A selective bibliography on the lexicon in MT

Ann Copestake
CSLI, Stanford University
aac@csli.stanford.edu

This bibliography includes references to some of the recent work in the following areas: lexicalist and semantic transfer based MT approaches, lexical semantics in MT, acquisition of monolingual and multilingual lexical information. For lexical semantics and acquisition from a monolingual perspective, see also the bibliography for the 1996 ACL tutorial on Lexical Semantics: available via my home page: `http://www-csli.stanford.edu/users/aac`. Note also the Journal of Machine Translation Special issue on Building Lexicons for MT: 9:3–4, 10:1–2, plus Klavans and Tzoukermann in 10:3–4. References below are listed with URLs where I'm aware of them: `cmp-lg` indicates the Computation and language electronic archive which can be accessed via `http://xxx.lanl.gov/cmp-lg/`.

Some other sources of online information and resources, which may be useful:

- Bibliography of NLP at University of Essex:
  `http://clwww.essex.ac.uk/search/`
  (Searchable via the WWW - other bibliographic databases also linked.) This is the best online bibliography for MT that I'm aware of, though it's apparently not fully up-to-date.
- Chris Manning's list of NLP resources, now at:
  `http://www.arts.su.edu.au/Arts/departs/ling/manning/statnlp.html`
  This has a very good set of links to sources of corpora etc.
- SIGLEX — ACL special interest group on the lexicon:
  `http://www.seas.upenn.edu/~gemilord/siglex2.html`
- LDC — mostly AmE spoken corpora, but also some text corpora and COMLEX dictionaries:
  `http://www.ldc.upenn.edu`.
- CRL — no longer funded, but useful wordlists and tools still available, concentrating much more on the lexicon than LDC does:
  `http://crl.nmsu.edu/clr/CLR.html`
- ELRA — European languages resources association:
  `http://www.icp.grenet.fr/ELRA/home.html`
- MULTEXT — freely available tools and corpora for non-commercial purposes:
  `http://www.lpl.univ-aix.fr/projects/multext/`

References

Ageno, A., I. Castellon, G. Rigau, H. Rodriguez, M.F. Verdejo, M.A. Marti and M. Taule (1992) 'SEISD: An environment for extraction of semantic information from on-line dictionaries', *Proceedings of the 3rd Conference on Applied Natural Language Processing (ANLP-92),* Trento, Italy, pp. 253–255.

Ageno, A., F. Ribas, G. Rigau, H. Rodriguez, A. Samiotou (1994) 'TGE: Tlinks Generation Environment', *Proceedings of the 15th International Congress on Computational Linguistics (COLING 94),* Kyoto, Japan.

Alshawi, H. (ed.) (1992) *The Core Language Engine,* MIT Press, Cambridge, Mass.

Alshawi, H. (1994) 'Qualitative and quantitative models of speech translation', *Proceedings of the Workshop on the Balancing Act, ACL-94,* Las Cruces, New Mexico.
`cmp-lg/9408014`

Alshawi, H. (1996) 'Head automata and bilingual tiling: translation with minimal representation', *Proceedings of the 34th Annual Meeting of the Association for Computational Linguistics (ACL-96),* Santa Cruz, California, pp. 167–176.

Alshawi, H., D. Carter, M. Rayner and B. Gambäck (1991) 'Translation by Quasi Logical Form transfer', *Proceedings of the 29th Annual Meeting of the Association for Computational Linguistics (ACL-91),* Berkeley, California, pp. 161–168.
`http://www.cam.sr.com/tr/crc021/paper.ps.Z`

Arnold, D. and L. Sadler (1990) 'The theoretical basis of MiMo', *Machine Translation,* **5 (3)**, 195–222.

Beaven, J. L. (1992) 'Lexicalist unification-based machine translation', PhD thesis, University of Edinburgh.

Black, E., F. Jelinek, J. Lafferty, D.M. Magerman, R. Mercer, S. Roukos (1993) 'Towards history-based grammars: using richer models for probabilistic parsers', *Proceedings of the 31st Annual Meeting of the Association of Computational Linguistics,* Ohio State University, Columbus, Ohio, pp. 31–37.

Boguraev, B. and E. J. Briscoe (eds) (1989) *Computational lexicography for natural language processing,* Longman, London.

Brent, M. (1991) 'Automatic acquisition of subcategorization frames from untagged text', *Proceedings of the 29th Annual Meeting of the Association for Computational Linguistics,* Berkeley, CA, pp. 209–214.

Brent, M. (1993) 'From grammar to lexicon: unsupervised learning of lexical syntax', *Computational Linguistics,* **19.3**, 243–262.

Briscoe, E.J. and J. Carroll (1995) 'Towards automatic extraction of argument structure from corpora', ACQUILEX II Working Paper.

Brown, P. F., J. Cocke, S.A. DellaPietra, V.J. DellaPietra, F. Jelinek, J.D. Lafferty, R.L. Mercer and P.S. Roosin (1990) 'A statistical approach to Machine Translation', *Computational Linguistics,* **16(2)**, 79–85.

Brown, P. F., J .Lai and R.L. Mercer (1991) 'Aligning sentences in parallel corpora', *Proceedings of the 29th Annual Meeting of the Association for Computational Linguistics (ACL-91),* Berkeley, California, pp. 169–176.

Carroll, J. and C. Grover (1989) 'The derivation of a large computational lexicon for English from LDOCE' in B. Boguraev and E. J. Briscoe (ed.), *Computational lexicography for natural language processing,* Longman, London, pp. 117–134.

Catizone, R., G. Russel and S. Warwick (1989) 'Deriving translation data from bilingual texts', *Proceedings of the First International Acquisition Workshop,* Detroit.

Church, K. (1993) '*char_align*: a program for aligning parallel texts at the character level', *Proceedings of the 31st Annual Meeting of the Association for Computational Linguistics (ACL-93),* Columbus, Ohio, pp. 1–8.

Copestake, A. (1990) 'An approach to building the hierarchical element of a lexical knowledge base from a machine readable dictionary', *Proceedings of the First International Workshop on Inheritance in Natural Language Processing,* Tilburg, The Netherlands, pp. 19–29. `http://www.cl.cam.ac.uk/Research/NL/acquilex/papers.html`

Copestake, A. (1995) 'Semantic transfer for Verbmobil', ACQUILEX II working paper 62 and Verbmobil report 93 `http://www.cl.cam.ac.uk/Research/NL/acquilex/papers.html`.

Copestake, A. and A. Sanfilippo (1993) 'Multilingual lexical representation', *Proceedings of the AAAI Spring Symposium on Building lexicons for machine translation,* Stanford, CA. `http://www.cl.cam.ac.uk/Research/NL/acquilex/papers.html`

Copestake, A., E.J. Briscoe, P. Vossen, A. Ageno, I. Castellon, F. Ribas, G. Rigau, H. Rodriguez, A. Samiotou (1995) 'Acquisition of lexical translation relations from MRDs', *Machine Translation,* **9(3–4)**, 183–219. `http://www.cl.cam.ac.uk/Research/NL/acquilex/papers.html`

Copestake, A., D. Flickinger, R. Malouf, S. Riehemann, I. Sag (1995) 'Translation using Minimal Recursion Semantics', *Proceedings of the The Sixth International Conference on Theoretical and Methodological Issues in Machine Translation (TMI95),* Leuven, Belgium. `http://www.cl.cam.ac.uk/Research/NL/acquilex/papers.html`

Dagan, I. and A. Itai (1994) 'Word-sense disambiguation using a second language monolingual corpus', *Computational Linguistics,* **20(4)**, 563–596.

Dorna, M. and M. Emele (1996) 'Efficient implementation of a semantic-based transfer approach', *Proceedings of the 12th European Conference on Artifical Intelligence (ECAI 96),* Budapest. `cmp-lg/9607010`

Dorna, M. and M. Emele (1996) 'Semantic-based transfer', *Proceedings of the International Conference on Computational Linguistics, Coling 96,* Copenhagen. `cmp-lg/9607009`

Dorr, B.J. (1994) 'Machine Translation divergences: a formal description and a proposed solution', *Computational Linguistics,* **20 (4)**, 597–633.

Dorr, B.J., J. Garman, A. Weinberg (1995) 'From syntactic encodings to thematic roles: building lexical entries for interlingual MT', *Machine Translation,* **9**, 221–250.

Estival, D., A. Ballim, G. Russell and S. Warwick (1990) 'A syntax and semantics for feature structure transfer', *Proceedings of the 3rd International Conference on theoretical and methodological issues in MT (TMI-90),* Austin,Texas, pp. 131–143.

Frank, A. and U. Reyle (1994) 'Principle based semantics for HPSG', Arbeitspapiere des Sonderforschungsbereichs 340, University of Stuttgart.

Fung, P. (1995) 'A pattern matching method for finding noun and proper noun translations from noisy parallel corpora', *Proceedings of the 33rd Annual Meeting of the Association for Computational Linguistics (ACL-95),* Cambridge, Mass.

Gale, W. and K. Church (1991) 'A program for aligning sentences in bilingual corpora', *Proceedings of the 29th Annual Meeting of the Association for Computational Linguistics (ACL-91),* Berkeley, California, pp. 177-184.

Grishman, R. & Sterling, J. (1992) 'Acquisition of selectional patterns', *Proceedings of the International Conference on Computational Linguistics, COLING-92,* Nantes, France, pp. 658–664.

Grishman, R., Macleod, C. & Meyers, A. (1994) 'Comlex syntax: building a computational lexicon', *Proceedings of the International Conference on Computational Linguistics, COLING-94,* Kyoto, Japan, pp. 268–272.

Helmreich, S., L. Guthrie and Y. Wilks (1993) 'The use of machine readable dictionaries in the Pangloss project', Paper presented at the AAAI Spring Symposium on Building Lexicons for Machine Translation, Stanford University.

Isabelle, P., M. Dymetman and E. Macklovitch (1988) 'CRITTER: a translation system for agricultural market reports', *Proceedings of the 12th International Conference on Computational Linguistics (COLING-88),* Budapest.

Johnston, M., B. Boguraev and J. Pustejovsky (1995) 'The acquisition and interpretation of complex nominals', *Proceedings of the AAAI Spring Symposium on representation and acquisition of lexical knowledge,* Stanford, CA.
`http://www.cs.brandeis.edu/~johnston/home.html`

Kaplan, R., K. Netter, J. Wedekind and A. Zaenen (1989) 'Translation by structural correspondences', *Proceedings of the 4th Conference of the European Chapter of the Association for Computational Linguistics (EACL-89),* Manchester, UK, pp. 272–281.

Kay, M. (1996) 'Chart generation', *Proceedings of the 34th Annual Meeting of the Association for Computational Linguistics (ACL-96),* Santa Cruz, CA, pp. 200–204.

Kay, M. and M. Röscheisen (1988) 'Text-translation alignment', Technical report P90-00143.

Kay, M., J.M. Gawron and P. Norvig (1994) *Verbmobil: A translation system for face-to-face dialog,* CSLI, Stanford.

Klavans, J. (1988) 'COMPLEX: A computational lexicon for natural language systems', *Proceedings of the 12th International Conference on Computational Linguistics (Coling-88),* Budapest, Hungary.

Klavans, J. and E. Tzoukermann (1996) 'Combining corpus and machine-readable dictionary data for building bilingual lexicons', *Machine Translation,* **10 (3–4)**, 1–34.

Landsbergen, J. (1987) 'Isomorphic grammars and their use in the ROSETTA translation system' in M. King (ed.), *Machine Translation Today: The State of the Art,* Edinburgh University Press, pp. 351–372.

Lonsdale, D., T. Mitamira and E. Nyberg (1995) 'Acquisition of large lexicons for practical knowledge-based MT', *Machine Translation,* **9**, 251–283.

Manning, C. (1993) 'Automatic acquisition of a large subcategorisation dictionary from corpora', *Proceedings of the 31st Annual Meeting of the Association for Computational Linguistics (ACL-93),* Columbus, Ohio.

Marti, M.A. and C. Soler (1994) 'Dealing with lexical mismatches', *Proceedings of the Euralex 94,* Amsterdam, The Netherlands.

Melamed, I. D. (1996) 'Automatic construction of clean broad-coverage translation lexicons', *Proceedings of the AMTA-96,* Montreal.

Neff, M. and M. McCord (1990) 'Acquiring lexical data from machine-readable dictionary resources for Machine Translation', *Proceedings of the Third International Conference on Theoretical and Methodological Issues in Machine Translation of Natural Languages (TMI-90),* Austin, Texas.

Neff, M., B. Blaser, J-M Lange, H. Lehmann and I. Dominguez (1993) 'Get it where you can: acquiring and maintaining bilingual lexicons for Machine Translation', Paper presented at the AAAI Spring Symposium on Building Lexicons for Machine Translation, Stanford University.

Nirenburg, S. and V. Raskin (1988) 'The Subworld Concept Lexicon and the Lexicon Management System', *Computational Linguistics,* **13(3-4)**, 276–289.

Nirenburg, S. and L. Levin (1992) 'Syntax-driven and ontology-driven lexical semantics' in J. Pustejovsky and S. Bergler (ed.), *Lexical semantics and knowledge representation: Proceedings of the first SIGLEX Workshop, Berkeley, 1991,* Springer-Verlag (Lecture notes in Artificial Intelligence), pp. 5–20.

Pollard, C. and I.A.Sag (1994) *Head-driven phrase structure grammar,* Chicago University Press, Chicago.

Popowich, F. (1994) 'Improving the efficiency of a generation algorithm for Shake and Bake Machine Translation using Head-Driven Phrase Structure Grammar', Technical report CMPT-TR 94-07, School of Computer Science, Simon Fraser University.

Poznanski, V., J.L. Beaven and P. Whitelock (1995) 'An efficient generation algorithm for lexicalist MT', *Proceedings of the 33rd Annual Meeting of the Association for Computational Linguistics (ACL-95),* Cambridge, Mass.

Poznanski, V. and A. Sanfilippo (1993) 'Detecting dependencies between semantic verb subclasses and subcategorization frames in text corpora', *Proceedings of the SIGLEX ACL Workshop on the Acquisition of Lexical Knowledge from Text,* Boguraev, B. & Pustejovsky, J. eds.

Rayner, M., and P Bouillon (1995) 'Hybrid Transfer in an English-French Spoken Language Translator', *Proceedings of the IA '95,* Montpellier, France.
`http://www.cam.sr.com/tr/crc056/paper.ps.Z`

Rayner, M., P Bouillon and D Carter (1995) 'Using Corpora to Develop Limited-Domain Speech Translation Systems', Translating and the Computer 17 (ASLIB).
`http://www.cam.sr.com/tr/crc059/paper.ps.Z`

Rayner, M., I Bretan, M Wiren, S Rydin and E Beshai (1996) 'Composition of transfer rules in a multi-lingual MT system', *Proceedings of the Workshop on Future Issues for Multilingual Text Processing,* Cairns, Australia.
`http://www.cam.sr.com/tr/crc063/paper.ps.Z`

Reichenbach, H. (1947) *Elements of Symbolic Logic,* Dover.

Resnik, P. (1995) 'Disambiguating noun groupings with respect to WordNet senses', *Proceedings of the 3rd workshop on very large corpora,* MIT, Cambridge, Mass.

Ribas, P. (1994) 'An experiment on learning appropriate selection restrictions from a parsed corpus', *Proceedings of the International Conference on Computational Linguistics, COLING-94,* Kyoto, Japan.

Sanfilippo, A., E. J. Briscoe, A. Copestake, M. A. Marti and A. Alonge (1992) 'Translation equivalence and lexicalization in the ACQUILEX LKB', *Proceedings of the 4th International Conference on Theoretical and Methodological Issues in Machine Translation (TMI-92),* Montreal, Canada.

Sanfilippo, A. and V. Poznanski (1992) 'The acquisition of lexical knowledge from combined machine-readable dictionary sources', *Proceedings of the 3rd Conference on Applied Natural Language Processing (ANLP-92),* Trento, Italy, pp. 80–88.

Shieber, S.M. (1993) 'The problem of logical form equivalence', *Computational Linguistics,* **19 (1)**, 179–190.

Smajda, F. (1993) 'Retrieving collocations from text: Xtract', *Computational Linguistics,* **19 (1)**, 143–177.

Smajda, F., K. McKeown and V. Hatzivassiloglou (1996) 'Translating collocations for bilingual lexicons: a statistical approach', *Computational Linguistics,* **22 (1)**, 1–38.

Sutcliffe, R. (1993) 'Using Distributed Patterns as Language Independent Lexical Representations', Paper presented at the AAAI Spring Symposium on Building Lexicons for Machine Translation, Stanford University.

Talmy, L. (1985) 'Lexicalization patterns: semantic structure in lexical form' in T. Shopen (ed.), *Language typology and syntactic description 3. Grammatical categories and the lexicon,* Cambridge University Press, pp. 57–149.

Trujillo, I.A. (1995) 'Lexicalist Machine Translation of Spatial Prepositions', PhD dissertation, University of Cambridge.

Ushioda, A., D. Evans, T. Gibson and A. Waibel (1993) 'The automatic acquisition of frequencies of verb subcategorization frames from tagged corpora' in Boguraev, B. & Pustejovsky, J. (ed.), *SIGLEX ACL Workshop on the Acquisition of Lexical Knowledge from Text,* Columbus, Ohio, pp. 95–106.

Vossen, P. (1992) 'The automatic construction of a knowledge base from dictionaries: a combination of techniques' in H. Tommola, K. Tarantola, T. Salmin Tolonen, J. Schop (ed.), *Euralex '92 Proceedings I-II,* Tampere, Finland.

Vossen, P. (1993) 'Extracting equivalence relations for a multilingual lexical knowledge base', Acquilex 2 working paper No. 14: University of Amsterdam.

Whitelock, P. (1992) 'Shake-and-bake translation', *Proceedings of the 14th International Conference on Computational Linguistics (COLING-92),* Nantes, France.

Wilks, Y., D. Fass, C-M. Guo, J. McDonald, T. Plate and B. Slator (1989) 'A tractable machine dictionary as a resource for computational semantics' in B. Boguraev and E.J. Briscoe (ed.), *Computational lexicography for natural language processing,* Longman, London, pp. 193–231.

Wu, D. (1990) 'Probabilistic unification-based integration of syntactic and semantic preferences for nominal compounds', *Proceedings of the 13th International Conference on Computational Linguistics (Coling90),* Helsinki, pp. 413–418.

Wu, D. and X. Xia (1995) 'Large-scale automatic extraction of an English-Chinese translation lexicon', *Machine Translation,* **9(3–4)**, 285–313.

Zernik, U. (ed.) (1992) *Lexical acquisition: using on-line resources to build a lexicon,* Lawrence Erlbaum, New Jersey.