructure for advanced treebanking. In Jan Hajič, Koenraad De Smedt, Marko Tadić & António Branco (eds.), *META-RESEARCH Workshop on Advanced Treebanking at LREC'2012*, 22–29. Istanbul, Turkey.
- Rosén, Victoria, Paul Meurer & Koenraad De Smedt. 2007. Designing and implementing discriminants for LFG grammars. In Tracy Holloway King & Miriam Butt (eds.), *The Proceedings of the LFG'07 Conference*, 397–417. Stanford: CSLI Publications.
- Rosén, Victoria, Gyri Smørdal Losnegaard, Koenraad De Smedt, Eduard Bejček, Agata Savary, Adam Przepiórkowski, Petya Osenova & Verginica Barbu Mititelu. 2015. A Survey of Multiword Expressions in Treebanks. In *Proc. Of 14th international Workshop on Treebanks and Linguistic Theories (TLT 2015)*, 179–193. Warsaw, Poland.
- Rosén, Victoria, Martha Thunes, Petter Haugereid, Gyri Smørdal Losnegaard, Helge Dyvik, Paul Meurer, Gunn Inger Lyse & Koenraad De Smedt. 2016. The enrichment of lexical resources through incremental parsebanking. *Language Resources and Evaluation* 50(2). 291–319.
- Sag, Ivan, Timothy Baldwin, Francis Bond, Ann Copestake & Dan Flickinger. 2002. Multiword expressions: A pain in the neck for NLP. In *Proceedings of the 3rd International Conference on Computational Linguistics and Intelligent Text Processing* (Lecture Notes in Computer Science 2276), 1–15. Springer.
- Sagot, Benoît & Pierre Boullier. 2005. From raw corpus to word lattices: Robust pre-parsing processing with SxPipe. *Archives of Control Sciences* 15(4). 653–662.
- Salton, G., A. Wong & C. S. Yang. 1975. A vector space model for automatic indexing. *Communications of the ACM* 18(11). 613–620.
- Samaridi, Niki & Stella Markantonatou. 2014. Parsing Modern Greek verb MWEs with LFG/XLE grammars. In *10th Workshop on Multiword Expressions (MWE 2014), European chapter of the Association for Computational Linguistics 2014*, 33–37. Gothenburg, Sweden.
- Savary, Agata. 2008. Computational inflection of Multi-Word units: A contrastive study of lexical approaches. *Linguistic Issues in Language Technology* 1(2). 1–53.
- Schenk, André. 1995. The syntactic behavior of idioms. In M. Everaert, E. van der Linden, A. Schen & R. Schreuder (eds.), *Idioms: Structural and psychological perspectives*, 253–272. Hillsdale: Lawrence Erlbaum Associates.



- Schneider, Gerold. 2008. *Hybrid Long-Distance functional dependency parsing*. Institute of Computational Linguistics, University of Zurich Doctoral Thesis.
- Schneider, Gerold. 2012. Using semantic resources to improve a syntactic dependency parser. In *Proceedings of Semantic Relations II workshop (SEM-II) at LREC 2012*, 67–76. Istanbul, Turkey.
- Schneider, Gerold. 2014. Improving PP attachment in a hybrid dependency parser using semantic, distributional, and lexical resources. In *Second PARSEME meeting*. Athens, Greece.
- Schneider, Nathan, Emily Danchik, Chris Dyer & Noah A. Smith. 2014. Discriminative lexical semantic segmentation with gaps: Running the MWE gamut. *Transactions of the Association for Computational Linguistics (TACL)* 2. 193–206.
- Schneider, Nathan, Spencer Onuffer, Nora Kazour, Emily Danchik, Michael T. Mordowanec, Henrietta Conrad & Noah A. Smith. 2014. Comprehensive annotation of multiword expressions in a social web corpus. In *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC 2014)*, 456–461. Reykjavik.
- Schwartz, Ariel S. & Marti A. Hearst. 2003. A simple algorithm for identifying abbreviation definitions in biomedical text. In *Proceedings of the PAC on Bio-computing*, 451–462.
- Seddah, Djamé, Reut Tsarfaty, Sandra Kübler, Marie Candito, Jinho Choi, Richárd Farkas, Jennifer Foster, Iakes Goenaga, Koldo Gojenola, Yoav Goldberg, Spence Green, Nizar Habash, Marco Kuhlmann, Wolfgang Maier, Joakim Nivre, Adam Przepiorkowski, Ryan Roth, Wolfgang Seeker, Yannick Versley, Veronika Vincze, Marcin Woliński, Alina Wróblewska & Eric Villemonte de la Clérgerie. 2013. Overview of the SPMRL 2013 Shared Task: A Cross-Framework Evaluation of Parsing Morphologically Rich Languages. In *Proceedings of the Fourth International Workshop on Statistical Parsing of Morphologically-Rich Languages (SPRML IV)*. Seattle, WA.
- Seki, Hiroyuki, Takashi Matsumura, Mamoru Fujii & Tadao Kasami. 1991. On multiple context-free grammars. *Theoretical Computer Science* 88(2). 191–229.
- Sekine, Satoshi & Michael Collins. 1997. *EVALB bracket scoring program*. <http://www.cs.nyu.edu/cs/projects/proteus/evalb>.
- Semmar, Nasredine, Christophe Servan, Gaël De Chalendar, Benoît Le Ny & Jean-Jacques Bouzaglou. 2010. A hybrid word alignment approach to improve translation lexicons with compound words and idiomatic expressions. In *Proceedings of the 32nd Translating and the Computer Conference (ASLIB 2010)*. London, UK.

- Semmar, Nasredine, Christophe Servan, Meriama Laib, Dhouha Bouamor & Morgane Marchand. 2019. Extracting and aligning multiword expressions from parallel corpora. In Yannick Parmentier & Jakub Waszczuk (eds.), *Representation and parsing of multiword expressions: Current trends*, 241–270. Berlin: Language Science Press.
- Seretan, Violeta. 2011. *Syntax-Based collocation extraction* (Text, Speech and Language Technology 44). Dordrecht: Springer.
- Seretan, Violeta. 2013. On collocations and their interaction with parsing and translation. *Informatics* 1(1). 11–31.
- Seretan, Violeta & Eric Wehrli. 2007. Collocation translation based on sentence alignment and parsing. In *Actes de la 14ème conférence sur le Traitement Automatique des Langues Naturelles (TALN 2007)*, 401–410. Toulouse, France.
- Sheinflux, Livnat Herzig, Tali Arad Greshler, Nurit Melnik & Shuly Wintner. 2019. Verbal multiword expressions: Idiomaticity and flexibility. In Yannick Parmentier & Jakub Waszczuk (eds.), *Representation and parsing of multiword expressions: Current trends*, 35–68. Berlin: Language Science Press.
- Shieber, Stuart M. 1984. The design of a computer language for linguistic information. In *Proceedings of the 10th International Conference on Computational Linguistics and 22nd Annual Meeting of the Association for Computational Linguistics (ACL 1984)*, 362–366. Stanford, CA. <http://www.aclweb.org/anthology/P84-1075>.
- Shieber, Stuart M. 1986. *An introduction to Unification-Based approaches to grammar* (CSLI Lecture Notes Series 4). Stanford, CA: CSLI.
- Siklósi, Borbála, Attila Novák & Gábor Prószéky. 2014. Resolving abbreviations in clinical texts without pre-existing structured resources. In *4th Workshop on Building and Evaluating Resources for Health and Biomedical Text Processing (BioTxtM), LREC*, 69–75. Reykjavik, Iceland.
- Sinclair, John. 1991. *Corpus, concordance, collocation*. Oxford: Oxford University Press.
- Steedman, Mark. 2000. *The syntactic process*. Cambridge, MA, USA: MIT Press.
- Steinberger, Ralf, Maud Ehrmann, Júlia Pajzs, Mohamed Ebrahim, Josef Steinberger & Marco Turchi. 2013. Multilingual media monitoring and text analysis—Challenges for highly inflected languages. In *International Conference on Text, Speech and Dialogue*, 22–33.
- Steinberger, Ralf, Aldo Podavini, Alexandra Balahur, Guillaume Jacquet, Hristo Tanev, Jens Linge, Martin Atkinson, Michele Chinosiand, Vanni Zavarella, Yaniv Steiner & Erik van der Goot. 2015. Observing trends in automated multilingual media analysis. In *Proceedings of the Symposium on New Frontiers of Automated Content Analysis in the Social Sciences (ACA'2015)*, 1–8.

- Steinberger, Ralf, Bruno Pouliquen, Mijail Kabadjov & Erik van der Goot. 2011. JRC-Names: A freely available, highly multilingual named entity resource. In *Proceedings of the 8th International Conference Recent Advances in Natural Language Processing (RANLP'2011)*, 104–110. Hissar, Bulgaria.
- Steinberger, Ralf, Bruno Pouliquen & Erik van der Goot. 2009. An introduction to the Europe Media Monitor family of applications. In *Proceedings of the SIGIR 2009 Workshop (SIGIR-CLIR'2009)*, 1–8. Boston, USA.
- Taghva, Kazen & Jeff Gilbreth. 1999. Recognizing acronyms and their definitions. *International Journal on Document Analysis and Recognition* 1(4). 191–198.
- Tesnière, Lucien. 1959. *Éléments de syntaxe structurale*. no address: Klincksieck.
- Tiedemann, Jörg. 2012. Parallel data, tools and interfaces in OPUS. In *Proceedings of the Eighth International Conference on Language Resources and Evaluation (LREC-2012)*, 2214–2218. Istanbul, Turkey.
- Titone, Debra A. & Cynthia M. Connine. 1994. Descriptive norms for 171 idiomatic expressions: Familiarity, compositionality, predictability, and literal-ity. *Metaphor and Symbol* 9(4). 247–270.
- Tjong Kim Sang, Erik F. 2002. Introduction to the CoNLL-2002 shared task: Language-independent named entity recognition. In *Proceedings of the 6th Conference on Natural Language Learning (CoNLL) – Volume 20*, 1–4. Taipei, Taiwan.
- Tomasello, Michael. 1998. Cognitive linguistics. In W. Bechtel & G. Graham (eds.), *A companion to cognitive science*, no pages. no address: Basil Blackwell.
- Trawinski, Beata, Manfred Sailer, Jan-Philipp Soehn, Lothar Lemnitzer & Frank Richter. 2008. Cranberry Expressions in English and in German. In *Proceedings of the LREC Workshop Towards a Shared Task for Multiword Expressions (MWE 2008)*, 35–38.
- Tu, Yuancheng & Dan Roth. 2011. Learning English light verb constructions: Contextual or statistical. In *Proceedings of the ACL 2011 Workshop on MWEs*, 31–39. Portland, OR.
- Tuşiş, Dan & Radu Ion. 2007. Parallel corpora, alignment technologies and further prospects in multilingual resources and technology infrastructure. In *Proceedings of the 4th International Conference on Speech and Dialogue Systems*, 183–195.
- Ullman, Edvin & Joakim Nivre. 2014. Paraphrasing Swedish compound nouns in Machine Translation. In *Proceedings of the 10th Workshop on Multiword Expressions (MWE)*, 99–103. Gothenburg, Sweden: Association for Computational Linguistics. <http://www.aclweb.org/anthology/W14-0817>.
- Vechtomova, Olga. 2005. The role of multi-word units in interactive information retrieval. In *European Conference on Information Retrieval*, 403–420.

- Villavicencio, Aline, Valia Kordoni, Yi Zhang, Marco Idiart & Carlos Ramisch. 2007. Validation and evaluation of automatically acquired multiword expressions for grammar engineering. In *Proceedings of the 2007 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning (EMNLP-CoNLL)*, 1034–1043. Prague, Czech Republic: Association for Computational Linguistics. <http://www.aclweb.org/anthology/D/D07/D07-1110>.
- Vincze, Veronika, István Nagy T. & Gábor Berend. 2011. Multiword expressions and named entities in the Wiki50 corpus. In *Proceedings of the International Conference Recent Advances in Natural Language Processing 2011*, 289–295. Hissar, Bulgaria: Association for Computational Linguistics.
- Vincze, Veronika, János Zsibrita & István Nagy T. 2013. Dependency parsing for identifying Hungarian light verb constructions. In *Proceedings of the Sixth International Joint Conference on Natural Language Processing (IJCNLP)*, 207–215. Nagoya, Japan: Asian Federation of Natural Language Processing. <http://aclweb.org/anthology/I13-1024>.
- Vintar, Špela & Darja Fišer. 2008. Harvesting Multi-Word expressions from parallel corpora. In *Proceedings of the 6th edition of the Language Resources and Evaluation Conference (LREC 2008)*, 1091–1096.
- Waszczuk, Jakub & Agata Savary. 2015. *Modeling syntactic properties of MWEs in LFG*. La Valletta. 4th PARSEME general meeting.
- Webelhuth, Gert & Farrell Ackerman. 1999. A lexical-functional analysis of predicate topicalization in German. *American Journal of Germanic Linguistics and Literatures* 11(01). 1–61.
- Weeds, Julie, James Dowdall, Gerold Schneider, Bill Keller & David Weir. 2007. Using distributional similarity to organise BioMedical terminology. In Fidelia Ibekwe-SanJuan, Anne Condamines & M. Teresa Cabré Castellví (eds.), *Application-Driven terminology engineering*, no pages. Amsterdam/Philadelphia: Benjamins.
- Wehrli, Eric. 2007. Fips: A “deep” linguistic multilingual parser. In *Proceedings of the ACL 2007 Workshop on Deep Linguistic Processing*, 120–127. Prague, Czech Republic: Association for Computational Linguistics. <http://www.aclweb.org/anthology/W/W07/W07-1216>.
- Wehrli, Eric & Luka Nerima. 2013. Anaphora resolution, collocations and translation. In J. Monti, R. Mitkov, G. Corpas Pastor & V. Seretan (eds.), *Proceedings of the Workshop on Multi-word Units in Machine Translation and Translation Technology (MUMTT’2013)*. Nice.

- Wehrli, Eric & Luka Nerima. 2015. The Fips multilingual parser. In N. Gala, R. Rapp & G. Bel-Enquix (eds.), *Language Production Cognition, and the Lexicon, festschrift in honour of Michael Zock*, 473–489. **no address**: Springer.
- Weller, Marion & Ulrich Heid. 2010. Extraction of German multiword expressions from parsed corpora using context features. In Nicoletta Calzolari, Khalid Choukri, Bente Maegaard, Joseph Mariani, Jan Odijk, Stelios Piperidis, Mike Rosner & Daniel Tapias (eds.), *Proceedings of LREC 2010*, 3195–3201. Valletta: European Language Resources Association.
- White, Michael. 2006. Efficient realization of coordinate structures in Combinatory Categorical Grammar. *Research on Language and Computation* 4(1). 39–75.
- Wray, Alison. 2008. *Formulaic language: Pushing the boundaries*. **no address**: Oxford University Press.
- Wren, Jonathan D. & Harold R. Garner. 2002. Heuristics for identification of Acronym-Definition patterns within text: Towards an automated construction of comprehensive Acronym-Definition dictionaries. *Methods of Information in Medicine* 41(5). 426–434.
- Wu, Chien-Cheng & Jason S Chang. 2003. Bilingual collocation extraction based on syntactic and statistical analyses. In *Proceedings of Research on Computational Linguistics Conference XV (ROCLING'2003)*, 33–55. Hsinchu, Taiwan: The Association for Computational Linguistics & Chinese Language Processing (ACLCLP). <http://aclweb.org/anthology/O03-1003>.
- Yamada, Hiroyasu & Yuji Matsumoto. 2003. Statistical dependency analysis with support vector machines. In *Proceedings of the International Workshop on Parsing Technologies (IWPT)*, vol. 3, 195–206. Nancy, France.
- Yangarber, Roman, Winston Lin & Ralph Grishman. 2002. Unsupervised learning of generalized names. In *Proceedings of COLING: The 19th International Conference on Computational Linguistics*. Taipei, Taiwan.
- Zhang, Hao, Liang Huang, Daniel Gildea & Kevin Knight. 2006. Synchronous binarization for machine translation. In *Proceedings of the main conference on Human Language Technology conference of the North American chapter of the Association of Computational Linguistics*, 256–263. New York, New York: Association for Computational Linguistics.
- Zhang, Y. & S. Vogel. 2006. *Suffix array and its applications in empirical Natural Language Processing*. Technical Report CMU-LTI-06-010. Language Technologies Institute, School of Computer Science, Carnegie Mellon University, Pittsburgh, PA.
- Zhang, Yi & Valia Kordoni. 2006. Automated deep lexical acquisition for robust open texts processing. In *Proceedings of the 5th International Conference on Language Resource and Evaluation (LREC-2006)*, 275–280. Genoa, Italy.

## *Bibliography*

- Zhang, Yi & Valia Kordoni. 2008. Robust parsing with a large HPSG grammar. In *Proceedings of the Sixth International Language Resources and Evaluation Conference (LREC'08)*, 1888–1893. Marrakech, Morocco.
- Zhang, Yi, Valia Kordoni, Aline Villavicencio & Marco Idiart. 2006. Automated multiword expression prediction for grammar engineering. In *Proceedings of the Workshop on Multiword Expressions: Identifying and Exploiting Underlying Properties (MWE '06)*, 36–44. Stroudsburg, PA, USA: Association for Computational Linguistics.