

## Toan Q. Nguyen

---

CONTACT INFORMATION	355 Fitzpatrick Hall Notre Dame, IN 46556	<a href="mailto:toannq177@gmail.com">toannq177@gmail.com</a> <a href="https://github.com/tnq177">tnq177.github.io</a>
------------------------	--	--

RESEARCH INTERESTS	Neural Machine Translation (with focus on low-resource languages), Deep Learning
-----------------------	--

EDUCATION	<b>University of Notre Dame</b> , Notre Dame, IN Ph.D., Computer Science	Jan 2016 to present
	<ul style="list-style-type: none"><li>• Advisor: David Chiang, Ph.D</li></ul> <b>University of Southampton</b> , Southampton, UK	Sept 2010 to June 2013
	BEng, Electronic Engineering	
	<ul style="list-style-type: none"><li>• <i>First-class honours</i></li></ul>	

RESEARCH PROJECTS	<b>Transformers without Tears: Improving the Normalization of Self-Attention</b> We evaluate three simple, normalization-centric changes to improve Transformer training. Our changes include proper routing of the residual connection, replacing <i>every</i> <b>LayerNormalization</b> with $\ell_2$ normalization with a single scale parameter, and fixing the word embedding norm to 1. Putting together, these modifications allow sharper performance curve, stable training even with warm-up free, large learning rate. For low-resource, we achieve an average of +1.1 BLEU over SOTA baselines. For high-resource, our model performs comparable to the baseline. Additionally, our $\ell_2$ -based normalization uses less computation than <b>LayerNormalization</b> so training is faster (roughly 5%) (paper accepted to <i>IWSLT'19</i> ).
----------------------	--

### Improving lexical choice in NMT

We propose two solutions to alleviate the mistranslation for rare words issue in NMT. First, we argue that the output layer, which computes the inner product between the hidden state with all target word embeddings, rewards frequent words disproportionately. So we propose to fix all target word embedding to a certain value. Second, we integrate a simple lexical module which is jointly learned with NMT and whose output is used to bias NMT's prediction. Our experiments on 8 different language pairs show improvements of from 1.1 to 4.3 BLEUs (paper accepted to *NAACL'18*).

### Transfer Learning across Low-resource, Related Languages for NMT

Using Byte-Pair-Encoding, we generate data with overlapped vocabulary between low-resource, related languages. This data is then used to train a parent model on one language pair, and the trained model is later used to initialize training on another language pair. Our method shows improvements of up to 4.3 BLEUs over a strong BPE baseline for Uyghur-English (paper accepted to *IJCNLP'17*).

RESEARCH EXPERIENCE	<b>Research Assistant</b>	Jan 2016 - present
	Natural Language Processing Group Department of Computer Science and Engineering University of Notre Dame Supervisor: David Chiang	
	<b>Research Intern</b>	Jun 2020 - Sept 2020
	Google Translate team Host: Colin Cherry & George Foster	
	<b>Applied Scientist Intern</b>	Jun 2019 - Sept 2019
	Amazon AWS AI Transcribe Science team Mentor: Julian Salazar Manager: Katrin Kirchhoff	
	<b>Visiting Student</b>	Sep 2018 - Dec 2018
	CILVR Group Center for Data Science New York University Supervisor: Kyunghyun Cho	
TEACHING EXPERIENCE	<b>Teaching Assistant</b>	Fall 2016
	CSE 40868/60868 - Introduction to Neural Networks Instructor: Adam Czajka, Ph.D Division of Computer Science and Engineering, University of Notre Dame	
	<b>Teaching Assistant</b>	Spring 2016
	CSE 40535/60535 - Computer Vision Instructor: Adam Czajka, Ph.D Division of Computer Science and Engineering, University of Notre Dame	
PROFESSIONAL EXPERIENCE	<b>Software Developer</b> , East Agile	Mar 2015 to November 2015
	Develop web-based application to convert complicated Google Analytics information to simple and meaningful reports with focus on small businesses.	
	<b>Software Developer</b> , Tripchum	August 2014 to March 2015
	Develop and deploy multiple experimental tools such as Hubot for Google Analytics information analysis and notification on Hipchat, Twitter Direct Message integration with in-house chatting and logging system.	
PROGRAMMING SKILLS	Proficient: Python, Pytorch, Tensorflow	
	Basic: C, C++, Javascript, Django, NodeJS, OpenCV	
PUBLICATIONS	1. Julian Salazar, Davis Liang, <b>Toan Q. Nguyen</b> , and Katrin Kirchhoff. Masked language model scoring. In <i>Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics</i> , pages 2699–2712, Online, July 2020. Association for Computational Linguistics	
	2. <b>Toan Q. Nguyen</b> and Julian Salazar. Transformers without tears: Improving the normalization of self-attention. In <i>Proceedings of the International Workshop on Spoken Language Translation</i> , 2019	

3. Kenton Murray, Jeffery Kinnison, **Toan Q. Nguyen**, Walter Scheirer, and David Chiang. Auto-sizing the transformer network: Improving speed, efficiency, and performance for low-resource machine translation. In *Proceedings of the 3rd Workshop on Neural Generation and Translation*, pages 231–240, Hong Kong, November 2019. Association for Computational Linguistics
4. Antonios Anastasopoulos, Alison Lui, **Toan Q. Nguyen**, and David Chiang. Neural machine translation of text from non-native speakers. In *Proc. NAACL HLT*, 2019
5. **Toan Q. Nguyen** and David Chiang. Improving lexical choice in neural machine translation. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers)*, pages 334–343. Association for Computational Linguistics, 2018
6. **Toan Q. Nguyen** and David Chiang. Transfer learning across low-resource, related languages for neural machine translation. In *Proceedings of the Eighth International Joint Conference on Natural Language Processing (Volume 2: Short Papers)*, pages 296–301, Taipei, Taiwan, November 2017. Asian Federation of Natural Language Processing
7. Leon Cheung, Thamme Gowda, Ulf Hermjakob, Nelson H S Liu, Jonathan May, Alexandra Mayn, Nima Pourdamghani, Michael Pust, Kevin Knight, Nikolaos Malandrakis, Pavlos Papadopoulos, Anil Ramakrishna, Karan Singla, Víctor Martínez, Colin Vaz, Doan Can, S. Narayanan, Kenton Murray, **Toan Nguyen**, David Chiang, Xiaoman Pan, Boliang Zhang, Ying Chuan Lin, Di Lu, Lifu Huang, Kevin Blissett, Tongtao Zhang, Ondrej Glembek, Murali Karthick Baskar, Santosh Kesiraju, Lukás Burget, Karel Benes, Igor Szoke, Karel Veselý, Camille Goudeseune, M. H. Johnson, Leda Sari, Wenda Chen, and Angli Liu. Elisa system description for lorehlt 2017. 2017

#### AWARDS

- Outstanding Graduate Student Teacher Award, Notre Dame Learning | Kaneb Center, Spring 2021
- Outstanding Graduate Teaching Assistant, University of Notre Dame, Department of Computer Science and Engineering, Spring 2017
- [The Vietnam Education Foundation Fellowship](#), Cohort 2015
- GD Sims Prize, University of Southampton, School of Electronics and Computer Science, 2011
- Honourable Mention, Asian Physics Olympiad, Ulan Bator, Mongolia, 2008

#### REFERENCES

David Chiang  
 Associate Professor  
 Department of Computer Science and Engineering  
 University of Notre Dame  
 Phone: 574-631-9441  
 E-mail: dchiang@nd.edu

Adam Czajka  
 Assistant Professor  
 Department of Computer Science and Engineering  
 University of Notre Dame  
 Phone: 574-631-7072  
 E-mail: aczajka@nd.edu