

Daniel Cer, *Curriculum Vitae*

CONTACT INFORMATION Google Inc., Attn. Daniel Cer (cer@), 1600 Amphitheatre Parkway Mountain View, CA 94043, USA
School of Information, University of California at Berkeley
Attn. Daniel Cer, 102 South Hall # 4600 Berkeley, CA 94720, USA
Phone: + 1 (650) 862-2394
Email: cer@google.com, dcer@berkeley.edu

RESEARCH INTERESTS natural language processing, representation learning, computational semantics, machine translation

EDUCATION Postdoc Stanford Natural Language Processing Group, Stanford University, 2011-2013
Ph.D. Computer Science, University of Colorado at Boulder, 2011
B.S. Computer Science with Cognitive Science Cert., University of Colorado at Boulder, 2002

PUBLICATIONS *Citations: +6,041 h-Index: 27 (Google Scholar, Aug 2021)*

Neural Retrieval for Question Answering with Cross-Attention Supervised Data Augmentation. Yinfei Yang, Ning Jin, Kuo Lin, Mandy Guo, Daniel Cer. In Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 2: Short Papers) (ACL/IJCNLP-2021). 2021

Crisscrossed captions: Extended intramodal and intermodal semantic similarity judgments for MSCOCO. Zarana Parekh, Jason Baldridge, Daniel Cer, Austin Waters, Yinfei Yang. In Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: Main Volume (EACL-2021). 2021

MultiReQA: A Cross-Domain Evaluation for Retrieval Question Answering Models. Mandy Guo, Yinfei Yang, Daniel Cer, Qinlan Shen, Noah Constant. In Proceedings of the Second Workshop on Domain Adaptation for NLP (AdaptNLP-2021). 2021

Multilingual universal sentence encoder for semantic retrieval. Yinfei Yang, Daniel Cer, Amin Ahmad, Mandy Guo, Jax Law, Noah Constant, Gustavo Hernandez Abrego, Steve Yuan, Chris Tar, Yun-Hsuan Sung, Brian Strope, Ray Kurzweil. In Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics: System Demonstrations (ACL-2020). 2020

ReQA: An Evaluation for End-to-End Answer Retrieval Models. Amin Ahmad, Noah Constant, Yinfei Yang, Daniel Cer. In Proceedings of the 2nd Workshop on Machine Reading for Question Answering (MRQA-2019). 2019

Learning cross-lingual sentence representations via a multi-task dual-encoder model. Muthuraman Chidambaram*, Yinfei Yang*, Daniel Cer*, Steve Yuan, Yun-Hsuan Sung, Brian Strope, Ray Kurzweil (* equal contribution). In Proceedings of the 4th Workshop on Representation Learning for NLP (RepL4NLP-2019). 2019

Improving multilingual sentence embedding using bi-directional dual encoder with additive margin softmax. Yinfei Yang, Gustavo Hernandez Abrego, Steve Yuan, Mandy Guo, Qinlan Shen, Daniel Cer, Yun-Hsuan Sung, Brian Strope, Ray Kurzweil. In Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence (IJCAI-19). 2019

Hierarchical Document Encoder for Parallel Corpus Mining. Mandy Guo, Yinfei Yang, Keith Stevens, Daniel Cer, Heming Ge, Yun-Hsuan Sung, Brian Strope, Ray Kurzweil. Proceedings of the Fourth Conference on Machine Translation (Volume 1: Research Papers) (WMT-2019). 2019

Universal sentence encoder for English. Daniel Cer, Yinfei Yang, Sheng-yi Kong, Nan Hua, Nicole Limtiaco, Rhomni St John, Noah Constant, Mario Guajardo-Cespedes, Steve Yuan, Chris Tar, Brian Strope, Ray Kurzweil. In Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing: System Demonstrations (EMNLP-2018). 2018

Learning semantic textual similarity from conversations. Yinfei Yang, Steve Yuan, Daniel Cer, Sheng-Yi Kong, Noah Constant, Petr Pilar, Heming Ge, Yun-Hsuan Sung, Brian Strope, Ray Kurzweil. In Proceedings of The Third Workshop on Representation Learning for NLP (ReplNLP-2018). 2018

Effective parallel corpus mining using bilingual sentence embeddings. Mandy Guo, Qinlan Shen, Yinfei Yang, Heming Ge, Daniel Cer, Gustavo Hernandez Abrego, Keith Stevens, Noah Constant, Yun-Hsuan Sung, Brian Strope, Ray Kurzweil. In Proceedings of the Third Conference on Machine Translation: Research Papers (WMT-2018). 2018

SemEval-2017 Task 1: Semantic Textual Similarity - Multilingual and Crosslingual Focused Evaluation. Daniel Cer, Mona Diab, Eneko Agirre, Inigo Lopez-Gazpio, Lucia Specia. In Proceedings of the 11th International Workshop on Semantic Evaluation (SemEval-2017). 2017

SemEval-2016 Task 1: Semantic Textual Similarity, Monolingual and Cross-Lingual Evaluation. Eneko Agirre, Carmen Banea, Daniel Cer, Mona Diab, Aitor Gonzalez-Agirre, Rada Mihalcea, German Rigau, Janyce Wiebe. In Proceedings of the 10th International Workshop on Semantic Evaluation (SemEval-2016). 2016

SemEval-2015 Task 2: Semantic Textual Similarity, English, Spanish and Pilot on Interpretability. Eneko Agirre, Carmen Banea, Claire Cardie, Daniel Cer, Mona Diab, Aitor Gonzalez-Agirre, Weiwei Guo, Iigo Lopez-Gazpio, Montse Maritxalar, Rada Mihalcea, German Rigau. In Proceedings of the 9th International Workshop on Semantic Evaluation (SemEval 2015). 2015

An empirical comparison of features and tuning for phrase-based machine translation. Spence Green, Daniel Cer, Christopher D. Manning. In Proceedings of the Ninth Workshop on Statistical Machine Translation (WMT 2014). 2014

Phrasal: A toolkit for new directions in statistical machine translation. Spence Green, Daniel Cer, Christopher D. Manning. Proceedings of the Ninth Workshop on Statistical Machine Translation (WMT 2014). 2014

SemEval-2014 Task 10: Multilingual Semantic Textual Similarity. Eneko Agirre, Carmen Banea, Claire Cardie, Daniel Cer, Mona Diab, Aitor Gonzalez-Agirre, Weiwei Guo, Rada Mihalcea, German Rigau, Janyce Wiebe. In Proceedings of the 8th International Workshop on Semantic Evaluation (SemEval 2014). 2014

Positive Diversity Tuning for Machine Translation System Combination. Daniel Cer, Christopher D. Manning and Dan Jurafsky. In Proceedings of the Eighth Workshop on Statistical Machine Translation (WMT 2013). 2013

Bilingual word embeddings for phrase-based machine translation. Will Y. Zou, Richard Socher, Daniel Cer, Christopher D. Manning. In Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing (EMNLP 2013). 2013

Fast and Adaptive Online Training of Feature-Rich Translation Models. Spence Green, Sida Wang, Daniel Cer and Christopher D. Manning. In Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (ACL 2013). 2013

**SEM 2013 shared task: Semantic Textual Similarity.* Eneko Agirre, Daniel Cer, Mona Diab, Aitor Gonzalez-Agirre and Weiwei Guo. In Proceedings of the Second Joint Conference on Lexical and Computational Semantics (*SEM 2013). 2013

Feature-rich phrase-based translation: Stanford University's submission to the WMT 2013 translation task. Spence Green, Daniel Cer, Kevin Reschke, Rob Voigt, John Bauer Sida Wang, Natalia Silveira, Julia Neidert and Christopher D. Manning. In Proceedings of the Eighth Workshop on Statistical Machine Translation (WMT 2013). 2013

SemEval-2012 Task 6: A Pilot on Semantic Textual Similarity. Eneko Agirre, Daniel Cer, Mona Diab and Aitor Gonzalez-Agirre. In Proceedings of the Sixth International Workshop on Semantic Evaluation (SemEval 2012). 2012

- Stanford: Probabilistic Edit Distance Metrics for STS.* Mengqiu Wang and Daniel Cer. In Proceedings of the Sixth International Workshop on Semantic Evaluation (SemEval 2012). 2012
- The Best Lexical Metric for Phrase-Based Statistical MT System Optimization.* Daniel Cer, Daniel Jurafsky and Christopher D. Manning. In Proceedings of the North American Association of Computational Linguistics (NAACL-10). 2010
- Parsing to Stanford Dependencies: Trade-offs between speed and accuracy.* Daniel Cer, Marie-Catherine de Marneffe, Dan Jurafsky and Christopher D. Manning. In Proceedings of Language Resources Evaluation Conference (LREC-10). 2010
- Phrasal: A Toolkit for Statistical Machine Translation with Facilities for Extraction and Incorporation of Arbitrary Model Features.* Daniel Cer, Michel Galley, Daniel Jurafsky and Christopher Manning. In Proceedings of the North American Association of Computational Linguistics - Demo Session (NAACL-10). 2010
- Measuring Machine Translation Quality as Semantic Equivalence: A Metric Based on Entailment Features.* Sebastian Padó, Daniel Cer, Michel Galley, Dan Jurafsky and Christopher D. Manning. Machine Translation. Nov 2009
- Stanford University's Arabic-to-English Statistical Machine Translation System for the 2009 NIST Evaluation.* Michel Galley, Spence Green, Daniel Cer, Pi-Chuan Chang and Christopher D. Manning. The 2009 NIST Open Machine Translation Evaluation Meeting. 2009
- Regularization and Search for Minimum Error Rate Training.* Daniel Cer, Daniel Jurafsky and Christopher D. Manning. In Proceedings of the Third Workshop on Statistical Machine Translation. 2008
- Stanford University's Chinese-to-English Statistical Machine Translation System for the 2008 NIST Evaluation.* Michel Galley, Pi-Chuan Chang, Daniel Cer, Jenny R. Finkel and Christopher D. Manning. The 2008 NIST Open Machine Translation Evaluation Meeting. 2008
- Aligning semantic graphs for textual inference and machine reading.* Marie-Catherine de Marneffe, Trond Grenager, Bill MacCartney, Daniel Cer, Daniel Ramage, Chloe Kiddon and Christopher D. Manning. AAAI Spring Symposium at Stanford. 2007
- Learning Alignments and Leveraging Natural Logic.* Nathanael Chambers, Daniel Cer, Trond Grenager, David Hall, Chloe Kiddon, Bill MacCartney, Marie-Catherine de Marneffe, Daniel Ramage, Eric Yeh and Christopher D. Manning. In Proceedings of the ACL-PASCAL Workshop on Textual Entailment and Paraphrasing. 2007
- Neural mechanisms of binding in the hippocampus and neocortex: Insights from computational models.* Daniel Cer and Randall C. O'Reilly. In H.D. Zimmer & A. Mecklinger & U. Lindenberger (Eds) Handbook of Binding and Memory, Oxford: Oxford University Press. 2006
- Learning to recognize features of valid textual entailments.* Bill MacCartney, Trond Grenager, Marie-Catherine de Marneffe, Daniel Cer and Christopher D. Manning. In Proceedings of the North American Association of Computational Linguistics (NAACL-06). 2006
- Learning to distinguish valid textual entailments.* Marie-Catherine de Marneffe, Bill MacCartney, Trond Grenager, Daniel Cer, Anna Rafferty and Christopher D. Manning. In Proceedings of the Second PASCAL Challenges Workshop. 2006
- The Detection of Emphatic Words Using Acoustic and Lexical Features.* Jason M. Brenier, Daniel Cer and Daniel Jurafsky. In Proceedings of EUROSPEECH. 2005
- Effectiveness of Neural Network Learning Rules Generated by a Biophysical Model of Synaptic Plasticity.* Dave Jilk, Daniel M. Cer and Randall C. O'Reilly. Technical report, Department of Psychology, University of Colorado, Boulder. 2003

Supervised and unsupervised models for propositional analysis. Simon Dennis, Dan Jurafsky and Dan Cer. Presentation at the Workshop on Syntax, Semantics and Statistics at the Neural Information Processing Society Conference. 2003

Issues in Recognition of Spanish-Accented Spontaneous English. Ayako Ikeno, Bryan Pellom, Dan Cer, Ashley Thornton, Jason Brenier, Dan Jurafsky, Wayne Ward and William Byrne. In Proceedings of the IEEE ISCA & IEEE Workshop on Spontaneous Speech Processing and Recognition. 2003

TEACHING

University of California at Berkeley, Berkeley, CA, USA

Lecturer, School of Information

2018 – present

Natural Language Processing with Deep Learning (W266)

RESEARCH
EXPERIENCE

Google, Inc., Mountain View, CA, USA

Senior Research Scientist, Google Research

2021 – present

Multilingual representation learning using deep learning methods for natural language processing tasks including question answering, semantic retrieval, bi-text retrieval and classification.

Software Engineer, Google Research

2015 – 2021

Representation learning for natural language processing with a focus on deep learning methods and computational semantics. Machine translation domain adaptation for localization materials.

Software Engineer, YouTube / Display Ads

2013 – 2015

Machine learning lead for YouTube channel classifications. Machine learning for video classification.

Stanford University, Stanford, CA, USA

Postdoctoral Researcher, Stanford NLP Group

2011 – 2013

Semantic textual similarity. Bilingual word embeddings. Positive diversity training for machine translation system combination. Machine translation models with large discriminative feature sets.

Research Assistant, Stanford NLP Group

2005 – 2011

Machine learning for phrase-based machine translation systems, including creating the state-of-the-art Phrasal machine translation package.

Google, Inc., Mountain View, CA, USA

Intern, Google Research - Machine Translation Team

Summer-Fall 2004

Implementation of language modeling data structures for use in machine translation.

University of Colorado, Boulder, CO, USA

Research Assistant, Center for Spoken Language Research

2003 – 2005

Machine learning for speech recognition tasks, and development of a parallel/distributed acoustic model trainer for the Sonic speech recognition system. Joint appointment with Computational Cognitive Neuroscience Lab.

Research Assistant, Computational Cognitive Neuroscience Lab

2001 - 2005

Modeling of human language processing and symbolic binding using biologically motivated neural networks with Prof. Randall C. O'Reilly.

SERVICE

Area Chair NAACL 2021 (Senior Area Chair: Sentence-level semantics), ACL 2020 (NLP Applications), CoNLL 2020, NAACL 2019 (Machine Translation), ACL 2019 (Machine Translation), EMNLP 2019 (Machine Translation and Multilinguality)

Reviewing/Program Committee EMNLP 2020, WMT 2020, ACL-IJCNLP 2020, CoNLL 2019, WMT 2019, ACL 2018, EMNLP 2018, AMTA 2018, COLING 2018 (awarded outstanding reviewer), CoNLL 2018, NAACL/HLT 2018, WMT 2018, ACL 2017, EMNLP 2017, EACL 2017, EAMT 2017, IJCNLP 2017, MT Summit XVI 2017, SemEval 2017, WMT 2017, AMTA 2016, NAACL-HLT 2016, SemEval 2016, WMT 2016 ACL-IJCNLP 2015, CoNLL 2015, EMNLP 2015, MT Summit XV 2015, NAACL 2015, SemEval 2015, WMT 2015, ACL 2014, AMTA 2014, CoNLL 2014, EMNLP 2014, WMT 2014, EMNLP 2013, IJCNLP 2013, NAACL-HLT 2013, *SEM 2013, WMT 2013

Paper Mentor COLING 2018

Board Member Information Officer for Special Interest Group on the Lexicon of the Association for Computational Linguistics (ACL/SIGLEX), 2016-2019

Co-Chair 11th International Workshop on Semantic Evaluation (SemEval-2017), 10th International Workshop on Semantic Evaluation (SemEval-2016), 9th International Workshop on Semantic Evaluation (SemEval 2015)

Publication Shepherd 9th International Workshop on Semantic Evaluation (SemEval 2015)

SOFTWARE

Google Universal Sentence Encoder Multilingual sentence embedding models.
Co-origination and development with Yinfei Yang. (2018-present)

Stanford Phrasal Phrase-based machine translation system.
*Originating author of the Phrasal decoder, decoder features and decoding model training code.
Co-development of the Phrasal toolkit with Michel Galley and Spence Green. (2007-2013)*

PDP++ Neural-network simulation system for computational cognitive neuroscience.
Originating author of the distributed multi-host neural network model training routines for PDP++, and support for activation functions based on a multi-dimensional lookup table for use in better simulating biological neurons. (2003)

SONIC Large vocabulary speech recognition platform.
Originating author of the distributed multi-host acoustic model training routines for Sonic. (2003)

COMMUNITY
EDUCATIONAL
OUTREACH

Brief Introduction to Natural Language Processing. Daniel Cer. Silicon Valley Code Camp. 2010

PROFESSIONAL
SOCIETIES

Association for Computational Linguistics (ACL)
including Special Interest Group on the Lexicon (ACL-SIGLEX)
Association for Computing Machinery (ACM)
Association for the Advancement of Artificial Intelligence (AAAI)
Linguistic Society of America (LSA)

DATE UPDATED

Aug 22, 2021

REFEREES

Available on request.