

EDUCATION

PhD, IIT Bombay & Monash University, 2021

(Doctor of Philosophy)

Dissertation: Investigations into Distributional Similarity for Cognate Detection and Computational Phylogenetics

Committee: Prof. Pushpak Bhattacharyya, Prof. Gholamreza Haffari, Prof. Malhar Kulkarni, Prof. Mark Carman, Prof. Krithi Ramamritham

BTech in Computer Science & Engineering, Gautam Buddha Technical University, 2013

(Bachelor in Technology)

Thesis: "Mechanisms for Sense Discrimination", Advisor: Prof. Pushpak Bhattacharyya

AWARDS

Best Paper Honourable Mention among the Long Papers at EACL, 2021

Awarded the honourable mention among the best papers in the long papers category at the 16th conference of the European Chapter of the Association for Computational Linguistics (EACL).

Teaching Assistant of the Semester, Computer Science & Engineering Department, 2019

Software Systems Lab (CS699) with Prof. Kavi Arya

Teaching Assistant of the Semester, Computer Science & Engineering Department, 2017

Software Foundations Lab (CS251) with Prof. Kavi Arya

Best Poster Presentation, IITB-Monash Research Academy, 2016

amongst all the researchers from the institute in the early researcher's category.

Winner - Statistical Machine Translation Competition, ICON Conference, 2014

Our work using Hierarchical Phrase based Statistical Machine Translation outperformed all other systems.

CISCO Scholarship Fellow, 2016

Awarded additional 4-year scholarship from amongst all the applicants in the CSE Department.

ACM India and IARCS Travel Grant, 2017

Awarded the travel grant to present two papers at AAAI 2017 held in San Francisco, USA.

IIT Bombay Travel Grant, 2016

Awarded the travel grant to present three papers at GWC 2016 held in Bucharest, Romania.

ACADEMIC SERVICE

Novel Program Committee Board, International Joint Conference on Artificial Intelligence (IJCAI-22, 23, 24)

Invited on the novel program committee board at IJCAI-ECAI 2022, IJCAI 2023, and IJCAI 2024 for a three year tenure.

Senior Program Chair, International Joint Conference on Artificial Intelligence (IJCAI-21)

Promoted to the academic position of a senior program chair at IJCAI 2021 (required to desk accept/reject papers for a track).

Reviewer, Association for Computational Linguistics (ACL)

Reviewed papers in the Semantics, Cognitive Psycholinguistics, and Resources & Evaluation track for years **2018-2021**.

Reviewer, Association for the Advancement of Artificial Intelligence (AAAI)

Reviewed papers in the Main Conference and Demonstrations track for years **2019 – 2022 (already invited)**

Reviewer, Empirical Methods in Natural Language Processing (EMNLP)

Reviewed papers in the Cognitive Psycholinguistics/Semantics track for years **2019 - 2021**

Reviewer, International Conference on Computational Linguistics (COLING)

Reviewed papers in the Semantics track for years **2018 - 2020**

Reviewer, SIGNLL Conference on Computational Natural Language Learning (CoNLL)

Reviewed papers in the main conference for years **2019 - 2021**

Reviewer, Global Wordnet Conference (GWC)

Reviewed papers in the main conference for years **2018 - 2021**.

Reviewer (Journals), LiLT Journal (Wordnet Issue) / Eye Tracking Research and Applications / Computer Speech and Language / Patter Recognition Letters / Transactions on Asian and Low-Resource Language Information Processing

CONFERENCE TUTORIALS AND INVITED TALKS

1. Diptesh Kanojia, **Adversarial Evaluation of Quality Estimation**, Microsoft Research (India), December 2021.
2. Rudra Murthy, Tamali Banerjee, Jyotsana Khatri, Diptesh Kanojia and Pushpak Bhattacharyya, **Unsupervised Machine Translation Demystified**, 17th International Conference on Natural Language Processing (ICON 2020), Dec. 2020.
3. Diptesh Kanojia, **Use of Distributed Semantics for Cognate Detection and Computational Phylogenetics**, Invited talk at Data61, Commonwealth Scientific and Industrial Research (CSIRO), Sydney, Australia, April 2020.
4. Diptesh Kanojia, **Natural Language Processing and its intersection with Data Science**, Invited Talk at Techfest, IIT Bombay, January 2018.
5. Diptesh Kanojia, **Natural Language Processing Fundamentals and Python**, Invited Talk at VIVA IET Group of Colleges, July 2016.

MAIN CONFERENCE / JOURNAL PUBLICATIONS (SIGNIFICANT PUBLICATIONS)

1. Diptesh Kanojia, Marina Fomicheva, Tharindu Ranasinghe, Frédéric Blain, Constantin Orăsan, Lucia Specia. **Pushing the Right Buttons: Adversarial Evaluation of Quality Estimation**. Main Conference proceedings of the Sixth Conference on Machine Translation (WMT 2021), November 2021.
2. Anirudh Mittal, Pranav Jeevan, Prerak Gandhi, Diptesh Kanojia, Pushpak Bhattacharyya. **'So You Think You're Funny?': Rating the Humour Quotient in Standup Comedy**. Proceedings of the 26th Conference of the Empirical Methods for Natural Language Processing (EMNLP 2021), November 2021.
3. Diptesh Kanojia, Prashant Sharma, Pushpak Bhattacharyya, Gholamreza Haffari and Malhar Kulkarni, **Cognition-aware Cognate Detection**, 16th conference of the European Chapter of the Association for Computational Linguistics (EACL 2021), April 2021.
4. Girishkumar Ponkiya, Diptesh Kanojia, Pushpak Bhattacharyya and Girish Palshikar, **FrameNet-Assisted Noun Compound Interpretation**, Findings of the Joint Conference of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (ACL-IJCNLP 2021), August, 2021
5. Diptesh Kanojia, Raj Dabre, Shubham Dewangan, Pushpak Bhattacharyya, Gholamreza Haffari, Malhar Kulkarni, **Harnessing Cross-lingual Features to Improve Cognate Detection for Low-resource Languages**, 28th International Conference on Computational Linguistics (COLING 2020), December 2020.
6. Diptesh Kanojia, Pushpak Bhattacharyya, Malhar Kulkarni, Gholamreza Haffari, **Challenge Datasets of Cognate and False Friend Pairs for Indian Languages**, Language Resources and Evaluation Conference (LREC 2020), May 2020.
7. Akash Sheoran, Diptesh Kanojia, Aditya Joshi, Pushpak Bhattacharyya, **Recommendation Chart of Domains for Cross-Domain Sentiment Analysis: Findings of A 20 Domain Study**, Language Resources and Evaluation Conference (LREC 2020), May 2020.
8. Diptesh Kanojia, Abhijeet Dubey, Malhar Kulkarni, Pushpak Bhattacharyya, Reza Haffari, **Utilizing Word Embeddings based Features for Phylogenetic Tree Generation of Sanskrit Texts**, 6th International Sanskrit Computational Linguistics Symposium (ISCLS 2019), October 2019.
9. Diptesh Kanojia, Kevin Patel, Pushpak Bhattacharyya, Malhar Kulkarni, Reza Haffari, **Utilizing Wordnets for Cognate Detection among Indian Languages**, 10th Global WordNet Conference (GWC 2019), July 2019.
10. Sandeep Mathias, Diptesh Kanojia, Kevin Patel, Samarth Agarwal, Abhijit Mishra, Pushpak Bhattacharyya, **Eyes are the Windows to the Soul: Predicting the Rating of Text Quality Using Gaze Behaviour**, The 56th Annual Meeting of The Association for Computational Linguistics (ACL 2018), July 2018.
11. Diptesh Kanojia, Kevin Patel, Pushpak Bhattacharyya, **Indian Language Wordnets and their Linkages with Princeton WordNet**, Language Resources and Evaluation Conference (LREC 2018), May 2018.
12. Abhijit Mishra, Diptesh Kanojia, Kuntal Dey, Seema Nagar and Pushpak Bhattacharyya, **Scanpath Complexity: Modeling Reading Effort using Gaze Information**, Thirty-first Association for the Advancement of Artificial Intelligence Conference (AAAI 2017), San Francisco, California, USA, February 2017.



13. Abhijit Mishra, Diptesh Kanojia, Seema Nagar, Kuntal Dey, and Pushpak Bhattacharyya, **Leveraging Cognitive Features for Sentiment Analysis**, The SIGNLL Conference on Computational Natural Language Learning (CoNLL 2016), Berlin, Germany, August 2016.
14. Abhijit Mishra, Diptesh Kanojia, Seema Nagar, Kuntal Dey, and Pushpak Bhattacharyya, **Harnessing Cognitive Features for Sarcasm Detection**, Association for Computational Linguistics (ACL 2016), Berlin, August 2016.
15. Diptesh Kanojia, Aditya Joshi, Pushpak Bhattacharyya, and Mark J. Carman, **That'll do fine!: A coarse lexical resource for English-Hindi Machine Translation, using polylingual topic models**, Language Resource and Evaluation Conference (LREC 2016), Portorož, Slovenia, May 2016.
16. Shehzaad Dhuliawala, Diptesh Kanojia, and Pushpak Bhattacharyya, **SlangNet: A WordNet like resource for English Slang**, Language Resource and Evaluation Conference (LREC 2016), Portorož, Slovenia, May 2016.
17. Abhijit Mishra, Diptesh Kanojia, and Pushpak Bhattacharyya, **Predicting Readers' Sarcasm Understandability by Modelling Gaze Behaviour**, Thirtieth Association for the Advancement of Artificial Intelligence Conference (AAAI 2016), Phoenix, Arizona, U.S.A., February 2016.
18. Diptesh Kanojia, Shehzaad Dhuliawala, and Pushpak Bhattacharyya, **A picture is worth a thousand words: Using OpenClipArt library for enriching IndoWordNet**, Global WordNet Conference (GWC 2016), Bucharest, Romania, January 2016.
19. Diptesh Kanojia, Shehzaad Dhuliawala, Abhijit Mishra, Naman Gupta and Pushpak Bhattacharyya, **TransChat: Cross-Lingual Instant Messaging for Indian Languages**, International Conference on Natural Language Processing (ICON 2015), Goa, India, December 2015.
20. Diptesh Kanojia, Aditya Joshi, Pushpak Bhattacharyya, and Mark J. Carman, **Using multilingual topic models for improved alignment in English-Hindi MT**, International Conference on Natural Language Processing (ICON 2015), Goa, India, December 2015.
21. Diptesh Kanojia, Manish Shrivastava, Raj Dabre, & Pushpak Bhattacharyya, **PaCMan: Parallel Corpora Management Workbench**, International Conference on Natural Language Processing (ICON 2014), Goa, India, December 2014.
22. Diptesh Kanojia, Pushpak Bhattacharyya, Raj Dabre, Siddhartha Gunti & Manish Shrivastava, **Do not do processing, when you can look up: Towards a Discrimination Net for WSD**, Global Wordnet Conference (GWC 2014), Tartu, Estonia, January 2014.
23. Salil Joshi, Diptesh Kanojia, and Pushpak Bhattacharyya, **More than meets the eye: Study of Human Cognition in Sense Annotation**, North American Chapter of Association of Computational Linguistics (NAACL 2013), Atlanta, U.S.A., June 2013.
24. Arindam Chatterjee, Salil Joshi, Diptesh Kanojia, Akhlesh Meena and Pushpak Bhattacharyya, **A Study of the Sense Annotation Process: Man v/s Machine**, Global WordNet Conference (GWC 2011), Matsue, Japan, January, 2012.

CONFERENCE DEMONSTRATION TRACK PUBLICATIONS

25. Aditya Joshi, Diptesh Kanojia, Pushpak Bhattacharyya, Mark Carman, **Sarcasm Suite: A browser-based engine for sarcasm detection and generation**, Thirty-first Association for the Advancement of Artificial Intelligence Conference: Demonstrations Track (AAAI 2017), San Francisco, California, USA, February 2017.
26. Diptesh Kanojia, Arindam Chatterjee, Salil Joshi, and Pushpak Bhattacharyya, **Discrimination Net for Hindi**, International Conference for Computational Linguistics (COLING 2012), Mumbai, India, December 2012.

CONFERENCE WOKRSHOP & COMPETITION PUBLICATIONS

Complete list is available on Google Scholar. Please follow [the link here for the complete list](#).

WORK EXPERIENCE

| | |
|---|--|
| Postdoctoral Research Fellow (2021-Present) | <ul style="list-style-type: none">▪ Currently working as a Postdoctoral Research Fellow at the Centre for Translation Studies, School of Literature and Languages, University of Surrey.▪ EU Grant accepted for project bid - "MENTAL HEALTH 4 ALL".▪ Part of multiple teams for writing project grants (CHANSE Bid, EPSRC DE Network Plus). |
| Research Engineer, CFILT, IITB (2013 - 2015) | <ul style="list-style-type: none">▪ Research Engineer at Centre for Indian Languages Technology, IIT Bombay.▪ Cognitive studies for Sentiment detection, Sarcasm Understandability, and Detection, through Gaze Tracking.▪ Multilingual Topic Modelling to improve Indian Language Machine Translation.▪ PanchaBhoota - Hierarchical Phrase based Statistical Machine Translation system for five language pairs.▪ Core Systems Administrator for Centre for Indian Language Technology handling / troubleshooting / maintaining both server systems, and desktop computers. |

Diptesh Kanojia

<https://dipteshkanojia.github.io>

Email: dipteshkanojia@gmail.com, d.kanojia@surrey.ac.uk



IITB-Monash
Research Academy



POSITIONS OF RESPONSIBILITY

| Teaching Assistant & Mentorship | IIT Bombay | Jan 2017 – Present |
|---|--|--------------------|
| CS684 CS699 CS251 CS213 CS226 | <ul style="list-style-type: none">Teaching Assistant for "Software Lab (PG)", and "Embedded Systems" by Prof. Kavi Arya.Teaching Assistant for the course "Software Systems Lab" run by Prof. Kavi Arya.Teaching Assistant for the course "Software Architecture" run by Prof. R. K. Joshi.Teaching Assistant for the course "Data Structures" conducted by Prof. R.K. Shyamsundar.Teaching Assistant for the course "Digital Logic Design" conducted by Prof. Supratik Chakraborty. | |
| Mentor, SoS | <ul style="list-style-type: none">Mentored a group of undergraduate students for "Summer of Science (SoS)" organized by Maths and Physics Club, IIT Bombay for projects based on Machine learning. | |

| Web Nominee & Design Lead | PG Academic Council, IIT Bombay | Mar 2017 – Mar 2018 |
|--------------------------------------|---|---------------------|
| Head, Design Team & Doctoral Council | <ul style="list-style-type: none">Worked with the Institute Doctoral Council as an internal group lead and conceptualized the PhD Qualifier Document, also lead the design team for various activities.Discussed the issue of tagging for PhD Students and proposed an alternative scheme for PG Students where Best X out of Y courses can be used to calculate CPI for IITB-Monash Research Academy Students.Created an online interface for student grievances to be later clubbed with a central portal.Proposed more Faculty Advisors for PhD Students (X faculty on every Y students).Direct Transfer of CISCO Fellowship to student accounts instead of the release of a cheque. | |

| Research Scholars Representative | CSE Department, IIT Bombay | July 2016 – July 2017 |
|----------------------------------|--|-----------------------|
| Research Scholars Representative | <ul style="list-style-type: none">Representing research scholars in the Departmental Post-Graduate Committee (DPGC) meetings.Organizing various meetings between potential employers, and interested research scholars.Helping organize intra-departmental meets between faculty-RS, and among RS, if need be.Organizing monthly 'Chai-meet' for RS, where they interact and share ideas across their research areas.Organizing RISC as a core-group member. | |

| Academy Student Leadership Team | IITB-Monash Research Academy | Nov 2016 – Present |
|---------------------------------|---|--------------------|
| Member, Student Leadership Team | <ul style="list-style-type: none">Organization of the "Pizza pe charcha" event for discussion of research happening among the IITB-Monash Research Academy students.Initiating an internal placement cell for the IITB-Monash Research Academy graduates.Discussion on various issues which are commonly faced by the Academy students. | |

PREVIOUS RESEARCH WORK & PROJECTS

| "WordNet and its Applications" | Summer Internship CSE Department, IIT Bombay | Guide: Prof. Pushpak Bhattacharyya July 2011 – Aug 2011 |
|--------------------------------|--|--|
| IWSD Algorithm | <ul style="list-style-type: none">Work based on Error Analysis of the IWSD algorithm, divided in various parts, viz. POS based, Ontology based, Ablation test on IWSD parameters.Aim being the importance of contextual part of the scoring function of IWSD algorithm. | |

| "Man vs. Machine Annotation for WSD" | Research Project CSE Department, IIT Bombay | Guide: Prof. Pushpak Bhattacharyya Jan 2012 – Mar 2012 |
|---|---|---|
| Study of Man / Machine Annotation Process | <ul style="list-style-type: none">Our work aimed at understanding the meaning and significance of "contextual evidence", with respect to the human sense disambiguation task.The experiments probed into the disambiguation time across different POS categories and ontological categories, they also investigated into the set of clues in the neighbouring words of a pre-defined target word. | |

| "Mechanisms for Sense Discrimination" | Minor Research Project CSE Department, IIT Bombay | Guide: Prof. Pushpak Bhattacharyya Jun 2012 – Aug 2012 |
|--|---|---|
| Development of "Sense Discrimination Tool" | <ul style="list-style-type: none">Emphasized on importance of context and clue words around a target word synset word.Developed "Sense Discrimination Tool" which facilitates the addition of clue words by lexicographers manually into the WordNet database.Laid groundwork for automation of the tool using clues present in the database, using various heuristics. | |

| "Mechanisms for Sense Discrimination" | Major Research Project CSE Department, IIT Bombay | Guide: Prof. Pushpak Bhattacharyya Jan 2013 – Mar 2013 |
|---|---|---|
| Improvements in "Sense Discrimination Tool" and its Future Implications | <ul style="list-style-type: none">Improved the context word searching scenario by providing the lexicographers with a rich corpus.Facilitated the clue addition to database by including Google Hindi Transliteration API within the tool, and by including Concordancer Corpus search.Automated the clue searching by providing with an automatically generated possible clue set.Work on automatic clue addition in future using PMI and G² indexing of the clues.Laid groundwork for a possible future rule based framework to be called, "Discrimination Net". | |