DIPANJAN DAS

office:
Google Inc.
111 8th Avenue,
4th Floor,
New York, NY 10011 USA

web: http://www.dipanjandas.com
email: dipanjand@google.com

EMPLOYMENT

Google AI	Nov. 2016 - *present*	**Staff Research Scientist** and **Technical Lead/Manager**
New York, New York	Nov. 2013 - Nov. 2016	Senior Research Scientist
	June 2012 - Nov. 2013	Research Scientist
		Areas: natural language processing and machine learning

EDUCATION

Carnegie Mellon University	August, 2008 - May, 2012	**Ph.D.** *Department:* Language Technologies Institute *Areas:* natural language processing *Advisor:* Noah A. Smith
Carnegie Mellon University	August, 2006 - August, 2008	**M.S.** *Department:* Language Technologies Institute *Areas:* natural language processing *Advisor:* Alexander I. Rudnicky
Indian Institute of Technology, Kharagpur	July, 2001 - August, 2005	**Bachelor of Technology** *Department:* Computer Science and Engg. *Thesis:* Shallow Parsing for Free-Word Order Languages [33]

HONORS AND AWARDS

- Google Outstanding Publication in Natural Language Processing, 2011, 2013.
- Best paper award nomination, Conference on Empirical Methods in Natural Language Processing, 2013.
- Best long paper award, 49th Annual Meeting of the Association for Computational Linguistics, 2011.
- Best undergraduate thesis award, Computer Science and Engineering, Class of 2005, IIT Kharagpur.
- Dr. B.C. Roy Memorial Gold Medal for best performance in academics and co-curricular activities, Class of 2005, IIT Kharagpur. (Awarded to one student per year in the institute.)

Research Experience

Google AI May–August 2010, June 2012 – present
Research Team, New York
Topics: Weakly supervised learning for very large scale language understanding; query understanding, question answering, ranking; semi-supervised learning for part-of-speech tagging, syntactic parsing, information extraction, and semantic analysis.

Carnegie Mellon University 2008 – 2012
Language Technologies Institute, School of Computer Science
Topics: semantic and syntactic parsing, unsupervised and semi-supervised structure prediction, semi-supervised learning, text-driven forecasting, distributed machine learning.

Carnegie Mellon University 2006 – 2008
Language Technologies Institute, School of Computer Science
Topics: text summarization, information summarization from structured databases and logs, multimedia information retrieval.

Stony Brook University 2005 – 2006
Computer Science Department
Topics: non-visual web access, assistive technologies.

Indian Institute of Technology, Kharagpur 2004 – 2005
Department of Computer Science and Engineering
Topics: rule based shallow parsing for Indo-European languages.

Teaching Experience

Probabilistic Graphical Models (10-708) Fall 2010
Teaching Assistant for Noah A. Smith
Machine Learning Department, Carnegie Mellon University
Duties: taught recitations, held office hours, graded exams and homeworks

Introduction to Computer Music (15-322) Spring 2009
Teaching Assistant for Roger Dannenberg
Computer Science Department, Carnegie Mellon University
Duties: held office hours, graded examinations and homeworks

Computer Science I (CSE 114) Fall 2005, Spring 2006
Teaching Assistant for Michael Tashbook
Computer Science Department, Stony Brook University
Duties: taught recitations, held office hours, graded exams and homeworks

Programming and Data Structures (CS13002) Spring 2005
Teaching Assistant for Anupam Basu
Department of Computer Science and Engineering, IIT Kharagpur
Duties: organized programming laboratory hours

Invited lectures:
New York University (CSCI-GA.3033-001), Fall 2016, Lexical Semantics
New York University (CSCI-GA.3033-001), Fall 2015, Lexical Semantics
New York University (CSCI-GA.3033-001), Fall 2013, Part-of-Speech Tagging
New York University (CSCI-GA.3033-001), Fall 2012, Semantic Parsing
Carnegie Mellon University (11-411), Spring 2009, Dependency parsing
Carnegie Mellon University (11-762), Fall 2008, Dependency parsing

---

SELECTED TALKS AND PRESENTATIONS

Attention-Based Neural Models for Natural Language Inference.[6, 10, 34]
– Invited talk at the Columbia University Business School.
New York, USA, July, 2017.

Two Case Studies in Semantic Inference.[2, 10]
– Invited talk at the NYU Semantics Group, Department of Linguistics.
New York, USA, September, 2016.
– Invited talk at the Language Technologies Institute Colloquium, Carnegie Mellon University.
Pittsburgh, October, 2016.
– Invited talk at the Department of Computer Science, University of North Carolina, Chapell Hill.
Chapel Hill, NC, October, 2016.

Cross-Lingual Learning For Natural Language Syntax.[5, 15, 21]
– Invited talk at the Lisbon Machine Learning School.
Lisbon, Portugal, July, 2014.

Statistical Models for Frame-Semantic Parsing.[14, 4]
– Invited talk at "Frame Semantics in Natural Language Processing: A Workshop in Honor of Chuck Fillmore".
Baltimore, Maryland, June, 2014.

Weakly-Supervised Learning of Natural Language Syntax. [5, 15, 21]
– Invited talk at the Jornadas da Engenharia Electrotécnica e de Computadores.
Lisbon, Portugal, March, 2014.

Robust Shallow Semantic Parsing of Text. [4]
– Invited talk at the International Computer Science Institute FrameNet Workshop.
Berkeley, September, 2013.
– Invited talk at Toyota Technological Institute, Chicago.
Chicago, February, 2012.
– Invited talk at Google Research.
New York, February, 2012.
– Invited talk at Language Weaver.
Los Angeles, January, 2012.
– Invited talk at the University of Washington.
Seattle, January, 2012.
– Invited talk at Microsoft Research.
Seattle, January, 2012.
– Invited talk at Yahoo Research.
Santa Clara, December, 2011.

Multilingual Guidance for Unsupervised Linguistic Structure Prediction. [20, 21]
– Invited talk at the Language Technologies Institute Colloquium, Carnegie Mellon University.
Pittsburgh, January, 2011.
– Invited talk at the Department of Computer Science, University of Massachusetts, Amherst.
Amherst, October, 2011.
– Invited talk at the Center for Language and Speech Processing Seminar, Johns Hopkins University.
Baltimore, September, 2011.

Unsupervised Part-of-Speech Tagging with Bilingual Graph-Based Projections [21]
– Invited talk at Carnegie Mellon University in Qatar.
Doha, October 2011.
– Best Paper Session, ACL Conference.
Portland, June 2011.

Probabilistic Frame-Semantic Parsing [25]
– NAACL Conference.
Los Angeles, June 2010.
– Google Research.
New York, June, 2010.

Paraphrase Identification as Probabilistic Quasi-Synchronous Recognition. [27]
– ACL-IJCNLP Conference.
Singapore, August, 2009.
– Invited talk at Microsoft Research.
Bangalore, India. May, 2009.

PUBLICATIONS

Total Citations: 4127, h-index: 26, i10-index: 31. (Source: Google Scholar)

PhD thesis

[1] Dipanjan Das. *Semi-Supervised and Latent-Variable Models for Natural Language Semantics*. Committee: Noah A. Smith (chair), William Cohen, Lori Levin and Dan Roth, Carnegie Mellon University, 2012.

Journal Articles

[2] Siva Reddy, Oscar Täckström, Michael Collins, Tom Kwiatkowski, Dipanjan Das, Mark Steedman and Mirella Lapata, *Transforming Dependency Structures to Logical Forms for Semantic Parsing*. Transactions of the Association for Computational Linguistics 4. 2016.

[3] Oscar Täckström, Kuzman Ganchev and Dipanjan Das, *Efficient Inference and Structured Learning for Semantic Role Labeling*. Transactions of the Association for Computational Linguistics 3. 2015.

[4] Dipanjan Das, Desai Chen, André F.T. Martins, Nathan Schneider and Noah A. Smith. *Frame-Semantic Parsing*. Computational Linguistics 40:1. 2014.

[5] Oscar Täckström, Dipanjan Das, Slav Petrov, Ryan T. McDonald and Joakim Nivre, *Token and Type Constraints for Cross-Lingual Part-of-Speech Tagging*. Transactions of the Association for Computational Linguistics 1. 2013.

ArXiv Preprints

[6] Kenton Lee, Shimi Salant, Tom Kwiatkowski, Ankur Parikh, Dipanjan Das and Jonathan Berant, *Learning Recurrent Span Representations for Extractive Question Answering*. arXiv:1611.01436.

Refereed Conference Publications

[7] Manaal Faruqui and Dipanjan Das, *Identifying Well-formed Natural Language Questions*. Proceedings of the Conference on Empirical Methods in Natural Language Processing (**EMNLP 2018**). November 2–4, 2018. Brussels, Belgium.

[8] Manaal Faruqui, Ellie Pavlick, Ian Tenney and Dipanjan Das, *WikiAtomicEdits: A Multilingual Corpus of Wikipedia Edits for Modeling Language and Discourse*. Proceedings of the Conference on Empirical Methods in Natural Language Processing (**EMNLP 2018**). November 2–4, 2018. Brussels, Belgium.

[9] Jan A. Botha, Manaal Faruqui, John Alex, Jason Baldridge and Dipanjan Das, *Learning To Split and Rephrase From Wikipedia Edit History*. Proceedings of the Conference on Empirical Methods in Natural Language Processing (**EMNLP 2018**). November 2–4, 2018. Brussels, Belgium.

[10] Ankur P. Parikh, Oscar Täckström, Dipanjan Das and Jakob Uszkoreit, *A Decomposable Attention Model for Natural Language Inference*. Proceedings of the Conference on Empirical Methods in Natural Language Processing (**EMNLP 2016**). November 2–4, 2016. Austin, Texas, USA.

[11] Nicholas FitzGerald, Oscar Tckstrm, Kuzman Ganchev and Dipanjan Das, *Semantic Role Labeling with Neural Network Factors*. Proceedings of the Conference on Empirical Methods in Natural Language Processing (**EMNLP 2015**). September 19–21, 2015. Lisbon, Portugal.

[12] Yoav Artzi, Dipanjan Das and Slav Petrov, *Learning Compact Lexicons for CCG Semantic Parsing*. Proceedings of the Conference on Empirical Methods in Natural Language Processing (**EMNLP 2014**). October 25–29, 2014. Doha, Qatar.

[13] Jason Mann, David Zhang, Lu Yang, Dipanjan Das and Slav Petrov, *Enhanced Search with Wildcards and Morphological Inflections in the Google Books Ngram Viewer*. Proceedings of the Annual Meeting of the Association for Computational Linguistics: Human Language Technologies (**ACL 2014**). June 22–27, 2014. Baltimore, Maryland, USA.

[14] Karl Moritz Hermann, Dipanjan Das, Jason Weston and Kuzman Ganchev, *Semantic Frame Identification with Distributed Word Representations*. Proceedings of the Annual Meeting of the Association for Computational Linguistics: Human Language Technologies (**ACL 2014**). June 22–27, 2014. Baltimore, Maryland, USA.

[15] Kuzman Ganchev and Dipanjan Das, *Cross-Lingual Discriminative Learning of Sequence Models with Posterior Regularization*. Proceedings of the Conference on Empirical Methods in Natural Language Processing (**EMNLP 2013**). October 18-21, 2013. Seattle, Washington, USA. (**Best paper award nomination**)

[16] Ryan McDonald, Joakim Nivre, Yvonne Quirmbach-Brundage, Yoav Goldberg, Dipanjan Das, Kuzman Ganchev, Keith Hall, Slav Petrov, Hao Zhang, Oscar Tckstrm, Claudia Bedini, Nria Bertomeu Castell and Jungmee Lee, *Universal Dependency Annotation for Multilingual Parsing*. Proceedings of the Annual Meeting of the Association for Computational Linguistics (**ACL 2013**). August 4-9, 2013. Sofia, Bulgaria.

[17] Dipanjan Das, André F.T. Martins and Noah A. Smith. *An Exact Dual Decomposition Algorithm for Shallow Semantic Parsing with Constraints*. Proceedings of the First Joint Conference on Lexical and Computational Semantics (***SEM 2012**). June 7-8, 2012. Montreal, Canada.

[18] Dipanjan Das and Noah A. Smith. *Graph-Based Lexicon Expansion with Sparsity-Inducing Penalties*. Proceedings of the North American Chapter of the Association for Computational Linguistics Human Language Technologies Conference (**NAACL-HLT 2012**). June 3-8, 2012. Montreal, Canada.

[19] Slav Petrov, Dipanjan Das and Ryan McDonald. *A Universal Part-of-Speech Tagset*. Proceedings of the Conference on Language Resources and Evaluation (**LREC 2012**). May 21-27, 2012. Istanbul, Turkey.

[20] Shay B. Cohen, Dipanjan Das and Noah A. Smith. *Unsupervised Structure Prediction with Non-Parallel Multilingual Guidance*. Proceedings of the Conference on Empirical Methods in Natural Language Processing (**EMNLP 2011**). July 27–31, 2011. Edinburgh, Scotland, UK.

[21] Dipanjan Das and Slav Petrov. *Unsupervised Part-of-Speech Tagging with Bilingual Graph-based Projections.* Proceedings of the Annual Meeting of the Association for Computational Linguistics: Human Language Technologies (**ACL 2011**). June 19–24, 2011. Portland, Oregon, USA. (**Best long paper award**)

[22] Kevin Gimpel, Nathan Schneider, Brendan O'Connor, Dipanjan Das, Daniel Mills, Jacob Eisenstein, Michael Heilman, Dani Yogatama, Jeffrey Flanigan and Noah A. Smith. *Part-of-Speech Tagging for Twitter: Annotation, Features, and Experiments.* Proceedings of the Annual Meeting of the Association for Computational Linguistics (**ACL 2011**). June 19–24, 2011. Portland, Oregon, USA.

[23] Dipanjan Das and Noah A. Smith. *Semi-Supervised Frame-Semantic Parsing for Unknown Predicates.* Proceedings of the Annual Meeting of the Association for Computational Linguistics: Human Language Technologies (**ACL 2011**). June 19–24, 2011. Portland, Oregon, USA.

[24] Kevin Gimpel, Dipanjan Das and Noah A. Smith. *Distributed Asynchronous Online Learning for Natural Language Processing.* Proceedings of the Conference on Computational Natural Language Learning (**CoNLL 2010**). July 15–16, 2010. Uppsala, Sweden.

[25] Dipanjan Das, Nathan Schneider, Desai Chen and Noah A. Smith. *Probabilistic Frame-Semantic Parsing.* Proceedings of the North American Chapter of the Association for Computational Linguistics Human Language Technologies Conference (**NAACL-HLT 2010**). June 1–6, 2010. Los Angeles, California, USA.

[26] Mahesh Joshi, Dipanjan Das, Kevin Gimpel and Noah A. Smith. *Movie Reviews and Revenues: An Experiment in Text Regression.* Proceedings of the North American Chapter of the Association for Computational Linguistics Human Language Technologies Conference (**NAACL-HLT 2010**). June 1–6, 2010. Los Angeles, California, USA.

[27] Dipanjan Das and Noah A. Smith. *Paraphrase Identification as Probabilistic Quasi-Synchronous Recognition.* Proceedings of the Joint Conference of the Annual Meeting of the Association for Computational Linguistics and the International Joint Conference on Natural Language Processing (**ACL-IJCNLP 2009**). Aug 2–7, 2009. Suntec, Singapore.

[28] André F.T. Martins, Dipanjan Das, Noah A. Smith and Eric P. Xing. *Stacking Dependency Parsers.* Proceedings of Conference on Empirical Methods in Natural Language Processing (**EMNLP 2008**). Oct 25–27, 2008. Waikiki, Honolulu, Hawaii, USA.

[29] Dipanjan Das, Datong Chen and Alexander G. Hauptmann. *Improving Multimedia Retrieval with a Video OCR.* Proceedings of IS&T/SPIE Annual Symposium on Electronic Imaging (**EI 2008**). Jan 28–31, 2008. San Jose, California, USA.

[30] Dipanjan Das, Mohit Kumar and Alexander I. Rudnicky. *Automatic Extraction of Briefing Templates.* Proceedings of the International Joint Conference on Natural Language Processing (**IJCNLP 2008**). Jan 7–12, 2008. Hyderabad, India.

[31] Jalal Mahmud, Yevgen Borodin, Dipanjan Das, and I.V. Ramakrishnan. *Combating Information Overload in Non-Visual Web Access Using Context.* Proceedings of the 2007 International Conference on Intelligent User Interfaces (**IUI 2007**). Jan 28–31, 2007, Honolulu, Hawaii, USA.

[32] Jalal Mahmud, Yevgen Borodin, Dipanjan Das, and I.V. Ramakrishnan. *Improving Non-visual Web Access Using Context.* Proceedings of the ACM SIGACCESS Conference on Computers & Accessibility (**ASSETS 2006**). Oct 23–25, 2006, Portland, Oregon, USA.

[33] Dipanjan Das, Monojit Choudhury, Sudeshna Sarkar and Anupam Basu. *An Affinity Based Greedy Approach towards Chunking for Indian Languages.* Proceedings of the International Conference on Natural Language Processing 2005 (**ICON 2005**). 18–20 December, 2005, Kanpur, India.

Refereed Workshop Publications:

[34] Gaurav Singh Tomar, Thyago Duque, Oscar Täckström, Jakob Uszkoreit and Dipanjan Das. *Neural Paraphrase Identification of Questions with Noisy Pretraining*. Proceedings of the First Workshop on Subword and Character LEvel Models in NLP (**SCLeM 2017**). September 7, 2017. Copenhagen, Denmark.

[35] Desai Chen, Nathan Schneider, Dipanjan Das and Noah A. Smith. *SEMAFOR: Frame argument resolution with log-linear models.* Proceedings of the International (ACL) Workshop on Semantic Evaluations (**SemEval 2010**). July 11–16, 2010. Uppsala, Sweden. (**Best performance in shared task**)

[36] Mohit Kumar, Dipanjan Das, Sachin Agarwal and Alexander I. Rudnicky. *Non-textual Event Summarization by Applying Machine Learning to Template-based Language Generation*. Proceedings of 2009 Workshop on Language Generation and Summarization at ACL-IJCNLP (**UCNLG+Sum 2009**). August 6, 2009.

Selected Unrefereed Reports:

[37] Nathan Schneider, Rebecca Hwa, Philip Gianfortoni, Dipanjan Das, Michael Heilman, Alan W. Black, Frederick L. Crabbe, and Noah Smith. *Visualizing Topical Quotations Over Time to Understand News Discourse*. CMU Technical Report, CMU-LTI-10-013, 2010.

[38] Dipanjan Das, Nathan Schneider, Desai Chen and Noah A. Smith. *SEMAFOR 1.0: A Probabilistic Frame-Semantic Parser*. CMU Technical Report, CMU-LTI-10-001, 2010.

[39] Dipanjan Das and André F.T. Martins. *A Survey on Automatic Text Summarization.* Literature Survey for the Language and Statistics II course at Carnegie Mellon University, 2007.

Service and Professional Activities

Area Chair of the International Conference on Learning Representations (ICLR) 2019.

Area Chair of the International Conference on Learning Representations (ICLR) 2018.

Co-Chair of the Text Mining and NLP Applications Area for EMNLP 2017.

Co-Chair of the Semantics Area for NAACL 2015.

Attendee of the 2nd Heidelberg Laureate Forum, 2014.

Journal Reviewing: *Natural Language Engineering*, *Transactions on Knowledge Discovery from Data*, *Computational Linguistics*, *Journal of Artificial Intelligence Research*, *Transactions of the Association for Computational Linguistics*.

Reviewer: ACL (2010-2018), COLING (2010, 2014), EACL (2014), EMNLP (2009, 2011-2016, 2018), ICML (2015), IJCNLP (2011), LREC (2010, 2012, 2014), NAACL (2012-2018), NIPS (2011-2013, 2018), WSDM (2015).

Organizer for the Large-Scale Lunch at Carnegie Mellon University (2009-2010).

Open Source Software

SEMAFOR (main developer): Semantic Analyzer of Frame Representations [4]. State-of-the-art shallow semantic analyzer. Downloadable from `http://www.ark.cs.cmu.edu/SEMAFOR`.

TWITTER NLP TOOLS (with 9 others): Part-of-speech tagger for Twitter data [22]. Downloadable from `http://www.ark.cs.cmu.edu/TweetNLP`.

MSTPARSERSTACKED (with André Martins): Stacked model for dependency parsing [28]. Downloadable from `http://www.ark.cs.cmu.edu/MSTParserStacked`.

INTERNS MENTORED

1. Hao Peng (University of Washington) - Semi-Parametric Methods for Generation, Summer 2018
   - Currently Ph.D. student at the University of Washington
2. Luheng He (University of Washington) - Efficient Models for Question Rewriting, Summer 2017
   - Currently Research Scientist at Google AI
3. Karthik Narasimhan (MIT) - Reinforcement Learning for Question Answering, Summer 2016
   - Currently Assistant Professor at Princeton
4. Nicholas FitzGerald (University of Washington) - Distributed Representations for Semantics, Summer 2014, 2015
   - Currently Ph.D. student at the University of Washington
5. Yoav Artzi (University of Washington) - Distributed Grammar Induction for Semantic Parsing, Summer 2013
   - Co-hosted with Slav Petrov
   - Currently Assistant Professor at Cornell Tech and Cornell University
6. Jason Mann (Columbia University) - Novel Features in Google Books NGram Viewer, Summer 2013
   - Co-hosted with Slav Petrov
   - Currently Software Engineer at Google
7. David Zhang (University of Southern California) - Novel Features in Google Books NGram Viewer, Summer 2013
   - Co-hosted with Slav Petrov
   - Currently Software Engineer at Google
8. Lu Yang (Cornell University) - Novel Features in Google Books NGram Viewer, Summer 2013
   - Co-hosted with Slav Petrov
   - Currently Software Engineer at Google

New York, August 23, 2018