

# Devendra Singh Sachan

Semantic Scholar: <https://bit.ly/3290U3Z>

Google Scholar: [goo.gl/nAF9ew](https://scholar.google.com/citations?user=nAF9ew)

Email : [sachande@mila.quebec](mailto:sachande@mila.quebec)

[devendra.sachan@mail.mcgill.ca](mailto:devendra.sachan@mail.mcgill.ca)

Mobile : +1-438-866-5202

## EDUCATION

---

- **Doctor of Philosophy in Computer Science** Sept 2019 – Present  
McGill University, Mila-Québec AI Institute; GPA: 4/4  
*Advisor: Joelle Pineau, William Hamilton* Montréal, QC, Canada
- **Master's in Computational Data Science** Aug 2016 – Dec 2017  
School of Computer Science, Carnegie Mellon University; GPA: 4/4+  
*Advisor: Eric Nyberg* Pittsburgh, PA, USA
- **Bachelor of Technology in Electronics and Communication Engineering** Aug 2009 – Jun 2013  
Indian Institute of Technology, Guwahati  
*Advisor: Amit Sethi* Assam, India

## RESEARCH INTERESTS

---

Natural Language Processing, Deep Learning, Machine Learning

## WORK EXPERIENCE

---

- **Facebook** Remote Internship  
*FAIR Labs* Fall 2021  
*Research Scientist Intern* Host: Luke Zettlemoyer  
– Ongoing research on language models.
- **Amazon** Remote Internship  
*AWS Language Technologies* Summer 2021  
*Applied Scientist Intern* Host: Marcello Federico  
– Research on domain-adaptation methods without finetuning for neural machine translation.  
– Worked on retriever-based approaches for machine translation.
- **Nvidia** Remote Internship  
*Applied Deep Learning Research Team* Summer 2020  
*Research Intern* Host: Bryan Catanzaro  
– Research on improving pre-training methods for generative language models such as T5.  
– Implemented learnable document-retrieval approaches such as REALM and DPR and trained them on web-scale text collections in a distributed setting consisting of 500+ GPUs.  
– Internship paper accepted for oral presentation at ACL 2021.
- **Petuum** Pittsburgh, PA, USA  
*Applied Machine Learning Research* Feb 2018 - Aug 2019  
– Research on neural network models for language understanding tasks such as sequence labelling, entity linking, text classification, paraphrase generation, and question-answering.  
– Designed a framework for retrieval-based chatbots that was involved in the launch of Neurobots platform.
- **Machine Learning Department, CMU** Pittsburgh, PA, USA  
*Graduate Research Assistant with Eric Xing, Pengtao Xie* Jan 2017 - Dec 2017  
– Research on the construction of medical knowledge graph.
- **Enixta Innovations** Hyderabad, AP, India  
*Machine Learning Research Engineer* Oct 2014 - Jun 2016

- Designed end-to-end pipeline for reviews text mining that included identification of keyword phrases, tokenization, word embeddings, word-sense disambiguation, and aspect-based sentiment analysis.
- Building a pipeline for a content-based recommendation engine using product specifications.

## PUBLICATIONS

---

1. **D. S. Sachan**, S. Reddy, W. Hamilton, C. Dyer, and D. Yogatama. “*End-to-End Training of Multi-Document Reader and Retriever for Open-Domain Question Answering*”. In Proceedings of the 35<sup>th</sup> Conference on Neural Information Processing Systems (**NeurIPS 2021**) [PDF]
2. **D. S. Sachan**, M. Patwary, M. Shoeybi, N. Kant, W. Ping, W. Hamilton, and B. Catanzaro, 2020. “*End-to-End Training of Neural Retrievers for Open-Domain Question Answering*”. In Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics (**ACL 2021**) (**Oral Presentation**) [PDF]
3. **D. S. Sachan**, Y. Zhang, P. Qi, and W. Hamilton, 2020. “*Do Syntax Trees Help Pre-trained Transformers Extract Information?*” In Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics (**EACL 2021**) (**Oral Presentation**) [PDF]
4. **D. S. Sachan**, M. Zaheer, R. Salakhutdinov. “*Revisiting LSTM Networks for Semi-Supervised Text Classification via Mixed Objective Function*”. In Proceedings of the 33<sup>rd</sup> AAAI Conference on Artificial Intelligence (**AAAI 2019**) (**Spotlight Presentation**) [PDF]
5. Z. Hu, H. Shi, Z. Yang, B. Tan, T. Zhao, J. He, W. Wang, X. Yu, L. Qin, D. Wang, X. Ma, H. Liu, X. Liang, W. Zhu, **D. S. Sachan**, E. P. Xing. “*Texar: A Modularized, Versatile, and Extensible Toolkit for Text Generation*”. In Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics (**ACL 2019**) (**Outstanding Paper Nomination**) [PDF]
6. **D. S. Sachan**, G. Neubig. “*Parameter Sharing Methods for Multilingual Self-Attentional Translation Models*”. In Proceedings of the 3<sup>rd</sup> Conference on Machine Translation (**EMNLP-WMT 2018**) (**Oral Presentation**) [PDF]
7. **D. S. Sachan**, M. Zaheer, R. Salakhutdinov. “*Investigating the Working of Text Classifiers*”. In Proceedings of the 27<sup>th</sup> International Conference on Computational Linguistics (**COLING 2018**) [PDF]
8. **D. S. Sachan**, P. Xie, M. Sachan, E. P. Xing. “*Effective Use of Bidirectional Language Modeling for Transfer Learning in Biomedical Named Entity Recognition*”. In Proceedings of the 3<sup>rd</sup> Machine Learning for Healthcare Conference (**MLHC 2018**) (**Spotlight Presentation**) [PDF]
9. M. Zaheer, S. Reddi, **D. S. Sachan**, S. Kale, S. Kumar. “*Adaptive Methods for Nonconvex Optimization*”. In Proceedings of the 32<sup>nd</sup> Conference on Neural Information Processing Systems (**NeurIPS 2018**) [PDF]
10. Y. Qi, **D. S. Sachan**, J. Padmanabhan, M. Felix, G. Neubig. “*When and Why are Pre-trained Word Embeddings Useful for Neural Machine Translation?*” In Proceedings of the Conference of the North American Chapter of the Association for Computational Linguistics (**NAACL 2018**) [PDF]
11. G. Neubig, M. Sperber, X. Wang, M. Felix, A. Matthews, S. Padmanabhan, Y. Qi, **D. S. Sachan**, P. Arthur, P. Godard, J. Hewitt, R. Riad, L. Wang. *XNMT: The eXtensible Neural Machine Translation Toolkit*. In Conference of the Association for Machine Translation in the Americas (**AMTA 2018**) [PDF]
12. **D. S. Sachan**, U. Tekwani, A. Sethi. “*Sports Video Event Classification from Multimodal Information using Deep Learning*”. AAAI Fall Symposium Series (**AAAI 2013**) [PDF]
13. **D. S. Sachan**, S. Dutta, T. S. Naveen, C. V. Jawahar. “*Segmentation of Degraded Malayalam Words: Methods and Evaluation*”. National Conference on Computer Vision, Pattern Recognition, Image Processing, and Graphics (**NCVPRIPG 2011**) [PDF]

## RESEARCH EXPERIENCE

---

- **End-to-End Training for Open-Domain Question Answering** External Collaboration  
*Advisor: Dani Yogatama (DeepMind)* Spring 2021
  - Proposed an EM algorithm for the end-to-end training of the Fusion-in-Decoder model. Obtained performance improvements by more than 2-3 EM points on three widely used datasets.
  - Paper accepted in NeurIPS 2021.
- **Graph-based Approaches in Neural NLP Models** Research Assistant  
*Advisor: William Hamilton* Fall 2019 and Spring 2020
  - Proposed two graph-attention-based approaches to incorporate syntax-knowledge from dependency trees into BERT models. Improved state-of-the-art scores by more than 3.5 F1 points on two widely used semantic role labeling datasets.
  - Worked on implementing novel Transformer models for the challenging task of multi-hop question generation. Obtained an improvement of more than 5.0 BLEU points over the previous methods.
- **Multilingual Neural Machine Translation** Capstone Project Work  
*Advisor: Graham Neubig* Spring 2017 and Fall 2017
  - Open-sourced a parallel multilingual translation dataset for 59 languages from the TED talk transcripts.
  - Implemented several modules towards the development of “XNMT” toolkit on DyNet framework.
  - Implemented the “Transformer” model and performed experiments for bilingual and multilingual translation tasks.
- **Text Classification on Covariate Shifted Data** Independent Study Course  
*Advisors: Ruslan Salakhutdinov, Manzil Zaheer* Fall 2017
  - Proposed an approach to construct more difficult datasets for text classification tasks in which the set of “lexicons” that highly correlate with the labels are disjoint between the training and test datasets.
  - In experiments, all algorithms performed substantially worse on the difficult dataset. Proposed approaches to regularize model training by anonymization of the lexicons split.
  - Proposed a mixed objective model for semi-supervised text classification task.
- **Transfer Learning for Biomedical Named Entity Recognition** Independent Study Course  
*Advisors: Eric Xing, Pengtao Xie, Mrinmaya Sachan* Spring 2017
  - Trained a bidirectional language model (BiLM) on unlabeled data to pretrain an NER model.
  - Obtained substantial improvements in the F1 scores on four benchmark datasets for biomedical NER.
  - BiLM weight transfer also lead to a faster model training and optimal learning curve.

## TEACHING EXPERIENCE

---

- **Computer Science Department, McGill University** Montréal, QC, Canada  
*TA for Jackie Cheung: COMP-424 Artificial Intelligence* Jan 2020 – April 2020  
Designing and evaluating course assignments, exams, conducting recitations, and holding office hours.
- **Computer Science Department, Université de Montréal** Montréal, QC, Canada  
*TA for Ioannis Mitliagkas: IFT 6390 - Fundamentals of Machine Learning* Aug 2019 – Dec 2019  
Designing course assignments, labs, exams, grading, conducting recitations, and holding office hours.
- **Machine Learning Department, CMU** Pittsburgh, PA, USA  
*TA for Aarti Singh and Pradeep Ravikumar: 10-725 Convex Optimization* Aug 2017 – Dec 2017  
Designing course assignments, exams, class quizzes, grading, mentoring students in course projects, and conducting office hours.

## HONORS

---

- Highlighted reviewer of ICLR 2022 (<https://iclr.cc/Conferences/2022/Reviewers>)
- Mila/McGill Graduate Research Scholarship, 2019-2023
- Received full scholarship for DPhil in Computer Science from University of Oxford.
- Received full scholarship for PhD in Computer Science from University of North Carolina at Chapel Hill.
- CMU, MCDS Research Scholarship, May-Aug 2017
- Academic scholarship recipient by Indian Air Force (IAFBA) for four consecutive years (2009-13).
- Won the regional quiz competition at Western Air Command organized by Indian Air Force.

## ACADEMIC TALKS

---

- End-to-End Training of Multi-Document Reader and Retriever for Open-Domain Question Answering  
1-hour (virtual) presentation in the Deepmind Language team, August 2021.  
1-hour (virtual) presentation in the Google N2Formal Reading Group, July 2021.
- Domain Adaptation in Neural Machine Translation without Finetuning  
30 minute (virtual) presentation in the AWS Language team, July 2021.
- End-to-End Training of Neural Retriever Models for Open-Domain Question Answering  
10 minute presentation at ACL 2021 (virtual meeting), Aug 2021.  
30 minute (virtual) presentation in the NVIDIA Applied Deep Learning Research Team, Dec 2020.
- Do Syntax Trees Help Pre-trained Transformers Extract Information?  
20 minute presentation at EACL 2021 (virtual meeting), April 2021.  
1-hour presentation in the group meeting, Mila, Montreal, CA, April 2020.
- Revisiting LSTM Networks for Semi-Supervised Text Classification via Mixed Objective Function  
AAAI 2019, Honolulu, HI, USA, Jan 2019
- Parameter Sharing Methods for Multilingual Self-Attentional Translation Models  
Petuum Inc, Pittsburgh, PA, USA, Nov 2018
- Effective Use of Bidirectional Language Modeling for Transfer Learning in Biomedical Named Entity Recognition  
MLHC 2018, Palo Alto, CA, USA, Aug 2018

## RELEVANT COURSEWORK

---

COMP-766	Graph Representation Learning	by William Hamilton
COMP-764	Language Understanding with Deep Learning	by Siva Reddy
10-701	Machine Learning (PhD)	by Eric Xing and Matthew Gormley
10-605	Machine Learning with Large Datasets	by William Cohen
11-661	Language and Statistics	by Roni Rosenfeld
11-641	Machine Learning for Text Mining	by Yiming Yang
10-708	Probabilistic Graphical Models	by Eric Xing
11-747	Neural Networks for NLP	by Graham Neubig
15-513	Introduction to Computer Systems	by Randy Bryant and Phil Gibbons
15-829	Performance Engineering of Software Systems	by Guy Blelloch and Phil Gibbons
11-632	MCDS Analytics Capstone	by Graham Neubig and Eric Nyberg
11-631	Data Science Seminar	by Eric Nyberg and Matthias Grabmair
11-634	Capstone Planning Seminar	by Eric Nyberg and Matthias Grabmair

## SKILLS

---

- **Programming Languages:** Python, C++, C, Java, MATLAB
- **Software:** PyTorch, TensorFlow, Keras, NumPy, Scikit-learn

## PROFESSIONAL SERVICES

---

- Program Committee Member of Conference in Machine Translation (WMT) 2018, 2019, Rep4NLP 2020
- Reviewer of ICLR (2022, 2021), NeurIPS (2022, 2021, 2020), NAACL 2021, ACL 2019, NAACL 2019, CoNLL 2018, EMNLP 2018
- Membership: ACM, AAAI, ACL