

# Kenneth Heafield

to@kheafield.com

<https://kheafield.com>

**RESEARCH INTERESTS** With applications to neural machine translation and language models, my group works on:  
Efficiency at every level of the stack from simpler neural networks to assembly and hardware co-design. Dominating the efficient translation shared task: [neural.mt/speed/](https://neural.mt/speed/).  
Big Open Data: Finding translations in petabytes of the web [paracrawl.eu](https://paracrawl.eu).  
Large Language Models: Training using millions of GPU hours [hplt-project.org](https://hplt-project.org).  
Low-resource: Exploiting additional types of data and inductive biases like trees.  
Analysis: Understanding transfer learning, measuring information in translation, and bias.  
I work closely with companies on funded projects and have my own 4-person company.

**EDUCATION** **PhD, Carnegie Mellon** 8/2008–9/2013  
Efficient Language Modeling Algorithms with Applications to Statistical Machine Translation  
Advised by Alon Lavie, Language Technologies Institute in the School of Computer Science.  
Created KenLM [neural.mt/code/kenlm/](https://neural.mt/code/kenlm/), an efficient  $N$ -gram language modeling toolkit.  
Trained a language model on a trillion words in 2013.

**Bachelor of Science, Caltech** 9/2003–3/2007  
Double major in Mathematics and Computer Science, with honors.

**EXPERIENCE** **Reader (Associate Professor), University of Edinburgh** 8/2020–Present  
**Lecturer (Assistant Professor), University of Edinburgh** 8/2015–8/2020  
I lead large grants with multiple partners totalling 35–45 FTE at a time:

*HPLT* Overall lead €1,119,312 local/€5,000,000 total 9/2022–8/2025  
Scaling translation and large language model training to 12 petabytes of web crawl, the entire digital copyright depositories of Finland and Norway, and millions of GPU hours: [hplt-project.org](https://hplt-project.org)  
Edinburgh, Prompsit, Charles University, Oslo, Helsinki, Turku, CESNET, Sigma2

*UTTER* Co-PI €1,188,182 local/€5,258,504 total 10/2022–9/2025  
Making multilingual and multimodal meeting assistants with transcription, translation, and summarization using large language models: [he-utter.eu](https://he-utter.eu)  
Edinburgh, Amsterdam, Unbabel, Instituto de Telecomunicações, Naver Labs Europe

*Bergamot* Overall lead €983,384 local/€2,999,096 total 1/2019–6/2022  
Launched local machine translation in Firefox [browser.mt](https://browser.mt) and local desktop translation software [translatelocally.com](https://translatelocally.com) using fast translation  
Edinburgh, Charles University, Sheffield, Tartu, Mozilla

*User-Focused Marian* Overall lead €233,074 local/€608,879 total 3/2020–2/2022  
Added run-time domain adaptation, forced translation, and 8-bit GPU support to the Marian machine translation toolkit: [marian-project.eu](https://marian-project.eu)  
Edinburgh, Unbabel, Tilde

*ParaCrawl* Overall lead €869,980 local/€2,383,039 total 9/2017–9/2021  
Created the largest parallel corpus for many languages by mining the web: [paracrawl.eu](https://paracrawl.eu)  
Edinburgh, Alicante, Prompsit, Omniscien, TAUS

*EuroPat* Overall lead €338,075 local/€695,890 total 9/2019–9/2021  
Mined patents for translations and released them for free: [europat.net](https://europat.net)  
Edinburgh, Omniscien, Prompsit

*SCRIPTS* Site co-PI £1,014,230 local/\$13,930,306 total 9/2017–2/2022  
Made cross-lingual search systems for low-resource languages in speech and text  
Columbia, Maryland, Yale, Edinburgh, Cambridge

Industry collaboration funded by Amazon, eBay, Facebook, Google, Huawei, Intel, Microsoft, Mozilla, and Oracle. Also faculty lead on industry engagement in our NLP PhD program.

**Director, Efficient Translation Limited** 8/2019–Present  
My four-person company commercializes efficient machine translation. Our clients (two large companies, one small company, and a government) use this for large-scale server-based translation or locally on desktops to preserve privacy.

**Research Scientist, Meta (Contract Worker via PRO)** 1/2021–10/2022  
Worked 20% time on the No Language Left Behind project, which scaled machine translation to 200 languages: [arxiv.org/abs/2207.04672](https://arxiv.org/abs/2207.04672). Contributed to data mining and efficiency.

**Senior Research Scientist, Bloomberg** 8/2014–7/2015  
Machine translation lead. Created new models and investigated hardware acceleration for publication. Rewrote academic funding call. Solicited proposals, reviewed, and advocated funding for research proposals from universities.

**Postdoctoral Scholar, Stanford** 10/2013–7/2014  
Responsible for machine translation efforts at Stanford, including supervising two PhD students and four Master’s students. Research included web-scale text processing, algorithms for machine translation, and applications of neural networks.

**Research Associate, University of Edinburgh** 8/2011–12/2011; 8/2012–9/2013  
Created an efficient search algorithm for syntactic machine translation, made language model estimation efficient, contributed to the Moses machine translation system, and informally advised PhD students.

**Software Engineer, Google** 3/2007–8/2008  
Optimized language classification for card catalog information about books as part of the Google Books team. Created the ranking function for a search system in Picasa Web Albums. Lectured at MIT about Hadoop.

**Intern, Infosys Technologies** 7/2006–9/2006  
Travelled to Bangalore, India for an internship with the Software Engineering Technology Lab. Applied latent Dirichlet allocation to automatically organize source code.

**Undergraduate Researcher, Netlab at Caltech** 6/2005–6/2006  
Developed an error model for kernel principal component analysis (kPCA) and applied it to automatically analyze computer network traffic and flag possible attacks.

**Undergraduate Researcher, Galaxy Evolution Explorer** 6/2004–3/2007  
Found variable stars by creating a database of 193 million measurements from a satellite.

GRANTS	High Performance Language Technologies <i>Coordinator</i>	EU: €5,000,000 2022
	Unified Transcription and Translation for Extended Reality	EU: €5,258,504 2022
	Intel-Edinburgh Collaboration on Neural Machine Translation	Intel: \$170,000 2022
	Faster Multilingual Sentence Embeddings	Huawei: £135,000 2022
	Scaling fast translation to many languages	EPSRC: 190k GPU hours 2022
	Intel-Edinburgh Collaboration on Neural Machine Translation	Intel: \$175,000 2021
	Machine Translation Shared Task Efficiency Research Competition	Oracle Cloud: \$20,000 2021

Intel-Edinburgh Collaboration on Neural Machine Translation	Intel: \$150,000	2020
Training Machine Translation on the Web	EPSRC: 1 PB disk, 10m CPU, 250k GPU hours	2020
User-Focused Marian <i>Coordinator</i>	EU: €608,879	2020
Mining the Web for Translations	EPSRC: 1 PB disk, 3m CPU, 30k KNL, 50k GPU hours	2020
Efficient Distributed Machine Translation	EPSRC: 200k GPU hours	2020
Efficient Hardware-level Neural Network Inference	Intel: \$50,000	2019
Machine Translation Marathon 2019	EAMT: €4700	2019
GPUs for Machine Translation Marathon 2019	Google Cloud: \$5000	2019
Catering for Machine Translation Marathon 2019	Microsoft: £2,500	2019
What is transfer learning transferring?	eBay: \$30,000	2019
EuroPat: Unleashing European Patent Translations <i>Coordinator</i>	EU: €695,890	2019
Continued Web-Scale Provision of Parallel Corpora for European Languages <i>Coordinator</i>	EU: €889,649	2019
UKRI Centre for Doctoral Training in Natural Language Processing <i>Industry Lead</i>	EPSRC: £6,530,681	2019
Mining non-English Parallel Corpora	eBay: \$30,000	2018
Faster Marian on the CPU	eBay: \$30,000	2018
Compression in Distributed Machine Translation	EPSRC: 150,000 GPU hours	2018
Bergamot: Browser-based Multilingual Translation <i>Coordinator</i>	EU: €2,999,096	2018
Broader Web-Scale Provision of Parallel Corpora for European Languages	EU: €907,976	2018
ParaCrawl	EPCC: 500,000 CPU hours	2018
Fast Neurons on Xeons	Intel: £28,302	2018
Distributed Machine Translation	EPSRC: 150,000 GPU hours	2017
SCRIPTS: System for cross-language information processing, translation, and summarization	IARPA: \$13,930,306	2017
Provision of web-scale parallel corpora for official European languages	EU: €585,414	2017
Medical machine translation	EPSRC: £51,993	2017
Scalable recurrent neural networks	US DOE: 9,000,000 supercomputer hours	2017
Making web crawl a Turing resource	Alan Turing Institute: £26,137	2017
Cloud computing for MSc dissertations	Google Cloud Credits: \$17,500	2017
Open data: mining translations and transcripts from the web	Mozilla: \$64,143	2017
Training neural machine translation systems	Microsoft Azure Credits: \$1,164,244	2017
Mining and training on translations from the web	Microsoft Azure Credits: \$20,000	2017

Neural network primitives and distributed training	Intel: £28,302 2017
Phrase-based decoding	eBay: \$30,000 2016
Decoding methods for spelling correction and synonym generation	Facebook: \$50,000 2016
Local coarse-to-fine decoding for long-distance models	Google: \$57,724 2015
Faster decoding and better features via local coarse-to-fine	Amazon: \$70,949 2015
Faster machine translation and principled rule learning	Bloomberg: \$150,000 2014
Applying tera-scale language models to advance machine translation	NSF: 200,000 CPU hours 2013
Graduate Research Fellowship	NSF: \$121,500 2008

GRANTS *At Bloomberg, I rewrote their grant program and made the case for several gifts in 2015.*

CHAMPIONED Christopher Manning. Natural Language Processing and Machine Learning. \$75,000.  
 Shay Cohen. Latent-Variable Learning for Transition-Based Parsing. \$63,379.  
 Philipp Koehn. High Quality Parallel Corpus Extraction from the Web. \$50,000.  
 Lane Schwartz. US Machine Translation Marathon at UIUC. \$10,000.  
 Fei-Fei Li. SAILORS AI summer outreach. \$10,000.

STUDENTS Graduated PhDs as main supervisor:  
 Maximiliana Behnke → Deep Learning R&D Engineer, Intel  
 Anna Currey → Senior Applied Scientist, Amazon  
 Alham Fikri Aji → Assistant Professor, MBZUAI  
 Graduated PhDs as assistant supervisor:  
 Nikolay Bogoychev → Research Associate, Edinburgh  
 PhD as main supervisor: Laurie Burchell, Pinzhen Chen, Proyag Pal  
 PhD as assistant supervisor: Amna Shahab, Naums Mogers  
 PhD examiner: Dominik Wurzer (2017), Duygu Ataman (2019), Mohiuddin Muhammad Tasnim (2022)  
 MPhil (with Miles Osborne): Luke Shrimpton, 2016  
 MSc: supervised 14 theses

CONFERENCE *Download from [neural.mt/papers/](https://neural.mt/papers/).*

PAPERS Laurie V. Burchell, Alexandra Birch, Nikolay Bogoychev, and **Kenneth Heafield**. An Open Model and Dataset for Language Identification. *The 61st Annual Meeting of the Association for Computational Linguistics: ACL 2023*, Toronto, Canada, July, 2023.

Proyag Pal and **Kenneth Heafield**. Cheating to Identify Hard Problems for Neural Machine Translation. *Findings of the Association for Computational Linguistics: EACL 2023*, Dubrovnik, Croatia, May, 2023.

Pinzhen Chen and **Kenneth Heafield**. Approaching Neural Chinese Word Segmentation as a Low-Resource Machine Translation Task. *36th Pacific Asia Conference on Language, Information and Computation*, Online, October, 2022.

Elsbeth Turcan, David Wan, Faisal Ladhak, Petra Galuscakova, Sukanta Sen, Svetlana Tchistiakova, Weijia Xu, Marine Carpuat, **Kenneth Heafield**, Douglas Oard, and Kathleen McKeeown. Constrained Regeneration for Cross-Lingual Query-Focused Extractive Summarization. *29th International Conference on Computational Linguistics*, Gyeongju, Republic of Korea, October, 2022.

Proyag Pal and **Kenneth Heafield**. Cheat Codes to Quantify Missing Source Information in Neural Machine Translation. *2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, Seattle, Washington, July, 2022.

**Kenneth Heafield**, Elaine Farrow, Jelmer van der Linde, Gema Ramírez-Sánchez, and Dion Wiggins. The EuroPat Corpus: A Parallel Corpus of European Patent Data. *13th Language Resources and Evaluation Conference*, Marseille, France, June, 2022.

Nikolay Bogoychev, Jelmer van der Linde, and **Kenneth Heafield**. TranslateLocally: Blazing-fast translation running on the local CPU. *2021 Conference on Empirical Methods in Natural Language Processing*, Punta Cana, Dominican Republic, November, 2021.

Farhad Akhbardeh, Arkady Arkhangorodsky, Magdalena Biesialska, Ondřej Bojar, Rajen Chatterjee, Vishrav Chaudhary, Marta R. Costa-jussa, Cristina España-Bonet, Angela Fan, Christian Federmann, Markus Freitag, Yvette Graham, Roman Grundkiewicz, Barry Haddow, Leonie Harter, **Kenneth Heafield**, Christopher Homan, Matthias Huck, Kwabena Amponsah-Kaakyire, Jungo Kasai, Daniel Khashabi, Kevin Knight, Tom Kocmi, Philipp Koehn, Nicholas Lourie, Christof Monz, Makoto Morishita, Masaaki Nagata, Ajay Nagesh, Toshiaki Nakazawa, Matteo Negri, Santanu Pal, Allahsera Auguste Tapo, Marco Turchi, and Valentin Vydrin and Marcos Zampieri. Findings of the 2021 Conference on Machine Translation (WMT21). *Sixth Conference on Machine Translation*, Punta Cana, Dominican Republic, November, 2021.

**Kenneth Heafield**, Qianqian Zhu, and Roman Grundkiewicz. Findings of the WMT 2021 Shared Task on Efficient Translation. *Conference on Machine Translation at the 2021 Conference on Empirical Methods in Natural Language Processing*, Punta Cana, Dominican Republic, November, 2021.

Maximiliana Behnke, Nikolay Bogoychev, Alham Fikri Aji, **Kenneth Heafield**, Graeme Nail, Qianqian Zhu, Svetlana Tchistiakova, Jelmer van der Linde, Pinzhen Chen, Sidharth Kashyap, and Roman Grundkiewicz. Efficient Machine Translation with Model Pruning and Quantization. *Conference on Machine Translation at the 2021 Conference on Empirical Methods in Natural Language Processing*, Punta Cana, Dominican Republic, November, 2021.

Pinzhen Chen, Jindřich Helcl, Ulrich Germann, Laurie Burchell, Nikolay Bogoychev, Antonio Valerio Miceli Barone, Jonas Waldendorf, Alexandra Birch, and **Kenneth Heafield**. The University of Edinburgh’s English-German and English-Hausa Submissions to the WMT21 News Translation Task. *Conference on Machine Translation at the 2021 Conference on Empirical Methods in Natural Language Processing*, Punta Cana, Dominican Republic, November, 2021.

Maximiliana Behnke and **Kenneth Heafield**. Pruning Neural Machine Translation for Speed Using Group Lasso. *Sixth Conference on Machine Translation*, Punta Cana, Dominican Republic, November, 2021.

Adithya Renduchintala, Denise Diaz, **Kenneth Heafield**, Xian Li, and Mona Diab. Gender Bias Amplification During Speed-Quality Optimization in Neural Machine Translation. *59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 2: Short Papers)*, Online, August, 2021.

Maximiliana Behnke and **Kenneth Heafield**. Losing Heads in the Lottery: Pruning Transformer Attention in Neural Machine Translation. *2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, Online, November, 2020.

Tobias Domhan, Michael Denkowski, David Vilar, Xing Niu, Felix Hieber, and **Kenneth Heafield**. The Sockeye 2 Neural Machine Translation Toolkit at AMTA 2020. *14th Con-*

*ference of the Association for Machine Translation in the Americas (AMTA 2020)*, Virtual, October, 2020.

Alham Fikri Aji, Nikolay Bogoychev, **Kenneth Heafield**, and Rico Sennrich. In *Neural Machine Translation, What Does Transfer Learning Transfer? 58th Annual Meeting of the Association for Computational Linguistics*, Online, July, 2020.

Pinzhen Chen, Nikolay Bogoychev, **Kenneth Heafield**, and Faheem Kirefu. *Parallel Sentence Mining by Constrained Decoding. 58th Annual Meeting of the Association for Computational Linguistics*, Online, July, 2020.

Marta Bañón, Pinzhen Chen, Barry Haddow, **Kenneth Heafield**, Hieu Hoang, Miquel Esplà-Gomis, Mikel L. Forcada, Amir Kamran, Faheem Kirefu, Philipp Koehn, Sergio Ortiz Rojas, Leopoldo Pla Sempere, Gema Ramírez-Sánchez, Elsa Sarrías, Marek Strelec, Brian Thompson, William Waites, Dion Wiggins, and Jaume Zaragoza. *ParaCrawl: Web-Scale Acquisition of Parallel Corpora. 58th Annual Meeting of the Association for Computational Linguistics*, Online, July, 2020.

Alham Fikri Aji, **Kenneth Heafield**, and Nikolay Bogoychev. *Combining Global Sparse Gradients with Local Gradients in Distributed Neural Network Training. 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, Hong Kong, November, 2019.

Anna Currey and **Kenneth Heafield**. *Multi-Source Syntactic Neural Machine Translation. 2018 Conference on Empirical Methods in Natural Language Processing*, Brussels, Belgium, November, 2018.

Nikolay Bogoychev, Marcin Junczys-Dowmunt, **Kenneth Heafield**, and Alham Fikri Aji. *Accelerating Asynchronous Stochastic Gradient Descent for Neural Machine Translation. 2018 Conference on Empirical Methods in Natural Language Processing*, Brussels, Belgium, November, 2018.

Marcin Junczys-Dowmunt, Roman Grundkiewicz, Tomasz Dwojak, Hieu Hoang, **Kenneth Heafield**, Tom Neckermann, Frank Seide, Ulrich Germann, Alham Fikri Aji, Nikolay Bogoychev, André F. T. Martins, and Alexandra Birch. *Marian: Fast Neural Machine Translation in C++. Proceedings of ACL 2018, System Demonstrations*, Melbourne, Australia, July, 2018.

Marcin Junczys-Dowmunt, Roman Grundkiewicz, Shubha Guha, and **Kenneth Heafield**. *Approaching Neural Grammatical Error Correction as a Low-Resource Machine Translation Task. 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers)*, New Orleans, Louisiana, June, 2018.

Alham Fikri Aji and **Kenneth Heafield**. *Sparse Communication for Distributed Gradient Descent. 2017 Conference on Empirical Methods in Natural Language Processing*, Copenhagen, Denmark, September, 2017.

**Kenneth Heafield**, Chase Geigle, Sean Massung, and Lane Schwartz. *Normalized Log-Linear Language Model Interpolation is Efficient. 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, Berlin, Germany, August, 2016.

**Kenneth Heafield**, Rohan Kshirsagar, and Santiago Barona. *Language Identification and Modeling in Specialized Hardware. 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 2: Short Papers)*, Beijing, China, July, 2015.

**Kenneth Heafield**, Michael Kayser, and Christopher D. Manning. *Faster Phrase-Based Decoding by Refining Feature State. 52nd Annual Meeting of the Association for Computational*

*Linguistics (Volume 2: Short Papers)*, Baltimore, MD, USA, June, 2014.

Christian Buck, **Kenneth Heafield**, and Bas van Ooyen. N-gram Counts and Language Models from the Common Crawl. *Ninth International Conference on Language Resources and Evaluation (LREC'14)*, Reykjavik, Iceland, May, 2014.

**Kenneth Heafield**, Ivan Pouzyrevsky, Jonathan H. Clark, and Philipp Koehn. Scalable Modified Kneser-Ney Language Model Estimation. *51st Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, Sofia, Bulgaria, August, 2013.

**Kenneth Heafield**, Philipp Koehn, and Alon Lavie. Grouping Language Model Boundary Words to Speed K-Best Extraction from Hypergraphs. *2013 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, Atlanta, Georgia, USA, June, 2013.

**Kenneth Heafield**, Philipp Koehn, and Alon Lavie. Language Model Rest Costs and Space-Efficient Storage. *2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning*, Jeju Island, Korea, July, 2012.

**Kenneth Heafield** and Alon Lavie. Voting on N-grams for Machine Translation System Combination. *Ninth Conference of the Association for Machine Translation in the Americas*, Denver, Colorado, USA, November, 2010.

Girish Maskeri, Santonu Sarkar, and **Kenneth Heafield**. Mining Business Topics in Source Code using Latent Dirichlet Allocation. *1st India Software Engineering Conference*, Hyderabad, India, February, 2008. *10-year test of time award*.

Stanley Browne, Jonathan Wheatley, Barry Welsh, Mark Seibert, **Kenneth Heafield**, R. Michael Rich, and the GALEX Science Team. RR Lyrae Stars in the Far Ultraviolet: GALEX Observations Compared with Theoretical Predictions. *American Astronomical Society 207th Meeting*, Washington, DC, USA, June, 2006.

Barry Welsh, Jonathan Wheatley, **Kenneth Heafield**, Mark Seibert, Stanley Browne, and the GALEX Science Team. The Flaring UV Sky. *American Astronomical Society 205th Meeting*, San Diego, California, USA, January, 2005.

REFEREED  
WORKSHOP  
PAPERS

Nikolay Bogoychev, Maximiliana Behnke, Jelmer Van Der Linde, Graeme Nail, **Kenneth Heafield**, Biao Zhang, and Sidharth Kashyap. Edinburgh's Submission to the WMT 2022 Efficiency Task. *Seventh Conference on Machine Translation (WMT)*, Abu Dhabi, December, 2022.

**Kenneth Heafield**, Biao Zhang, Graeme Nail, Jelmer Van Der Linde, and Nikolay Bogoychev. Findings of the WMT 2022 Shared Task on Efficient Translation. *Seventh Conference on Machine Translation (WMT)*, Abu Dhabi, December, 2022.

Laurie Burchell, Alexandra Birch, and **Kenneth Heafield**. Exploring Diversity in Back Translation for Low-Resource Machine Translation. *Third Workshop on Deep Learning for Low-Resource Natural Language Processing*, Seattle, Washington, July, 2022.

Ulrich Germann, Roman Grundkiewicz, Martin Popel, Radina Dobreva, Nikolay Bogoychev, and **Kenneth Heafield**. Speed-optimized, Compact Student Models that Distill Knowledge from a Larger Teacher Model: the UEDIN-CUNI Submission to the WMT 2020 News Translation Task. *Fifth Conference on Machine Translation*, Online, November, 2020.

Alham Fikri Aji and **Kenneth Heafield**. Compressing Neural Machine Translation Models with 4-bit Precision. *Fourth Workshop on Neural Generation and Translation*, Online, July, 2020.

Nikolay Bogoychev, Roman Grundkiewicz, Alham Fikri Aji, Maximiliana Behnke, **Kenneth Heafield**, Sidharth Kashyap, Emmanouil-Ioannis Farsarakis, and Mateusz Chudyk. Edinburgh’s Submissions to the 2020 Machine Translation Efficiency Task. *Fourth Workshop on Neural Generation and Translation*, Online, July, 2020.

**Kenneth Heafield**, Hiroaki Hayashi, Yusuke Oda, Ioannis Konstas, Andrew Finch, Graham Neubig, Xian Li, and Alexandra Birch. Findings of the Fourth Workshop on Neural Generation and Translation. *Fourth Workshop on Neural Generation and Translation*, Online, July, 2020.

Young Jin Kim, Marcin Junczys-Dowmunt, Hany Hassan, Alham Fikri Aji, **Kenneth Heafield**, Roman Grundkiewicz, and Nikolay Bogoychev. From Research to Production and Back: Ludicrously Fast Neural Machine Translation. *3rd Workshop on Neural Generation and Translation*, Hong Kong, November, 2019.

Anna Currey and **Kenneth Heafield**. Zero-Resource Neural Machine Translation with Monolingual Pivot Data. *3rd Workshop on Neural Generation and Translation*, Hong Kong, November, 2019.

Alham Fikri Aji and **Kenneth Heafield**. Making Asynchronous Stochastic Gradient Descent Work for Transformers. *3rd Workshop on Neural Generation and Translation*, Hong Kong, November, 2019.

Anna Currey and **Kenneth Heafield**. Incorporating Source Syntax into Transformer-Based Neural Machine Translation. *Fourth Conference on Machine Translation (Volume 1: Research Papers)*, Florence, Italy, August, 2019.

Roman Grundkiewicz, Marcin Junczys-Dowmunt, and **Kenneth Heafield**. Neural Grammatical Error Correction Systems with Unsupervised Pre-training on Synthetic Data. *Fourteenth Workshop on Innovative Use of NLP for Building Educational Applications*, Florence, Italy, August, 2019.

Douglas Oard, Petra Galuscakova, Kathleen McKeown, Marine Carpuat, Ramy Eskander, **Kenneth Heafield**, Efsun Kayi, Chris Kedzie, Smaranda Muresan, Suraj Nair, Xing Niu, Dragomir Radev, Anton Ragni, Han-Chin Shing, Yan Virin, Weijia Xu, Rui Zhang, Elena Zotkina, Joseph Barrow, and Mark Gales. Surprise Languages: Rapid-Response Cross-Language IR. *The Ninth International Workshop on Evaluating Information Access (EVIA 2019)*, Tokyo, Japan, June, 2019.

Philipp Koehn, Huda Khayrallah, **Kenneth Heafield**, and Mikel L. Forcada. Findings of the WMT 2018 Shared Task on Parallel Corpus Filtering. *Third Conference on Machine Translation: Shared Task Papers*, Brussels, Belgium, October, 2018.

Barry Haddow, Nikolay Bogoychev, Denis Emelin, Ulrich Germann, Roman Grundkiewicz, **Kenneth Heafield**, Antonio Valerio Miceli Barone, and Rico Sennrich. The University of Edinburgh’s Submissions to the WMT18 News Translation Task. *Third Conference on Machine Translation: Shared Task Papers*, Brussels, Belgium, October, 2018.

Marcin Junczys-Dowmunt, **Kenneth Heafield**, Hieu Hoang, Roman Grundkiewicz, and Anthony Aue. Marian: Cost-effective High-Quality Neural Machine Translation in C++. *2nd Workshop on Neural Machine Translation and Generation*, Melbourne, Australia, July, 2018.

Roman Grundkiewicz and **Kenneth Heafield**. Neural Machine Translation Techniques for Named Entity Transliteration. *Seventh Named Entities Workshop*, Melbourne, Australia, July, 2018.

Hieu Hoang, Tomasz Dwojak, Rihards Krislauks, Daniel Torregrosa, and **Kenneth Heafield**.



Fast Neural Machine Translation Implementation. *2nd Workshop on Neural Machine Translation and Generation*, Melbourne, Australia, July, 2018.

Anna Currey and **Kenneth Heafield**. Unsupervised Source Hierarchies for Low-Resource Neural Machine Translation. *Workshop on the Relevance of Linguistic Structure in Neural Architectures for NLP*, Melbourne, Australia, July, 2018.

Anna Currey, Antonio Valerio Miceli Barone, and **Kenneth Heafield**. Copied Monolingual Data Improves Low-Resource Neural Machine Translation. *Second Conference on Machine Translation*, Copenhagen, Denmark, September, 2017.

Rico Sennrich, Alexandra Birch, Anna Currey, Ulrich Germann, Barry Haddow, **Kenneth Heafield**, Antonio Valerio Miceli Barone, and Philip Williams. The University of Edinburgh’s Neural MT Systems for WMT17. *Second Conference on Machine Translation*, Copenhagen, Denmark, September, 2017.

Nadir Durrani, Barry Haddow, Philipp Koehn, and **Kenneth Heafield**. Edinburgh’s Phrase-based Machine Translation Systems for WMT-14. *Ninth Workshop on Statistical Machine Translation*, Baltimore, MD, USA, June, 2014.

Julia Neidert, Sebastian Schuster, Spence Green, **Kenneth Heafield**, and Christopher D. Manning. Stanford University’s Submissions to the WMT 2014 Translation Task. *Ninth Workshop on Statistical Machine Translation*, Baltimore, MD, USA, June, 2014.

Nadir Durrani, Barry Haddow, **Kenneth Heafield**, and Philipp Koehn. Edinburgh’s Machine Translation Systems for European Language Pairs. *Eighth Workshop on Statistical Machine Translation*, Sofia, Bulgaria, August, 2013.

**Kenneth Heafield**, Hieu Hoang, Philipp Koehn, Tetsuo Kiso, and Marcello Federico. Left Language Model State for Syntactic Machine Translation. *International Workshop on Spoken Language Translation*, San Francisco, California, USA, December, 2011.

**Kenneth Heafield**. KenLM: Faster and Smaller Language Model Queries. *Sixth Workshop on Statistical Machine Translation*, Edinburgh, Scotland, United Kingdom, July, 2011.

**Kenneth Heafield** and Alon Lavie. CMU System Combination in WMT 2011. *Sixth Workshop on Statistical Machine Translation*, Edinburgh, Scotland, United Kingdom, July, 2011.

**Kenneth Heafield** and Alon Lavie. CMU Multi-Engine Machine Translation for WMT 2010. *Joint Fifth Workshop on Statistical Machine Translation and MetricsMATR*, Uppsala, Sweden, July, 2010.

**Kenneth Heafield**, Greg Hanneman, and Alon Lavie. Machine Translation System Combination with Flexible Word Ordering. *Fourth Workshop on Statistical Machine Translation*, Athens, Greece, March, 2009.

JOURNAL  
ARTICLES

Marcos Treviso, Ji-Ung Lee, Tianchu Ji, Betty van Aken, Qingqing Cao, Manuel R. Ciosici, Michael Hassid, **Kenneth Heafield**, Sara Hooker, Colin Raffel, Pedro H. Martins, André F. T. Martins, Jessica Zosa Forde, Peter Milder, Edwin Simpson, Noam Slonim, Jesse Dodge, Emma Strubell, Niranjan Balasubramanian, Leon Derczynski, Iryna Gurevych, and Roy Schwartz. Efficient Methods for Natural Language Processing: A Survey. *Transactions of the Association for Computational Linguistics*. March, 2023.

**Kenneth Heafield** and Alon Lavie. Combining Machine Translation Output with Open Source: The Carnegie Mellon Multi-Engine Machine Translation Scheme. *The Prague Bulletin of Mathematical Linguistics* 93. January, 2010.

Jonathan H. Clark, Jonathan Weese, Byung Gyu Ahn, Andreas Zollmann, Qin Gao, **Kenneth Heafield**, and Alon Lavie. The Machine Translation Toolpack for LoonyBin: Automated Management of Experimental Machine Translation HyperWorkflows. *The Prague Bulletin of Mathematical Linguistics* 93. January, 2010.

Barry Welsh, Johathan Wheatley, **Kenneth Heafield**, Mark Seibert, and the GALEX Science Team. The GALEX Ultraviolet Variability Catalog. *The Astronomical Journal* 130. 2005.

ARXIV NLLB Team, Marta R. Costa-jussà, James Cross, Onur Çelebi, Maha Elbayad, **Kenneth Heafield**, Kevin Heffernan, Elahe Kalbassi, Janice Lam, Daniel Licht, Jean Maillard, Anna Sun, Skyler Wang, Guillaume Wenzek, Al Youngblood, Bapi Akula, Loic Barrault, Gabriel Mejia Gonzalez, Prangthip Hansanti, John Hoffman, Semarley Jarrett, Kaushik Ram Sadagopan, Dirk Rowe, Shannon Spruit, Chau Tran, Pierre Andrews, Necip Fazil Ayan, Shruti Bhosale, Sergey Edunov, Angela Fan, Cynthia Gao, Vedanuj Goswami, Francisco Guzmán, Philipp Koehn, Alexandre Mourachko, Christophe Ropers, Safiyyah Saleem, Holger Schwenk, and Jeff Wang. No Language Left Behind: Scaling Human-Centered Machine Translation. arXiv preprint 2207.04672. July, 2022.

PATENTS Girish Maskeri Rama, **Kenneth Heafield**, and Santonu Sarkar. Identification of Topics in Source Code. US Patent 8209665 filed in 2009 and issued June, 2012.

Taylor Curtis and **Kenneth Heafield**. Systems and Methods for Identifying Similar Documents. US Patent 7958136 filed in 2008 and issued June, 2011.

SELECTED PRESS TechCrunch: Mozilla brings free, offline translation to Firefox 2022  
[techcrunch.com/2022/06/02/mozilla-brings-free-offline-translation-to-firefox](https://techcrunch.com/2022/06/02/mozilla-brings-free-offline-translation-to-firefox)

CNET: Meta’s ‘No Language Left Behind’ AI Can Now Translate 200 Languages 2022  
[cnet.com/news/social-media/metas-no-language-left-behind-ai-can-now-translate-200-languages](https://cnet.com/news/social-media/metas-no-language-left-behind-ai-can-now-translate-200-languages)

Slator: Blazing-Fast Machine Translation With Kenneth Heafield 2021  
[slator.com/blazing-fast-machine-translation-with-kenneth-heafield](https://slator.com/blazing-fast-machine-translation-with-kenneth-heafield)

Slator: Amazon Releases New Iteration of Neural Machine Translation Toolkit Sockeye 2020  
[slator.com/amazon-releases-new-iteration-of-neural-machine-translation-toolkit-sockeye](https://slator.com/amazon-releases-new-iteration-of-neural-machine-translation-toolkit-sockeye)

Trusted Reviews: Firefox is getting language translation and it’s safer than Google Chrome’s 2019  
[trustedreviews.com/news/firefox-translation-bergamot-mozilla-3948613](https://trustedreviews.com/news/firefox-translation-bergamot-mozilla-3948613)

ZDNet: Firefox to get page translation feature, like Chrome 2019  
[zdnet.com/article/firefox-to-get-page-translation-feature-like-chrome/](https://zdnet.com/article/firefox-to-get-page-translation-feature-like-chrome/)

New York Times: Is Translation an Art or a Math Problem? 2015  
[nytimes.com/2015/06/07/magazine/is-translation-an-art-or-a-math-problem.html](https://nytimes.com/2015/06/07/magazine/is-translation-an-art-or-a-math-problem.html)

INVITED TALKS Deep Learning Indaba, Dagstuhl Efficient and Equitable Natural Language Processing in the Age of Deep Learning, Meta-Forum, University of Southern California, Grammarly, European Union Workshop Showcasing Language Technology, eBay, Amazon Berlin, Nanyang Technological University, National University of Singapore, A\*STAR, European Union Workshop on Parallel Corpora, Intel, Microsoft, Alan Turing Institute, European Language Resource Consortium, University of Maryland, Georgetown, Facebook Paris, Google, Facebook Menlo Park, Facebook London, Microsoft, Amazon Seattle, Facebook London, University of Edinburgh, Indiana University, University of Southern California, Bloomberg, Queens College, Google, Apple, Numen, Xerox Research Centre Europe, Qatar Computing Research Institute, Carnegie Mellon–Qatar, Hong Kong University of Science and Technology

DAGSTUHL	<i>Invitation-only small-group meetings in computer science</i>	
	Efficient and Equitable Natural Language Processing in the Age of Deep Learning	2022
	Programming with “Big Code”	2015
ORGANIZER	UKRI Centre for Doctoral Training in NLP - Industry Event <a href="http://edin.ac/3NMbbsG">edin.ac/3NMbbsG</a>	2021, 2022
	WMT Efficiency Shared Task <a href="http://neural.mt/speed/2021">neural.mt/speed/2021</a>	2021, 2022
	ACL Senior Area Chair in Machine Translation <a href="http://2022.aclweb.org">2022.aclweb.org</a>	2022
	ACL Sponsorship Co-Chair in European Region <a href="http://aclweb.org">aclweb.org</a>	2021
	Machine Translation Half-Marathon <a href="http://unbabel.com/lp/events/mt-half-marathon">unbabel.com/lp/events/mt-half-marathon</a>	2021
	WNGT Efficiency Shared Task <a href="http://sites.google.com/view/wngt20/efficiency-task">sites.google.com/view/wngt20/efficiency-task</a>	2020
	Machine Translation Marathon <a href="http://statmt.org/mtm19">statmt.org/mtm19</a>	2019
REVIEWING	ACL, TACL, EMNLP, NAACL, EACL, IJCAI, ICML, AAAI, WMT, COLING, TALIP, MT Journal, CACM	
TUTORIALS	Language Modeling, Machine Translation Marathon	2011–2017
	Language Model Implementation, Machine Translation Marathon	9/2013
	Language Modeling with KenLM, Qatar Computing Research Institute	3/2013
	Chart Based Decoding, Machine Translation Marathon	9/2012
TEACHING	<i>Since 2017, I have been bought out by grants so teaching has been limited to project supervision.</i>	
	Extreme Computing, University of Edinburgh	Fall 2015, Fall 2016, Fall 2017
	Computer Programming Skills and Concepts, University of Edinburgh	Fall 2016
	Guest Course Lecture: Machine Translation, Carnegie Mellon	3/2013
	Guest Course Lecture: Advanced NLP, University of Edinburgh	10/2012
	Teaching Assistant: Language and Statistics, Carnegie Mellon	Spring 2012
	Teaching Assistant: Algorithms for NLP, Carnegie Mellon	Fall 2010
	Lecturer: Introduction to Hadoop, MIT	1/2008
AWARDS	Bloomberg BFIRST	2014
	Bloomberg award for Knowledge Discovery and Data Mining (KDD)	
	Student Travel Grant \$800 in travel funded by the EMNLP conference	2012
	Google Peer Bonus and Site Award For lecturing at MIT on Hadoop while a Software Engineer at Google	2008
	International Collegiate Programming Contest Regional Ranked third of fifty in a team of two instead of three	2006–2007

Carnation Scholarship Year of full Caltech tuition based on academic merit; 38 awarded per year	2005–06
Richard and Dena Krown Summer Undergraduate Research Fellowship \$5,000 for ten weeks of summer research in networking	2005
Summer Undergraduate Research Fellowship \$5,000 for ten weeks of summer research in astronomy data mininug	2004