Curriculum Vitae

Anoop Sarkar
School of Computing Science
8888 University Drive
Burnaby, BC V5A 1S6, Canada

Email: mailto:anoop@sfu.ca
Web: http://www.cs.sfu.ca/~anoop
Tel.: +1 (778)-782-2015

Contents

1 Summary of contributions

My area of research is in natural language processing (NLP) and computational linguistics (CL), in particular in the areas of statistical machine translation and statistical parsing of natural language.

Research impact

As of January 2017, I have published 68 journal articles, book chapters and peer-reviewed conference publications. My publications are largely in high quality conferences such as ACL, North American ACL, EMNLP, among others. The impact factor of conferences in my field is much higher than the journals.

I have been cited 2574 times, cited on average more than 200 times per year, and an h-index of 26 (as tracked by Google Scholar).

In 2013, I was awarded the prestigious **NSERC Discovery Accelerator Supplements** (DAS) award.

Professional activities

I was **local co-organizer** of the Annual Meeting of the Association for Computational Linguistics: ACL 2017 which is the premier international conference in Natural Language Processing which was held in Vancouver from July 30 to August 4, 2017.

I served as **Program Co-chair**, along with Joyce Chai at Michigan State, for the Meeting of the North American ACL: NAACL 2015 which is the premier North American conference in my field.

I have served as **Program Co-chair**, along with Mirella Lapata from Edinburgh University, of the 14th Conference on Computational Natural Language Learning (CoNLL) in 2010 (a very good second tier conference in NLP) and **Program Co-Chair** for the 9th International Workshop on Tree Adjoining Grammars in 2008.

I served as faculty advisor to the North American ACL Student Session in 2009 and Area Chair for Parsing in EMNLP in 2007 and Area Chair for Machine Learning for ACL in 2006.

External recognition

I am well known in the NLP research community. I was elected to two consecutive terms (four years in all) as **Secretary of the North American ACL** (NAACL) Board which organizes the annual conference in NLP in North America.

Service to SFU

In 2012-2013, I was responsible for the university approval and the launch of the Big Data Professional M.Sc. program, the newest addition to the SFU Computing graduate program. I served as the **Program Director** of the Big Data M.Sc. program for two years from 2013 to 2015.

From 2012 to 2014, I have served on the SFU Senate Undergraduate Awards Committee (SUAAC).

Graduate supervision

In the last two years (2016 and 2017) I have graduated 2 Ph.D. and 6 M.Sc. students. Both Ph.D. students have been placed in good positions: one as research scientist in Amazon Inc. and the other as tenure track faculty at the University of the Fraser Valley (UFV).

In all, I have graduated 6 Ph.D. and 21 M.Sc. students and supervised 2 post-doctoral fellows.

All four of my graduated Ph.D. students have placed in excellent jobs after their graduation. Reza Haffari is a tenure-track faculty in Monash University, Australia. Yudong Liu is a tenure-track faculty in Western Washington University, USA. Baskaran Sankaran is a research staff member at IBM Research at Yorktown Heights, NY, USA.

Majid Razmara is CTO in his own start-up company called MetaOptima, Inc. which made the British Columbia Digital Health Emerging Rocket List in 2017.

Some of my M.Sc. students have gone on to Ph.D. positions in places like EPFL, Zürich and University of Texas at Dallas. Other M.Sc. students are now employed in companies such as Accenture, Arista, Microsoft, Nuance, Lightspeed, SAP Business Objects, and Twitter.

Undergraduate teaching

I was awarded the **Excellence in Undergraduate Teaching** Award in 2007 by the SFU Undergraduate Computing Science Student Society (CSSS).

I have taught 31 courses at SFU. My average course evaluation score is 3.47 and the median is 3.47. My average instructor evaluation score is 3.63 and the median is 3.63. The scores are in the range $[0, 4]$.

Industrial grants

I received the **IBM Faculty Award** in 2008 and a **Google Faculty Award** in 2009 and again in 2014. I received an **NSERC CRD** (Collaborative Research and Development) grant with Boeing Inc. on text visualization research.

I have had several interactions with industry on applications of natural language processing, and technology transfer from NLP research to NLP industrial applications. For instance, in 2011 I worked with NLP Technologies – a Canadian company that provides translation services for Quebec courts, and from 2012 to 2014 with McKesson Inc. on NLP for radiology reports.

Research software

In 2012, we published the paper: Kriya - an end-to-end hierarchical phrase-based machine translation (MT) system. B. Sankaran, M. Razmara and A. Sarkar. *The Prague Bulletin of Mathematical Linguistics*. March 2012. This paper describes the statistical machine translation system developed in my lab. To our knowledge this is the only end-to-end **open-source statistical MT system** developed at a Canadian university.

We are one of the few Canadian universities to participate in the WMT-2012 shared task. In this international research competition our system outperformed a comparable system from Johns Hopkins university.

In 2012, we also collaborated with the Language Technology group (LTRC) from NRC to submit a joint submission to the 2012 NIST translation task for Arabic-English using NRC's Portage MT system, which is the most prestigious MT competition in North America.

2 Background

2.1 Education

2002 Ph.D. in Computer Science
Department of Computer and Information Sciences,
University of Pennsylvania, USA
Thesis: Combining Labeled and Unlabeled Data in Statistical Natural Language Parsing
Advisor: Prof. Aravind Joshi
My thesis was among the first to introduce semi-supervised learning techniques to the area of natural language parsing. It also introduced novel statistical parsing algorithms for Tree-adjoining grammars.

1991 B.Eng. in Computer Science
University of Poona, India
First Class with Distinction (ranked third overall)

2.2 Employment History

09/2015 - current Professor
School of Computing Science, Simon Fraser University

09/2009 - 09/2015 Associate Professor
School of Computing Science, Simon Fraser University

09/2002 - 09/2009 Assistant Professor
School of Computing Science, Simon Fraser University

06/2002 - 09/2002 Postdoctoral Fellow
Institute for Research in Cognitive Science (IRCS), University of Pennsylvania

08/1995 - 06/2002 Research Assistant
Department of Computer and Information Sciences, University of Pennsylvania

08/1991 - 08/1993 Research Associate
Center for Development in Advanced Computing (C-DAC), Poona, India

2.3 Awards and Research Grants

- Google Faculty Research Award, 2014-2015.
- NSERC Discovery Accelerator Supplements (DAS) Program, 2012-2015.
- NSERC Discovery Grants Program, 2003-2007, 2007-2012, 2012-2017.
- NSERC Strategic Project Grants, co-investigator on the GreenPhones project with Sasha Fedorova and Arrvindh Shriraman, 2012-2016.
- NSERC Engage Grant with McKesson Inc., 2012.
- NSERC Collaborative Research and Development (CRD) Grants, 2011-2012.
- Google Faculty Research Award, 2009-2010.
- IBM Faculty Award, 2008-2009.
- NSERC Research Tools and Instruments (RTI) Grants Program, 2008-2009.
- MITACS Accelerate Internship Grant, 2003, 2008, 2009, 2010, 2011, 2013
- MITACS Globalink Internship, 2009, 2010, 2011, 2012, 2014
- SFU President's Research Grant (SFU), 2002-2003.

3 Most Significant Contributions to Research and/or to Practical Applications

My papers below have had a substantial research impact:

1. F. Och, D. Gildea, S. Khudanpur, A. Sarkar, et. al. A Smorgasbord of Features for Statistical Machine Translation. HLT-NAACL 2004. **cited 206 times** (first paper on syntax for MT)
2. L. Shen, A. Sarkar and F. Och. Discriminative Reranking for Machine Translation. HLT-NAACL 2004. **cited 121 times** (introduced discriminative ranking for MT)
3. A. Sarkar. Applying Co-Training Methods to Statistical Parsing. NAACL 2001. **cited 116 times** (first paper on Co-Training for statistical parsing)
4. D. Song and A. Sarkar. Training Global Linear Models for Chinese Word Segmentation. Canadian AI 2009. **Best paper award**. (12 pages) (Exponentiated Gradient applied to NLP)
5. Kriya - an end-to-end hierarchical phrase-based MT system. B. Sankaran, M. Razmara and A. Sarkar. *The Prague Bulletin of Mathematical Linguistics*. March 2012. (16 pages) (*Kriya* is an end-to-end hierarchical phrase-based machine translation system developed in my lab, in collaboration with my students. To our knowledge this is the only end-to-end **statistical MT system developed at a Canadian university**).

3.1 Publications

Names in **bold** face are my students.

Best Paper Awards

1. **Best paper award**: **D. Song** and A. Sarkar. Training Global Linear Models for Chinese Word Segmentation. In *Proceedings of the 22nd Canadian Conference on Artificial Intelligence, Canadian AI 2009*. Kelowna, BC. May 25-27, 2009.
2. **Best Presentation Award**: **Y. Liu**, **G. Haffari** and Anoop Sarkar. Latent SVMs for Semantic Role Labeling using LTAG Derivation Trees. (Full Paper). The Pacific Northwest Regional NLP Workshop (NW-NLP 2010). April 23, 2010.

Journal Papers and Book Chapters

1. Joint Prediction of Word Alignment with Alignment Types. **Anahita Mansouri, Te Bu**, Anoop Sarkar. *Transactions of the ACL*. volume 5, 2017.
2. A Python-based Interface for Wide Coverage Lexicalized Tree-adjoining Grammars. **Ziqi Wang, Haotian Zhang**, Anoop Sarkar. *The Prague Bulletin of Mathematical Linguistics* 103. April 2015.
3. Kriya - an end-to-end hierarchical phrase-based MT system. Baskaran Sankaran, Majid Razmara and Anoop Sarkar. *The Prague Bulletin of Mathematical Linguistics*. March 2012.
4. A. Sarkar. Book Review of *Parsing Schemata for Practical Text Analysis* by Carlos Gómez Rodríguez. Computational Linguistics (to appear). 2011. doi: 10.1162/COLI_r_00081
5. N. Ueffing, **G. Haffari** and A. Sarkar. Transductive learning for statistical machine translation. In *Machine Translation*, DOI 10.1007/s10590-008-9036-3, Springer. 2008.
6. K. Forbes, E. Miltsakaki, R. Prasad, A. Sarkar, A. Joshi and B. Webber. D-LTAG System: Discourse parsing with a Lexicalized Tree Adjoining Grammar. *Journal of Logic, Language and Information*: Volume 12, Issue 3, pp. 261-279, Springer. 2003.
7. S. Wintner and A. Sarkar. A note on typing feature structures. *Computational Linguistics*. 28(3):389-397. MIT Press. 2002.
8. A. Sarkar. Syntax and Parsing. In *Multilingual Natural Language Applications: From Theory to Practice*. Daniel M. Bikel and Imed Zitouni, eds. Prentice Hall. 2012.

9. N. Ueffing, **G. Haffari** and A. Sarkar. Semi-supervised learning for machine translation. Book chapter in *Learning Machine Translation*, Eds. C. Goutte, N. Cancedda, M. Dymetman and G. Foster. MIT Press. 2008.
10. A. Sarkar. Combining SuperTagging with Lexicalized Tree-Adjoining Grammar parsing. Book chapter in *Complexity of Lexical Descriptions and its Relevance to Natural Language Processing: A Supertagging Approach*, Eds. S. Bangalore and A. Joshi. 2007.
11. A. Sarkar and A. Joshi. Tree-Adjoining Grammars and its application to statistical parsing. In *Data-oriented parsing*. Eds. R. Bod, R. Scha and K. Sima'an, CSLI Publications, Stanford, 2003.
12. C. Doran, B. A. Hockey, A. Sarkar, B. Srinivas and F. Xia. Evolution of the XTAG System. In *Tree Adjoining Grammars: Formal, Computational and Linguistic Aspects*. pp. 371-404. Eds. A. Abeille and O. Rambow, CSLI Publications, Stanford, 2000.
13. A. Sarkar. The conflict between future tense and modality: the case of will in English. In *Penn Working Papers in Linguistics* volume 5, number 2, pp. 91-117, 1998.

Refereed Conference and Workshop Papers

14. Lexicalized Reordering for Left-to-Right Hierarchical Phrase-based Translation. **Maryam Siahbani** and Anoop Sarkar. Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 2, Short Papers, 2017.
15. Evaluating the Value of Lensing Wikipedia During the Information Seeking Process. Hoeber, Orland and Sarkar, Anoop and **Vacariu, Andrei** and Whitney, Max and Gaikwad, Manali and Kaur, Gursimran. Proceedings of the ACM SIGIR Conference on Human Information Interaction and Retrieval (CHIIR), 2017.
16. Graph-based Semi-supervised Gene Mention Tagging. **Golnar Sheikhshab**, Elizabeth Starks, Aly Karsan, Anoop Sarkar and Inanc Birol. 15th Workshop on Biomedical Natural Language Processing BioNLP, Berlin, Germany. August 12, 2016.
17. Learning Segmentations that Balance Latency versus Quality in Spoken Language Translation **Hassan S. Shavarani**, **Maryam Siahbani**, **Ramtin Mehdizadeh Seraj**, Anoop Sarkar. Proceedings of the International Workshop on Spoken Language Translation (IWSLT), Da Nang, Vietnam. 2015.
18. Improving Statistical Machine Translation with a Multilingual Paraphrase Database. **Ramtin Mehdizadeh Seraj**, **Maryam Siahbani** and Anoop Sarkar. Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing. Lisbon, Portugal. September 1721, 2015.
19. Non-uniform Stochastic Average Gradient Method for Training Conditional Random Fields. **Mark Schmidt**, Reza Babanezhad, Mohamed Osama Ahemd, Aaron Defazio, **Ann Clifton** and Anoop Sarkar. In Proceedings of the 18th International Conference on Artificial Intelligence and Statistics (AISTATS). May 9-12, 2015. San Diego, California, USA. JMLR W & CP volume 38.
20. Two Improvements to Left-to-Right Decoding for Hierarchical Phrase-based Machine Translation. **Maryam Siahbani** and Anoop Sarkar. In Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP 2014). October 25-29, 2014. Doha, Qatar.
21. Expressive Hierarchical Rule Extraction for Left-to-Right Translation. **Maryam Siahbani** and Anoop Sarkar. In Proceedings of the 11th Biennial Conference of the Association for Machine Translation in the Americas (AMTA-2014). October 22-26, 2014. Vancouver, Canada.
22. Bayesian Iterative-cascade Framework for Hierarchical Phrase-based Translation. **Baskaran Sankaran** and Anoop Sarkar. In Proceedings of the 11th Biennial Conference of the Association for Machine Translation in the Americas (AMTA-2014). October 22-26, 2014. Vancouver, Canada.
23. Pivot-based Triangulation for Low-Resource Languages. **Rohit Dholakia** and Anoop Sarkar. In Proceedings of the 11th Biennial Conference of the Association for Machine Translation in the Americas (AMTA-2014). October 22-26, 2014. Vancouver, Canada.

24. Incremental Translation using a Hierarchical Phrase-based Translation System. **Maryam Siahbani, Ramtin Mehdizadeh Seraj, Baskaran Sankaran** and Anoop Sarkar. In Proceedings of the 2014 IEEE Spoken Language Technology Workshop (SLT 2014). December 7-10, 2014. Nevada, USA.
25. Ensemble Triangulation for Statistical Machine Translation. **Majid Razmara** and Anoop Sarkar. In the 6th International Joint Conference on Natural Language Processing: IJCNLP 2013. Oct 14-18, 2013. Nagoya, Japan.
26. Scalable Variational Inference for Extracting Hierarchical Phrase-based Translation Rules. **Baskaran Sankaran, Gholamreza Haffari** and Anoop Sarkar. In the 6th International Joint Conference on Natural Language Processing: IJCNLP 2013. Oct 14-18, 2013. Nagoya, Japan.
27. An Online Algorithm for Learning over Constrained Latent Representations using Multiple Views. **Ann Clifton, Max Whitney** and Anoop Sarkar. In the 6th International Joint Conference on Natural Language Processing: IJCNLP 2013. Oct 14-18, 2013. Nagoya, Japan.
28. Knowledge Base Population and Visualization Using an Ontology based on Semantic Roles. **Maryam Siahbani, Ravikiran Vadlapudi, Max Whitney** and Anoop Sarkar. In Proceedings of Automated Knowledge Base Construction (AKBC) 2013: The 3rd Workshop on Knowledge Extraction at CIKM 2013 in San Francisco, October 27-28, 2013.
29. Efficient Left-to-Right Hierarchical Phrase-based Translation with Improved Reordering. **Maryam Siahbani, Baskaran Sankaran** and Anoop Sarkar. In Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP 2013). Oct 18-21, 2013. Seattle, USA.
30. **Majid Razmara** and Anoop Sarkar. Stacking for Statistical Machine Translation. Short Paper. In Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (ACL 2013). Aug 4-9, 2013. Sofia, Bulgaria.
31. **Majid Razmara**, **Maryam Siahbani**, Reza Haffari and Anoop Sarkar. Graph Propagation for Paraphrasing Out-of-Vocabulary Words in Statistical Machine Translation. In Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (ACL 2013). Aug 4-9, 2013. Sofia, Bulgaria.
32. **Baskaran Sankaran**, Anoop Sarkar and Kevin Duh. Multi-Metric Optimization Using Ensemble Tuning. In Proceedings of the 2013 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL-HLT 2013). June 9-14, 2013. Atlanta, USA.
33. **Baskaran Sankaran**, Anoop Sarkar and Kevin Duh. Multi-objective Optimization Problems in Statistical Machine Translation. Poster at the 22nd International Conference on Multiple Criteria Decision Making (MCDM 2013). June 17-21, 2013. Málaga, Spain.
34. **M. Whitney** and A. Sarkar. Bootstrapping via Graph Propagation. In Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics (ACL-2012). July 9-11, 2012. Jeju Island. R. of Korea.
35. **M. Razmara**, G. Foster, **B. Sankaran** and A. Sarkar. Mixing Multiple Translation Models in Statistical Machine Translation. In Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics (ACL-2012). July 9-11, 2012. Jeju Island. R. of Korea.
36. **B. Sankaran** and A. Sarkar. Improved Reordering for Shallow-n Grammar based Hierarchical Phrase-based Translation. In Proceedings of the 2012 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL HLT 2012). Montreal, Quebec, Canada. June 3-8, 2012.
37. **B. Sankaran**, **M. Razmara**, A. Farzindar, W. Khreich, F. Popowich and A. Sarkar. Domain Adaptation Techniques for Machine Translation and their Evaluation in a Real-World Setting. In Proceedings of the 25th Canadian Conference on Artificial Intelligence. York University, Ontario. May 28-30, 2012.
38. **Baskaran Sankaran,** Gholamreza Haffari, and Anoop Sarkar. Compact Rule Extraction for Hierarchical Phrase-based Translation. In the 10th biennial conference of the Association for Machine Translation in the Americas (AMTA). San Diego, CA. 2012.

39. **Ravikiran Vadlapudi**, **Maryam Siabhani**, Anoop Sarkar and John Dill. LensingWikipedia: Parsing Text for the Interactive Visualization of Human History (poster). IEEE visWeek 2012 Posters, Seattle, WA, USA. Oct 2012.

40. **M. Razmara, B. Sankaran, A. Clifton** and Anoop Sarkar. Kriya - The SFU System for the Translation Task at WMT-12. In NAACL 2012 Seventh Workshop on Statistical Machine Translation, June 7-8, 2012. Montreal, Quebec, Canada.

41. **Ann Clifton** and Anoop Sarkar. Combining Morpheme-based Machine Translation with Post-processing Morpheme Prediction. In Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies. Portland, OR, USA. June 19-24, 2011.

42. Gholamreza Haffari, **Marzieh Razavi** and Anoop Sarkar. An Ensemble Model that Combines Syntactic and Semantic Clustering for Discriminative Dependency Parsing. Short Paper (poster). In Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies. Portland, OR, USA. June 19-24, 2011.

43. **B. Sankaran**, **G. Haffari** and Anoop Sarkar. Bayesian Extraction of Minimal SCFG Rules for Hierarchical Phrase-based Translation. Sixth Workshop on Statistical Machine Translation: WMT-2011. Edinburgh, UK July 30-31, 2011.

44. **B. Sankaran**, **A. Grewal** and A. Sarkar. Incremental Decoding for Phrase-Based Statistical Machine Translation. In Proceedings of the Joint Fifth Workshop on Statistical Machine Translation and MetricsMATR. Uppsala, Sweden. July 15, 2010.

45. **D. Song** and A. Sarkar. Training Global Linear Models for Chinese Word Segmentation. In *Proceedings of the 22nd Canadian Conference on Artificial Intelligence, Canadian AI 2009*. Kelowna, BC. May 25-27, 2009. **Best paper award**.

46. **G. Haffari**, M. Roy and A. Sarkar. Active Learning for Statistical Phrase-based Machine Translation. In *Proceedings of the annual meeting of the North American Chapter of the Association for Computational Linguistics - Human Language Technologies (NAACL HLT)*. Boulder, Colorado. May 31-June 5, 2009.

47. **G. Haffari** and A. Sarkar. Active Learning for Multilingual Statistical Machine Translation. In *Proceedings of the 47th annual meeting of the Association for Computational Linguistics and the 4th International Joint Conference on Natural Language Processing of the Asian Federation of Natural Language Processing (ACL-IJCNLP 2009)*. Singapore, August 2-7, 2009.

48. **G. Haffari** and A. Sarkar. Homotopy-based Semi-Supervised Hidden Markov Models for Sequence Labeling. In *Proc. of the 22nd International Conference on Computational Linguistics: COLING 2008*. Manchester, 18-22 August, 2008.

49. **G. Melli**, M. Ester and A. Sarkar. Recognition of Multi-sentence n-ary Subcellular Localization Mentions in Biomedical Abstracts. In *Proceedings of the 2nd International Symposium on Languages in Biology and Medicine (LBM) 2007*. Singapore, Dec 6-7, 2007.

50. **G. Haffari** and A. Sarkar. Analysis of semi-supervised learning with the Yarowsky algorithm. In *Proc. of the 23rd Conf. on Uncertainty in Artificial Intelligence*, UAI 2007. Vancouver, BC. July 19-22, 2007.

51. N. Ueffing, **G. Haffari** and A. Sarkar. Transductive learning for statistical machine translation. In *Proc. of the Annual Conf. of the Assoc. for Computational Linguistics*, ACL 2007, Prague, Czech Republic. June 25-27, 2007.

52. **Y. Liu** and A. Sarkar. Experimental evaluation of LTAG-based features for Semantic Role Labeling. In *Proc. of the Conf. on Empirical Methods in Natural Language Processing and the Conf. on Computational Natural Language Learning*, EMNLP-CoNLL 2007. Prague, Czech Republic. June 28-30, 2007.

53. **Z. Shi**, **G. Melli**, Y. Wang, **Y. Liu**, B. Gu, **M. K. Kashani**, A. Sarkar and F. Popowich. Question answering summarization of multiple biomedical documents. In *Proc. of the 20th Canadian Conf. on Artificial Intelligence*, Canadian AI 2007, Montreal, QC. May 28-30, 2007.

54. **Z. Shi**, A. Sarkar and F. Popowich. Simultaneous identification of biomedical named-entity and functional relations using statistical parsing techniques. In *Proc. of the Annual Conf. of the North American Chapter of the Assoc. for Computational Linguistics*, NAACL-HLT 2007, short paper, Rochester, USA. April 22-27, 2007.

55. **J. Birke** and A. Sarkar. A clustering approach for the nearly unsupervised recognition of non-literal language. In *Proc. of the 11th Conf. of the European Chapter of the Assoc. for Computational Linguistics*, EACL-2006. Trento, Italy. April 3-7, 2006.

56. **Z. Shi** and A. Sarkar. Intimate learning: A novel approach for combining labeled and unlabeled data. In Poster Track, *Nineteenth International Joint Conf. on Artificial Intelligence*: IJCAI-05. Edinburgh, UK. August 2-5, 2005.

57. **H. Shen** and A. Sarkar. Voting between multiple data representations for text chunking. In *Proc. of the Eighteenth Meeting of the Canadian Society for Computational Intelligence*, Canadian AI 2005. Victoria, BC, Canada. May 9-11, 2005. In Advances in Artificial Intelligence, Lecture Notes in Artificial Intelligence 3501, eds. Balász Kégl and Guy Lapalme. Springer.

58. L. Shen, A. Sarkar and F. Och. Discriminative re-ranking for machine translation. In the *Human Language Tech. Conf. and the 5th Meeting of the North American Assoc. for Computational Linguistics*: HLT-NAACL 2004. Boston, USA. May 2-7, 2004.

59. F. Och, D. Gildea, S. Khudanpur, A. Sarkar, K. Yamada, A. Fraser, S. Kumar, L. Shen, D. Smith, K. Eng, V. Jain, Z. Jin, D. Radev. A smorgasbord of features for statistical machine translation. In the *Human Language Tech. Conf. and the 5th Meeting of the North American Assoc. for Computational Linguistics*: HLT-NAACL 2004. Boston, USA. May 2-7, 2004.

60. L. Shen, A. Sarkar and A. Joshi. Using LTAG-based features in parse re-ranking. In the *2003 Conf. on Empirical Methods in Natural Language Processing*. Sapporo, Japan. July 11-12, 2003.

61. M. Steedman, R. Hwa, Stephen Clark, M. Osborne, A. Sarkar, J. Hockenmaier, P. Ruhlen, S. Baker, J. Crim. Example selection for bootstrapping statistical parsers. In the *Human Language Tech. Conf. and the 4th Meeting of the North American Assoc. for Computational Linguistics*: HLT-NAACL 2003. Edmonton, AB. May 27-June 1, 2003.

62. M. Steedman, M. Osborne, A. Sarkar, S. Clark, R. Hwa, J. Hockenmaier, P. Ruhlen, S. Baker, J. Crim. Bootstrapping statistical parsers from small data-sets. In *Proc. of the 11th Conf. of the European Assoc. for Computational Linguistics*: EACL 2003. Budapest, Hungary. April 12-17, 2003.

63. A. Sarkar and W. Tripasai. Learning verb argument structure from minimally annotated corpora. In *Proc. of the 18th International Conf. on Computational Linguistics: COLING 2002*. Taipei, Taiwan. August 2002.

64. A. Sarkar. Applying co-training methods to statistical parsing. In *Proc. of the 2nd Meeting of the North American Chapter of the Assoc. for Computational Linguistics: NAACL 2001*. pp. 175-182. Pittsburgh, USA, June 2001.

65. A. Sarkar and D. Zeman. Automatic Extraction of Subcategorization Frames for Czech. In *Proc. of the 18th International Conf. on Computational Linguistics: COLING 2000*. pp. 691-698. Saarbrücken, Germany, August 2000.

66. D. Zeman and A. Sarkar. Learning Verb Subcategorization from Corpora: Counting Frame Subsets. In *2nd International Conf. on Language Resources and Evaluation: LREC 2000*. pp. 227-233. Athens, Greece. May 31 - June 2, 2000.

67. A. Sarkar. Conditions on Consistency of Probabilistic Tree Adjoining Grammars. In *Proc. of the 36th Annual Meeting of the Assoc. for Computational Linguistics and 17th International Conf. on Computational Linguistics: COLING-ACL 1998*. pp. 1164-1170. Montreal, Quebec, 1998.

68. M.-J. Nederhof, A. Sarkar and G. Satta. Prefix Probabilities from Probabilistic Tree Adjoining Grammars. In *Proc. of the 36th Annual Meeting of the Assoc. for Computational Linguistics and 17th International Conf. on Computational Linguistics: COLING-ACL 1998*. pp. 953-959. Montreal, Quebec, 1998.

69. A. Sarkar. Separating Dependency from Constituency in a Tree Rewriting System. In *Proc. of the Fifth Meeting on Mathematics of Language*. pp. 153-160. Saarbrücken, Germany, August 1997.
70. A. Sarkar and A. Joshi. Coordination in Tree Adjoining Grammars: Formalization and Implementation. In *Proc. of 16th International Conf. on Computational Linguistics: COLING 1996*. pp. 610-615. Copenhagen, Denmark, 1996.
71. A. Sarkar. Incremental Parser Generation for Tree Adjoining Grammars. In *Proc. of the 34th Meeting of the Assoc. for Computational Linguistics: ACL 1996, Student Session*. pp. 375-376. University of California, Santa Cruz, June 1996.
72. A. Sarkar. Extending Kimmo's Two-Level Model. In *Proc. of the 31st Meeting of the ACL, Student Session*. pp. 304-306. Columbus, Ohio. June 1993.

Tutorials

73. Natural Language Processing and Text Visualization. A. Sarkar and **M. Siahbani**. Tutorial at the Canadian Visual Analytics Summer School: CANVAS 2013. July 15-19, Halifax, Nova Scotia, Canada.
74. Tutorial on Tree-adjoining Grammars. Anoop Sarkar. In 6th International Workshop Weighted Automata: Theory and Applications WATA 2012, May 29 - June 2, 2012. Dresden, Germany.
75. Tutorial on inductive semi-supervised learning methods: with applicability to natural language processing. A. Sarkar and **G. Haffari**. Tutorial at the *Human Language Tech. Conf. - North American chapter of the Assoc. for Computational Linguistics* annual meeting (HLT-NAACL) 2006. New York City, USA. June 4, 2006. *(peer reviewed)*
76. Parsing with Tree-adjoining Grammars. A. Sarkar. Tutorial at the *ACL/HCSNet Advanced Program in Natural Language Processing*, Melbourne, Australia. July 13-14, 2006. *(invited)*
77. Chunking and Statistical Parsing. A. Sarkar. Tutorial at a two-week course on corpus-based NLP at Anna University, Chennai, India. December, 2001. *(invited)*

Non-refereed Workshop Publications

78. **D. Song** and A. Sarkar. Training a perceptron with global and local features for Chinese word segmentation. Bake-off short paper. Proc. of the Sixth SIGHAN Workshop on Chinese Language Processing, IJCNLP 2008 workshop, Hyderabad, India. January 11-12, 2008.
79. **D. Song** and A. Sarkar. Voting between dictionary-based and sub-word tagging models for Chinese word segmentation. Bake-off short paper. Proc. of the *Fifth SIGHAN Workshop on Chinese Language Processing*, COLING-ACL 2006 workshop, Sydney, Australia. July 22-23, 2006.
80. **G. Melli**, **Z. Shi**, Y. Wang, **Y. Liu**, A. Sarkar and F. Popowich. Description of SQUASH, the SFU Question Answering Summary Handler for the DUC-2006 summarization task. In *Proc. of the Document Understanding Conf. 2006* (DUC-2006), New York City, USA, June 8-9, 2006.
81. **Z. Shi**, B. Gu, F. Popowich and A. Sarkar. Synonym-based query expansion and Boosting-based re-ranking: A two-phase approach for genomic information retrieval. In Proc. of the Fourteenth Text REtrieval Conf. (TREC 2005). NIST, Gaithersburg, USA, November 15-18, 2005.
82. **G. Melli**, Y. Wang, **Y. Liu**, **M. Kashani**, **Z. Shi**, B. Gu, A. Sarkar and F. Popowich. Description of SQUASH, the SFU Question Answering Summary Handler for the DUC-2005 summarization task. *Proc. of the Document Understanding Conf. 2005* (DUC-2005), Vancouver, B.C., October 2005.
83. B. Baldwin, J.C. Reynar, M. Collins, J. Eisner, A. Ratnaparkhi, J. Rosenzweig, A. Sarkar and B. Srinivas. Description of the University of Pennsylvania entry in the MUC-6 competition. *Proc. Sixth Message Understanding Conf.* (MUC-6), Maryland, October 1995.

Technical Reports

84. Experimental evaluation of LTAG-based features for semantic role labeling. **Y. Liu** and A. Sarkar. Technical Report TR 2007-03, School of Computing Science, Simon Fraser University, Burnaby, BC, Canada, April 2007. (longer version of our EMNLP-CoNLL 2007 paper)
85. Analysis of semi-supervised learning with the Yarowsky algorithm. **G. Haffari** and A. Sarkar. Technical Report TR 2007-07, School of Computing Science, Simon Fraser University, Burnaby, BC, Canada, March 2007. (longer version of our UAI 2007 paper)
86. F. Och, D. Gildea, S. Khudanpur, A. Sarkar, K. Yamada, A. Fraser, S. Kumar, L. Shen, D. Smith, K. Eng, V. Jain, Z. Jin, D. Radev. CLSP Final Report of the Johns Hopkins Summer Workshop 2003: Syntax for Statistical Machine Translation. 2003.
87. M. Steedman, S. Clark, R. Hwa, M. Osborne, A. Sarkar, J. Hockenmaier, P. Ruhlen, S. Baker, J. Crim. CLSP Final Report of the Johns Hopkins Summer Workshop 2002: Semi-Supervised Training for Statistical Parsing. 2002.
88. A Lexicalized Tree Adjoining Grammar for English. XTAG Research Group. Technical Report IRCS-01-03, IRCS, University of Pennsylvania, 2001.

3.2 Research Experience

89. 2012, WMT 2012 Shared Task
    We are one of the few Canadian universities to participate in the WMT-2012 shared task. In this international research competition our system outperformed a comparable system from Johns Hopkins university.
90. 2012, NIST OpenMT Competition
    In collaboration with the Language Technology group (LTRC) from NRC we made a joint submission to the 2012 NIST translation task for Arabic-English using NRC's Portage MT system, which is the most prestigious MT competition in North America.
91. Dec–Jun 2006, Document Understanding Conference (DUC) 2006
    Team Leader, Question Answering Combined with Multi-Document Summarization.
    The task was similar to the DUC 2005 effort (see below).
92. Jan–Jun 2005, Document Understanding Conference (DUC) 2005
    Team Leader, Question Answering Combined with Multi-Document Summarization.
    The task involved providing a 250 word summary answer to a provided question. The answer was produced from a small set of documents (20-50 per question). NIST ran the evaluation effort for all competing systems.
93. Jul–Aug 2003, Johns Hopkins Summer Workshop 2003
    Member, Syntax for Statistical Machine Translation.
    Along with a team of faculty members, graduate and undergraduate students, I explored statistical machine translation (MT) models that used various kinds of information from statistical parsers. The task was to find better translations within an n-best list produced by a baseline statistical MT system (we worked on the NIST 02/03 large data track for Chinese to English MT).
94. Jun–Jul 2002, Johns Hopkins Summer Workshop 2002
    Member, Weakly Supervised Learning for Wide Coverage Parsing.
    Along with Prof. Mark Steedman (Univ. of Edinburgh), I proposed a topic for inclusion in the 2002 workshop, and our proposal was funded as one of four out of twelve candidates. During the workshop, along with a team of faculty members, graduate and undergraduate students, I conducted experiments in improving accuracy of statistical parsing when faced with limited amounts of training data.
95. 1995–2002, The XTAG Project (`http://www.cis.upenn.edu/~xtag`)
    Lead developer of an open-source parser for a wide-coverage linguistic grammar of English.
    The XTAG project is an academic research project directed by Prof. Aravind Joshi. The project has built a

wide-coverage lexicalized Tree Adjoining Grammar (TAG) and associated parser for English which handles naturally occurring text. It has been used for both recognition and generation. I designed a parsing algorithm for TAG and implemented a parser which showed significant speedup over the previous implementation.

96. Aug–Oct 1995, Penn MUC-6 Team
Team Member
Designed and implemented modules for the Univ. of Penn. Co-reference Resolution System that stood third at the Sixth Message Understanding Conference in the co-reference task.

97. 1991–1993, Centre for Development of Advanced Computing (CDAC)
University of Poona Campus, Pune 411 007, India
A research lab run by the Dept. of Electronics, Govt. of India.
Research Associate (Natural Language Group)
Implemented a morphological analyzer which extended Kimmo's two-level model. Also worked on the implementation in Prolog of a tutoring system.

3.3 Invited Talks and Presentations

Invited Talks

98. *Machine Reading of Natural Language and Interactive Visualization*.
Invited talk. Michigan State University, Lansing MI. Nov 3, 2017.

99. *The challenge of simultaneous speech translation*.
Keynote talk. PACLIC 2017, Seoul, Korea. Oct 30, 2016.

100. *Segmentation and stream decoding for simultaneous speech translation*.
Invited talk. University of Washington, ECE Department. Feb 16, 2016.

101. *Interactive visualization of facts extracted from natural language*.
Invited talk. University of Colorado Boulder, CO. Feb 6, 2015.

102. *Pareto optimal machine translation: multiple objectives are better than one*.
IRMACS Interdisciplinary Colloquium - Fall 2013. SFU Burnaby campus, Canada. September 20, 2013.

103. *Ensemble decoding for statistical machine translation*.
Invited Talk. Carnegie Mellon University, Pittsburgh, PA. Dec 7, 2012.

104. *Bootstrapping via Graph Propagation*. Invited Talk.
University of Maryland, College Park, MD. Oct 23, 2012.
University of British Columbia, Vancouver, BC. May 14, 2012.

105. *Four methods for morpheme-based machine translation*.
Anoop Sarkar. Invited Talk at KAIST, July 12, 2012. Daejeon, Korea.

106. *Text visualization and NLP*. Anoop Sarkar.
Invited Talk at CANVAS, July 25, 2012. SFU Burnaby campus, Canada.

107. *Combining Morpheme-based machine translation with post-processing morpheme prediction*.
Language Technologies Research Centre and the Interactive Language Technologies Group of NRC-IIT. Gatineau, Quebec. Mar 29, 2011.

108. *Translation beyond words: unsupervised word segmentation and statistical machine translation*.
Department of Computer Science, University of Alberta. Nov 23, 2010.

109. *Latent TAG Derivations for Semantic Role Labeling*.
University of Heidelberg, Heidelberg, Germany. Feb 4, 2010.
Cambridge University, Cambridge, U.K. May 12, 2010.
Open University, Milton Keynes, U.K. April 15, 2010.

110. *Bootstrapping a classifier using the Yarowsky algorithm*
University of Edinburgh, Edinburgh, Scotland. Oct 2, 2009.

111. *Lexicalized Tree-adjoining Grammar applied to semantic role labeling*
LORIA/INRIA. Nancy, France. Jun 5, 2008.

112. *Extensions of Regular Tree Grammars and their relation to Tree-adjoining Grammars*
Information Sciences Institute (ISI), Marina del Rey, USA. 16 Aug, 2007.

113. *Context-free languages is to Regular tree languages as Tree-adjoining languages is to what?*
University of Pennsylvania, Philadelphia, USA. Apr 21, 2007.

114. *Semi-supervised learning for statistical machine translation*
Machine Learning for Multilingual Information Access, NIPS 2006 workshop, Vancouver, BC. Dec 9, 2006.

115. *Bootstrapping statistical parsers from small data-sets*
University of Alberta, Artificial Intelligence Lab Seminar, Edmonton, AB. Mar 7, 2003.

116. *Combining Structural and Statistical Information: Relevance for Efficient Processing*
Seminar on Efficient Processing with High-Level Grammatical Formalisms, Schloss Dagstuhl, Germany. Oct 21, 1999.

117. *LR-parsing of very large Tree Adjoining Grammars*
DFKI, Saarbrücken, Germany. Aug 28, 1997.

118. *Synchronous Tree Adjoining Grammars and their application to Machine Translation*
Centre for Development of Advanced Computing, Pune, India. Jul 15, 1996.

Presentations

119. *Computational Constraints on Linguistic Descriptions*
Defining Cognitive Science at SFU series. Simon Fraser University, Burnaby, BC. Sep 26, 2007.

120. *A Statistical Parser for Hindi in $\leq$ 2 Weeks*
Joint work with P. Kanade, T. P. Reddy, M. Parakh, V. Mehta. Presentation at the Computational Linguistics Lunch (CLunch) Seminar at the University of Pennsylvania, 16 Jan 2002.

121. *A Statistical Parser for Hindi*
Joint work with P. Kanade, T. P. Reddy, M. Parakh, V. Mehta. Presentation of a 2 week project at the Corpus-Based NLP Workshop, Anna University, Chennai, India, 30 Dec 2001.

122. *Co-Training Methods for Statistical Parsing using Lexicalized Grammars*
Presentation at the NLP Open House 2000, IBM T. J. Watson Research Center, Hawthorne, 31 Oct 2000.

123. *Typing as a means for validating feature structures*
A. Sarkar and S. Wintner. Presentation by S. Wintner at CLIN99, Utrecht University, Netherlands, 10 Dec 1999.

124. *Structural Language Modeling using Stochastic Tree Adjoining Grammars*
Presentation at the AT&T Student Research Day 1999, 29 Oct 1999.

125. *Handling Coordination in Tree Adjoining Grammars*
Tutorial session at the Fourth Workshop on Tree Adjoining Grammars, TAG+ 4, Philadelphia, PA, 31 July 1998.

126. *Grammar Inference using TAGs and Conditions on their Consistency*
Poster Presentation at the First Annual Northeast Cognitive Science Society (NECSS) Graduate Conf. Cornell University, Ithaca, NY, 1-2 May 1998.

4 Teaching

4.1 Graduate Student and Post-doc Supervision

Summary

Summary of senior supervisory duties:

Post-doc		Ph.D.		M.Sc.		Students Graduated
Current	Previous	Active	Graduated	Active	Graduated	
0	2	4	6	7	21	27 + 1 (co-senior supervisor)

I have also supervised 23 undergraduate student research projects.

Senior Supervision – Postdoctoral Fellows

Name	Thesis/Project Title	Began	Completed
Mark Schmidt	Advanced Annotation Models for Text Analysis	2013-3	2014-2
Maxim Roy	Advanced Annotation Models for Text Analysis	2011-3	2012-3

Senior Supervision – Graduated Students

Name	Degree	Thesis/Project Title	Began	Completed
Ann Clifton	Ph.D.	Morphologically complex languages in statistical machine translation	2010-3	2015-3
Maryam Siabhani	Ph.D.	Left to right decoding for statistical machine translation	2012-2	2016-2
Baskaran Sankaran	Ph.D.	Improvements to Hierarchical phrase-based translation	2008-1	2013-3
Majid Razmara	Ph.D.	Ensemble decoding for machine translation	2009-3	2013-2
Yudong Liu	Ph.D.	Semantic Role Labeling using Lexicalized Tree Adjoining Grammars	2002-3	2009-3
Gholamreza Haffari	Ph.D.	Semi-supervised learning for natural language processing	2005-1	2009-2

Kou, Xinxin	M.Sc.	Speed versus Accuracy in Neural Sequence Tagging for Natural Language Processing	2015-3	2017-3
Wu, Zhelun	M.Sc.	Decipherment of Evasive or Encrypted Offensive Text	2014-3	2016-3
Shavarani, Sayyedhassan	M.Sc.	Training Data Annotation for Segmentation Classification in Translation	2014-3	2016-1
Soleimani Nasab, Mohammad Mahdi	M.Sc.	On the Importance of Decoding in Semi-supervised Learning	2014-2	2016-2
Bu, Te	M.Sc.	Joint Prediction of Word Alignment and Alignment Types for Statistical Machine Translation	2013-3	2015-3
Mehdizadeh Seraj, Ramtin	M.Sc.	Paraphrases for Statistical Machine Translation	2013-3	2015-3
Sabharwal, Jasneet	M.Sc.	Bilingual Language Models using Word Embeddings for Machine Translation	2013-3	2016-2
Rohit Dholakia	M.Sc.	Real-world Use of Pivot Languages to Translate Low-resource Languages	2012-3	2014-1
Ravikiran Vadlapudi	M.Sc.	Verbose Labels for Semantic Roles	2011-3	2013-1
Max Whitney	M.Sc.	Bootstrapping via Graph Propagation	2010-3	2012-2
Marzieh Razavi	M.Sc.	Ensemble-based discriminative models for dependency parsing	2010-3	2012-2
Young-chan Kim	M.Sc.	Topic in Syntax-based Machine translation	2009-3	2012-1
Porus Patell	M.Sc.	Experiments on Phrasal Chunking in NLP using Exponentiated Gradient for Structured Prediction	2009-2	2011-1
Ann Clifton	M.Sc.	Latent variable discriminative learning for NLP	2008-3	2010-2
Ajeet Grewal	M.Sc.	Model Adaptation for Statistical Machine Translation	2007-3	2009-3
Dong Song	M.Sc.	Experimental comparison of discriminative learning approaches for Chinese word segmentation	2006-2	2008-2
Akshay Gattani	M.Sc.	Automated natural language headline generation	2005-3	2007-1
Julia Birke	M.Sc.	A clustering approach for the nearly unsupervised recognition of nonliteral language	2002-3	2005-2
Hong Shen	M.Sc.	Voting between multiple data representations for text chunking	2003-3	2004-2
Yingjian Zhang	M.Sc.	Prediction of financial time series with Hidden Markov Models	2002-3	2004-2

Co-Senior Supervision – Graduated Students

Name	Degree	Thesis/Project Title (co-supervised with)	Status	Began	Completed
Zhongmin Shi	Ph.D.	Functional relationship identification from bio-medical Literature (Senior supervisor: Prof. Fred Popowich)	Completed	2004-3	2007-2

4.2 Supervision of Undergraduate Students

Name	Degree	Funding Source	Date
Parmit Chilana	B.Sc. (SFU)	NSERC USRA	2003-1
Parmit Chilana	B.Sc. (SFU)	NSERC USRA	2003-2
Max Whitney	B.Sc. (SFU)	NSERC	2008-3
Max Whitney	B.Sc. (SFU)	NSERC	2009-2
Manaal Faruqui	B.Sc. (IIT Kharagpur)	MITACS Globalink	2009-2
Diptesh Chatterjee	B.Sc. (IIT Kharagpur)	MITACS Globalink	2010-2
Mengkyun Kong	B.Sc. (SFU DDP)	Google	2011-2
Yuhuan Jiang	B.Sc. (Chongqing Univ)	MITACS Globalink	2013-2
Haotian Zhang	B.Sc. (SFU DDP)	NSERC CRD	2013-2
Ziqi Wang	B.Sc. (SFU DDP)	NSERC CRD	2013-2
Ziqi Wang	B.Sc. (SFU DDP)	NSERC CRD	2014-1
Yulan Huang	B.Sc. (SFU DDP)	NSERC CRD	2014-1
Yulan Huang	B.Sc. (SFU DDP)	SFU VP-R USRA and NSERC	2014-2
Utkarsh Patange	B.Sc. (IIT Kanpur)	MITACS Globalink	2014-2
Sun Yan	B.Sc. (SFU DDP)	NSERC	2014-3
Di Fu	B.Sc. (SFU DDP)	NSERC	2014-3
Jin, Torres	B.Sc. (SFU DDP)	NSERC	2015-2
Vacariu, Andrei	B.Sc. (SFU)	NSERC	2015-2016
Wang, Yizhou	B.Sc. (SFU)	NSERC	2015-2016
La, Sean	B.Sc. (SFU)	NSERC	2016-1
Chen, (Kingston) Jingda	B.Sc	NSERC	2016-2 and 2016-3
Gu, (Jetic) Yilong	B.Sc.	NSERC	2016-2 to 2017-3
Ke, (Ryoma) Jiahao	B.Sc.	NSERC	2016-2
Liu, (Luke) Shiqi	B.Sc.	NSERC	2016-2
Li, (Jerry) Jiaqi	B.Sc.	NSERC	2016-2
Avvaru, Pravalika	B.Sc.	MITACS Globalink	2016-2
Pancholi, Chinmaya	B.Sc.	MITACS Globalink	2017-2

4.3 Semesterly Teaching at Simon Fraser University

Awards

I was awarded the **Excellence in Undergraduate Teaching** Award in 2007 by the SFU Undergraduate Computing Science Student Society (CSSS).

Courses Taught at Simon Fraser University

Legend

1: Spring, 2: Summer, 3: Fall.

The table below shows enrollments and below that are the evaluation scores for course (left) and instructor (right).

* Indicates that a graduate course was cross-listed with an undergraduate course in the same semester.

$(\_)$ The number in parentheses next to the course number refers to the course description listing below.

		Graduate Courses				Undergraduate Courses			
CMPT:		(3) 825	(4,5,6,2,1) 882	(7) 755	(8) 894	(9) 379	(10) 413	(11) 300	(12) 415
MACM:									
2002	3		6 2.83 2.67						
2003	1						27 3.75 3.5		
	2								1
	3	12 3.71 3.88							
2004	1				2	22 3.67 3.47	19 3.47 3.59		
	2				1				
	3			7 3.33 3.5		14 3.46 3.92			
2005	1						14 3.71 3.79		
	2								
	3					7 3.71 3.86			
2006	1	5 4.0 3.88						8 3.5 3.5	
	2								
	3			13 * 3.43 3.86		17 * 3.44 3.67			
2007	1		10 3.78 3.62				17 3.58 3.55		
	2								
	3					17 3.53 3.73			
2008	1	10 2.9 3.14			2		28 2.91 3.14		
2010	3	16 3.77 3.69				43 3.33 3.58			
2011	1						38 3.4 3.87		
	3		13 3.73 3.91			58 3.47 3.69			
2012	1						43 3.58 3.71		
	3	11 3.55 3.73				40 3.42 3.92			
2013	1						39 3.43 3.80		
	3					35 3.05 3.63			
2014	1	23 * 3.75 3.75					51 * 3.62 3.57		
	3	26 * 3.75 3.75					43 * 3.62 3.57		
2015	1		8 3.88 3.88						
2016	2					67 4.53 4.45			
	3	30 * 4.49 4.58	15 4.54 4.48				38 * 4.47 4.37		

Graduate Course Descriptions

1. CMPT 882: *Neural Machine Translation*. In Fall 2016 I taught a graduate seminar course on recent research in the area of neural networks for machine translation which included a basic introduction to recurrent neural networks, gated recurrent networks and more advanced topics such as encoder-decoder networks and attention models for such networks.
2. CMPT 882: *Decipherment*. In Spring 2015 I taught a graduate seminar course on computational decipherment methods with connections to cryptography and to the decipherment of ancient scripts.

3. CMPT 825: *Natural Language Processing*. In the Fall 2003 offering of this course I created a broad introduction to the field with an emphasis on hands-on experience with software and data-sets commonly used in research and development in this field. The large number of assignments (eight) was crucial in motivating the graduate students to apply theoretical knowledge learnt in class to the field of natural language processing. This course was taught again in Spring 2006 with an emphasis on statistical machine translation and in Spring 2008 with an emphasis on text mining techniques using shallow parsing and statistical parsing.
4. CMPT 882, Fall 2002, special topics course: *Statistical Learning of Natural Language*. This course was an advanced graduate course on machine learning and statistical methods in natural language processing. Lectures were based on research papers taken from journals and conferences. Many of these were very recent works which were extended by the students as part of their projects. Two of the final papers written were eventually accepted at a conference in the field.
5. CMPT 882, Spring 2007, special topics course: *Discriminative Methods in Machine Learning*. This course was an introductory graduate course on machine learning. The course focused on so-called "distribution-free" or discriminative methods, especially large-margin methods. The course covered the classical papers on the topic: covering the analysis of different learning algorithms and the trade-off between optimal learning vs. scalability of the learning model to large data-sets.
6. CMPT 882, Fall 2011, *Statistical Machine Translation*. This course was dedicated to modern methods in statistical machine translation (SMT). Graduate students learned the fundamentals of SMT and also learned how to use SMT toolkits like Moses, SRILM, and KenLM. By the end of the course, they had implemented core decoding algorithms for SMT and wrote their own code for unsupervised word alignment.
7. CMPT 755: *Compiler Theory*. This graduate course on compilers covers the theory and practice of writing a compiler. Students learnt about parsing and optimization in class and in their assignments they had to build a compiler from scratch without using tools like lex or yacc. While this course was built on my undergraduate class on compilers, the graduate students had to do additional work such as writing three survey papers. The Fall 2006 offering of this course was cross-listed with CMPT 379, the undergraduate compiler course, but the course content was the same as the Fall 2004 offering.
8. CMPT 894: *Directed Reading Course*. This course is a directed reading course for graduate students, involving weekly meetings, a syllabus of papers that are read and discussed, and in some cases also including assignments. In Spring 2008 and Summer 2004, the topic for the students involved was a broad introduction to *Natural Language Processing*. In Spring 2004, the topic for the student involved was *Algebraic Foundations of Statistical Parsing*.

Undergraduate Course Descriptions

9. CMPT 379: *Principles of Compiler Design*. When I first taught this course, I used an approach that focused on theory of compilation in class but which had detailed homework assignments that enabled students to build a working compiler entirely from scratch for the MIPS instruction set (no compiler construction tools were used). The daunting task of building a compiler from scratch was broken down into nine small assignments. Students who fell behind earlier were given solutions to enable them to catch up. In 2007 the entire course was then re-worked to use compiler tools such as lex and yacc, and students built a working compiler for my own invented programming language called **`Decaf`**. In 2011, the course was changed yet again to move away from MIPS to the use of LLVM libraries for code generation for the x86 instruction set.
10. CMPT 413: *Computational Linguistics*. I have continued to develop new material each time I have taught this fourth year undergraduate course. The course focuses on broad fundamentals of the field, taking into account the various strategies previously proposed in the ACL Teaching CL workshop. A total of nine assignments introduced Perl programming and used it to implement natural language processing tools such as taggers, parsers, feature unifiers, among others. In 2007, this course was completely re-worked again to use the open-source Python-based NLTK toolkit. All the assignments which were originally in Perl are now done in Python. A few bug-fixes and contributions were given back to the NLTK open-source project as a result of this course.

The lecture notes and/or the assignments for this course have been used by instructors at other universities, such as the University of British Columbia (UBC), Swarthmore College, and the University of Pennsylvania.

11. MACM 300: *Introduction to Formal Languages and Automata*. This course is an introduction to the mathematical foundations of computation. It covers three classes of models: models with finite amount of memory (finite-state automata); models with stack memory (push-down automata); and unrestricted models (Turing machines). It also covers the impact of formal language theory for many computer science applications: in compilers, natural language processing, and program verification.
12. CMPT 415: *Undergraduate Research Project*. This course is an optional final year project for undergraduate honors students. The Summer 2003 offering of this was a research project on the use of VoiceXML and the CMU Sphinx speech recognition system to build an interactive dialog system for human-computer interaction.

Course Materials

For each of my courses, I maintain a course web page containing the lecture notes, presentation slides, and homework assignments. The course web pages are available at: `http://www.cs.sfu.ca/~anoop/courses.html`

Instructors from other institutions, such as the University of British Columbia (UBC), Swarthmore College and the University of Pennsylvania, have used my material in their courses.

5 Service

5.1 Service to Simon Fraser University

Departmental and Senate Committees

	Dates	Committee	Duties
	2013-2014	—	Program Director of Big Data M.Sc.
	2013-2014	Senate Undergraduate Awards Committee	Grant all major and minor scholarships to SFU students
	2012-2013	Research & Industrial Relations	Distinguished Lecture Series
	2012-2013	Tenure and Promotion Committee	—
	2011-2012	Tenure and Promotion Committee	Alternate member
	2010-2012	Graduate Program	—
	2010-2011	—	CS Seminar Series
(1)	2007-2008 2006-2007 2005-2006 2004-2005	Hardware & Capital Resources	Web page
(2)	2004-2005	—	Distinguished Lecture Series
(3)	2003-2004	Tenure and Promotion	
(4)	2003-2004	Graduate Programs	CMPT 891
(5)	2002-2003	Undergraduate Programs	Recruitment

Descriptions

1. I served as chair of the web page sub-committee which deals with the development, maintenance and policy decisions concerning the department web page content and back-end software. As the only faculty member involved in the web page decisions, I also mediate between faculty, staff and the development team to ensure smooth operation of the web page.
2. Along with Valentine Kabanets, I organized this lecture series as an additional service to the department. Responsibilities included the choice of invited speakers, and coordinating all local arrangements, including the web page, for each talk.
3. During this year, the Tenure and Promotion Committee created the first draft of the written policies for the department's Tenure and Promotion criteria.
4. Organized the weekly Advanced Graduate Seminar (aka CMPT 891) for Fall 2003 and Spring 2004. Tasks included organizing the talk each week by faculty or visitors, coordinating with other departmental seminars, and assigning a grade to each graduate student for participation.
5. I served on the sub-committee on undergraduate recruiting efforts. The committee was charged with improving recruiting efforts to meet provincial recruiting requirements under the Double The Opportunity (DTO) initiative.

5.2 Service to the Academic Community

Professional Service

- Secretary of the North American chapter of the Association for Computational Linguistics (NAACL) from 2009-2013, elected by the membership in 2009 and 2011.
- Reviewer for Czech National Science Foundation (GACR), 2013.
- Reviewer for NSERC Strategic Project Grants, 2013
- Reviewer for NSERC Discovery Grants, 2012-2013

Grant Reviews

- MITACS College of Reviewers for the ACCELERATE program: April 2009 – April 2010
- Grant Reviewer, The Netherlands Organization for Scientific Research (NWO).
- Panel member and reviewer, NSF Career Review Panel (Human Language and Communication, Division of Information & Intelligent Systems) 2005.

Conference Chair

- Program Co-chair, North American ACL: NAACL 2015.
- Doctoral forum chair: SIAM International Conference on Data Mining: SDM 2015.
- Tutorials Chair, American Machine Translation Association: AMTA 2014.
- Invited Applications Papers Selection Committee, 2011 International Conference on Machine Learning: ICML 2001.
- Program Co-chair, CoNLL 2010 Conference on Natural Language Learning, Uppsala, Sweden, July 15-16, 2010.
- Faculty Advisor Program Co-chair, NAACL HLT 2009 Student Research Workshop, Boulder, Colorado, June 1-3, 2009.
- Program Co-chair, The 9th International Workshop on Tree Adjoining Grammar and Related Formalisms: TAG+9. University of Tübingen, June 2008.
- Area Chair, Parsing, Conf. on Empirical Methods in Natural Language Processing and the Conf. on Comp. Natural Language Learning: EMNLP-CoNLL 2007.
- Area Chair, Machine Learning Methods, Joint conference of the Intl. Committee on Comp. Ling. and the Assoc. for Comp. Ling.: ACL-COLING 2006.

Conference Organization

- Local co-organizer of the Annual Meeting of the Association for Computational Linguistics: ACL 2017 which is the premier international conference in Natural Language Processing held in Vancouver from July 30 to August 4, 2017.
- Organizing Committee of the 1st CAIAC/Precarn AI Challenge, 2008-2009. (a Netflix style challenge to encourage AI interest in Canadian high school, undergraduate and graduate students).
- Local Preparation and Student Volunteer Coordinator, Human Language Tech. Conf. and Conf. on Empirical Methods in Natural Language Processing: HLT/EMNLP 2005.
- Local Organizer, The 7th International Workshop on Tree Adjoining Grammars and Related Formalisms: TAG+7, Vancouver, BC, 2004.
- Organizing Committee, 6th International Workshop on Tree Adjoining Grammars and Related Formalisms: TAG+6 2002, Venice, Italy.

Journal Reviews

- The Python Papers (TPP) and The Python Papers Source Codes (TPPSC) (2009)
- Linguistic Issues in Language Technology (LiLT) (2008)
- IEEE Transactions on Speech and Audio Processing (2007)
- IEEE Transactions on Knowledge and Data Engineering (2004)
- Computational Intelligence (2003)
- Natural Language Engineering (2003, 2009)
- Computational Linguistics (2001)
- Transactions of the ACL (2013-2014)

Book Reviews

- CSLI Lecture Notes 118 (Center for the Study of Language and Information, Stanford, CA). (2002)
- Statistical machine translation book, Cambridge University Press. (2007-2008)

October 29, 2017