

**Antonios Anastasopoulos**  
<http://cs.cmu.edu/~aanastas>

---

CONTACT INFORMATION	Language Technologies Institute 5721 Gates Hillman Center Pittsburgh, PA 15213	Carnegie Mellon University 574-993-1434 <a href="mailto:aanastas@andrew.cmu.edu">aanastas@andrew.cmu.edu</a>
EDUCATION	<b>University of Notre Dame</b> Ph.D. in Computer Science	2014–2019
	<b>National Technical University of Athens</b> B.Sc - M.Sc. in <a href="#">Electrical and Computer Engineering</a>	2008–2014
RESEARCH & PROFESSIONAL EXPERIENCE	<b>Postdoctoral Research Associate</b> Language Technologies Institute, Carnegie Mellon University Advisor: <a href="#">Graham Neubig</a>	February 2019 – present
	<b>Research Assistant</b> University of Notre Dame Advisor: <a href="#">David Chiang</a>	August 2014 – February 2019
	<b>Research Intern</b> Google AI Speech & Language Supervisor: <a href="#">Shankar Kumar</a>	June – September 2018
	<b>Research Intern</b> Google AI Language Supervisors: <a href="#">Chris Alberti</a> and <a href="#">Manaal Faruqui</a>	May – August 2017
	<b>Visiting Researcher</b> Institute of Language, Cognition, and Computation, University of Edinburgh Advisors: <a href="#">Sharon Goldwater</a> and <a href="#">Adam Lopez</a>	January – March 2017
	<b>Research Intern</b> Fondazione Bruno Kessler, Trento Supervisors: <a href="#">Matteo Negri</a> and <a href="#">Marco Turchi</a>	June – August 2013
TEACHING EXPERIENCE	<b>Co-Instructor</b> Machine Translation and Sequence-to-Sequence Models Carnegie Mellon University (CS 11-731) with Graham Neubig Graduate level, 60 students.	Fall 2019
	<b>Co-Instructor</b> Neural Networks for NLP Carnegie Mellon University (CS 11-747) with Graham Neubig Graduate level, 90 students.	Spring 2019
	<b>Guest Lecturer</b> Natural Language Processing University of Notre Dame (CSE 40657/60657) with David Chiang Undergraduate/Graduate level, 35 students.	Fall 2016

## Teaching Assistant

Spring 2016

Theory of Computing

University of Notre Dame (CSE 30151) with David Chiang

Undergraduate level, 80 students.

## PUBLICATIONS

- [1] Keita Kurita, Anna Belova, and **Antonios Anastasopoulos**. “Towards Robust Toxic Content Classification”, to appear in *The AAAI-20 Workshop on Engineering Dependable and Secure Machine Learning Systems*, New York, February 2020.
- [2] **Antonios Anastasopoulos** and Graham Neubig. [Pushing the Limits of Low-Resource Morphological Inflection](#), in *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, Hong Kong, November 2019.
- [3] Zi-Yi Dou, **Antonios Anastasopoulos**, and Graham Neubig. [Unsupervised Domain Adaptation for Neural Machine Translation with Domain-Aware Feature Embeddings](#), in *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, Hong Kong, November 2019.
- [4] Zi-Yi Dou, Keyi Yu, and **Antonios Anastasopoulos**. [Investigating Meta-Learning Algorithms for Low-Resource Natural Language Understanding Tasks](#), in *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, Hong Kong, November 2019.
- [5] **Antonios Anastasopoulos**. [An Analysis of Source-Side Grammatical Errors in NMT](#), in *Proceedings of the 2nd BlackboxNLP workshop on analyzing and interpreting neural networks for NLP*, Florence, Italy, August 2019.
- [6] Mengzhou Xia, Xiang Kong, **Antonios Anastasopoulos**, and Graham Neubig. [Generalized Data Augmentation for Low-Resource Translation](#), in *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics (ACL)*, Florence, Italy, August 2019.
- [7] Yu-Hsiang Lin, Chian-Yu Chen, Jean Lee, Zirui Li, Yuyan Zhang, Mengzhou Xia, Shruti Rijhwani, Junxian He, Zhisong Zhang, Xuezhe Ma, **Antonios Anastasopoulos**, Patrick Littell, and Graham Neubig. [Choosing Transfer Languages for Cross-Lingual Learning](#), in *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics (ACL)*, Florence, Italy, August 2019.
- [8] Xian Li, Paul Michel, **Antonios Anastasopoulos**, Yonatan Belinkov, Nadir Durrani, Orhan Firat, Philipp Koehn, Graham Neubig, Juan Pino and Hassan Sajjad. [Findings of the First Shared Task on Machine Translation Robustness](#) in *Proceedings of the Fourth Conference on Machine Translation (WMT)*, Florence, Italy, August 2019.

- [9] Shuyan Zhou, Xiangkai Zeng, Yingqi Zhou, **Antonios Anastasopoulos**, and Graham Neubig. [Improving Robustness of Neural Machine Translation with Multi-task Learning](#) in *Proceedings of the Fourth Conference on Machine Translation (WMT)*, Florence, Italy, August 2019.
- [10] **Antonios Anastasopoulos**, Alison Lui, Toan Q. Nguyen, and David Chiang. [Neural Machine Translation of Text from Non-Native Speakers](#) in *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL-HLT)*, Minneapolis, Minnesota, June 2019.
- [11] Brian Thompson, Huda Khayrallah, **Antonios Anastasopoulos**, Arya McCarthy, Kevin Duh, Rebecca Marvin, Paul McNamee, Jeremy Gwinnup, Tim Anderson, and Philipp Koehn. [Freezing Subnetworks to Analyze Domain Adaptation in Neural Machine Translation](#) In *Proceedings of the Third Conference on Machine Translation, Volume 1: Research Papers (WMT)*, Brussels, Belgium, October 2018.
- [12] **Antonios Anastasopoulos** and David Chiang. [Leveraging Translations for Speech Transcription in Low-Resource Settings](#) In *Proceedings of Interspeech 2018*, Hyderabad, India, September 2018.
- [13] Marceley Zanon-Boito, **Antonios Anastasopoulos**, Marika Lekakou, Aline Villavicencio, and Laurent Besacier. [A small Griko-Italian speech translation corpus](#) In *the 6th Workshop on Spoken Language Technologies for Under-Resourced Languages (SLTU)*, Gurugram, India, September 2018.
- [14] **Antonios Anastasopoulos**, Marika Lekakou, Josep Quer, Eleni Zimianiti, Justin DeBenedetto, and David Chiang. [Part-of-Speech Tagging on an Endangered Language: A Parallel Griko-Italian Resource](#) In *Proceedings of the 27th International Conference on Computational Linguistics (COLING)*, Santa Fe, New Mexico, August 2018.
- [15] **Antonios Anastasopoulos** and David Chiang. [Tied Multitask Models for Neural Speech Transcription and Translation](#) In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL-HLT)*, New Orleans, Louisiana, June 2018.
- [16] **Antonios Anastasopoulos**, Sameer Bansal, Sharon Goldwater, Adam Lopez, and David Chiang. [Spoken Term Discovery for Language Documentation using Translations](#) In *the 1st Workshop on Speech-Centric Natural Language Processing (SCNLP)*, Copenhagen, Denmark, September 2017.
- [17] Graham Neubig, Chris Dyer, Yoan Goldberg, Steve Matthews, Waleed Ammar, **Antonios Anastasopoulos**, Miguel Ballesteros, David Chiang, Daniel Clothiaux, Trevor Cohn, Kevin Duh, Manaal Faruqui, Cynthia Gan, Dan Garrette, Yangfeng Ji, Lingpeng Kong, Adhiguna Kuncoro, Gaurav Kumar, Chaitanya Malaviya, Paul Michel, Yusuke Oda, Matthew Richardson, Naomi Saphra, Swabha Swayamdipta, Pengcheng Yin. [DyNet: The Dynamic Neural Network Toolkit](#), January 2017.

- [18] **Antonios Anastasopoulos**. [A case study on using speech-to-translation alignments for language documentation](#) In *the 2nd Workshop on Computational Methods for Endangered Languages (ComputEL-2)*, Honolulu, Hawaii, March 2017.
- [19] **Antonios Anastasopoulos**, David Chiang, and Long Duong. [An Unsupervised Probability Model for Speech-to-Translation Alignment of Low-Resource Languages](#) In *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing (EMNLP)* (pp. 1255-1263), Austin, Texas, November 2016.
- [20] Long Duong, **Antonios Anastasopoulos**, Trevor Cohn, Steven Bird, and David Chiang. [An Attentional Model for Speech Translation Without Transcription](#) In *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL-HLT)* (pp. 949-959), San Diego, California, June 2016.
- [21] Marco Turchi, **Antonios Anastasopoulos**, José Guilherme Camargo de Souza, and Matteo Negri. [Adaptive Quality Estimation for Machine Translation](#) In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (ACL)* (Vol. 1, pp. 710-720), 2014.
- [22] José Guilherme Camargo de Souza, Marco Turchi, **Antonios Anastasopoulos**, and Matteo Negri. [Online and Multitask learning for Machine Translation Quality Estimation in Real-world scenarios](#) In *1st Italian Conference on Computational Linguistics*, Pisa, Italy, 2014.
- [23] Stella Markantonatou, **Antonios Anastasopoulos**, and Yanis Maistros. [Explicit Objects and Polysemy](#). In *11th International Conference on Greek Linguistics (ICGL)*, Rhodes, Greece 2013.
- [24] **Antonios Anastasopoulos**, Stella Markantonatou, Yanis Maistros. [Using noun and clitic frequencies to study verb transitivity in Modern Greek](#) In *International Workshop on The relative frequencies of nouns, pronouns, and verbs in discourse adjacent to ALT10*, Leipzig, Germany, 2013.

PUBLICATIONS  
UNDER REVIEW  
AND PREPRINTS

- [1] Xinyi Wang, Hieu Pham, Paul Michel, **Antonios Anastasopoulos**, Graham Neubig, and Jaime Carbonell. “Optimizing Data Usage via Differentiable Rewards”, preprint on arXiv:1911.10088.
- [2] **Antonios Anastasopoulos**, Hilaria Cruz, and Gregory Stump. “A Resource for Studying the Chatino Verbal Morphology”, under review for LREC 2020.
- [3] Mingjun Duan, Carlos Fasola, Rodolfo Vega, **Antonios Anastasopoulos**, Lori Levin, Alan W Black, and Jaime Carbonell. “A Resource for Computational Experiments on Mapudungun”, under review for LREC 2020, preprint on arXiv:1912.01772.

- [4] David R. Mortensen, Xinjian Li, Patrick Littell, Alexis Michaud, **Antonios Anastasopoulos**, and Graham Neubig. “Multilingual Allophone Resource for (Near) Universal ASR”, under review for LREC 2020.
- [5] Yiyuan Li, **Antonios Anastasopoulos**, and Alan W Black. “Towards Minimal Supervision BERT-based Grammar Error Correction”, accepted at *AAAI-20 Student Abstract and Poster Program*.
- [6] Yiyuan Li, **Antonios Anastasopoulos**, and Alan W Black. “Towards Minimal Supervision BERT-based Grammar Error Correction”, accepted at *Natural Language, Dialog, and Speech (NDS)*, New York, November 2019.
- [7] Yiyuan Li, **Antonios Anastasopoulos**, and Alan W Black. “Comparison of Interactive Knowledge Base Spelling Correction Models for Low-Resource Languages”, accepted at *Natural Language, Dialog, and Speech (NDS)*, New York, November 2019.
- [8] **Antonios Anastasopoulos**, Shankar Kumar, and Hank Liao. “Neural Language Modeling with Visual Features” preprint on arXiv:1903.02930.

#### AWARDS

##### Research

- 2018–2019 Outstanding Research Assistant Award, 2019  
Notre Dame Department of Computer Science and Engineering

##### Teaching

- Outstanding Graduate Student Teacher Award, 2017  
Notre Dame Graduate School
- Outstanding Graduate Student Teacher Award, 2017  
Notre Dame Kaneb Center for Teaching and Learning

##### Scholarships

- Undergraduate Studies Scholarship 2008 - 2013  
by the John S. Latsis Public Benefit Foundation

#### MENTORING

- Mengzhou Xia (2019). Masters project at Carnegie Mellon University. *Neural Machine Translation for Low-Resource Languages, Prediction of NLP Performance*.
- Additi Chaudhary (2019). PhD at Carnegie Mellon University. *Morphological Analysis for Low-Resource Languages*.
- Xinyi Wang (2019). PhD at Carnegie Mellon University. *Data Selection for Multilingual Machine Translation*.
- Nikitha Murikinati (2019). CMU AI Mentoring program for undergrads from under-represented communities. *Neural Machine Translation*.
- Emanuele Bugliarello (2019). Masters project at Tokyo Institute of Technology. *Neural Machine Translation* (→ PhD at University of Copenhagen, Denmark).
- Xinyuan Zhao (2019). Masters independent study at Carnegie Mellon University. *Neural Post-Editing and Human-in-the-Loop for Machine Translation*.
- Vasiliki Mikedas-Mitsa (2019). Masters project at Universitat Pompeu Fabra, Spain. *Morphological Analysis in Griko*.

- Pin-Jung Chen, Keita Kurita, Mengzhou Xia (2019). Masters Capstone project at Carnegie Mellon University. *Cross-lingual Models for Morphological Analysis and Neural Machine Translation at Scale*.
- Zi-Yi Dou (2019). Masters project at Carnegie Mellon University. *Domain Adaptation for Neural Machine Translation*.
- Yiyuan Li (2019). Masters project at Carnegie Mellon University. *BERT-based Multilingual Unsupervised Grammar Error Correction*.
- Eleni Zimianiti (2018). Undergraduate thesis at University of Ioannina, Greece. *POS-tagging in Griko* ( $\rightarrow$  Masters at TU Delft).
- Alison Lui (2017). Undergraduate independent study at University of Notre Dame. *Neural Machine Translation for Ungrammatical Input* ( $\rightarrow$  Software Engineer at Google Translate).

SERVICE	Area Chair for the Widening NLP workshop (WiNLP)	2019-2020
	Co-Organizer of the Workshop on Language Technology for Language Documentation and Revitalization, Pittsburgh, PA, USA	August 2019
	Mentor for the CMU AI mentoring program	2019
	Area Chair (Resources and Evaluation) for ACL	2019
	Organizing Committee for MSLD 2018	May 2018
	ACL Student Research Workshop pre-submission Mentor	2017, 2019
	Reviewer for ACL, NAACL, EMNLP, COLING, SRW	2017 – Present
	Quality of Life co-chair, Notre Dame Graduate Student Union	2016 – 2017
	<ul style="list-style-type: none"> <li>• Organise events that promote graduate student welfare.</li> <li>• Focus on international and LGBTQ students' integration.</li> </ul>	
	Notre Dame International and Graduate Orientation Ambassador	2015, 2016
INVITED TALKS	<ul style="list-style-type: none"> <li>• <i>Tutorial: NLP meets Endangered Languages</i> the 28<sup>th</sup> Conference on Computational Linguistics, Barcelona, Spain</li> <li>• <i>Lecture: Deep Learning and Computation Graphs</i> Talk: <i>Translating Noisy Text</i> 5<sup>th</sup> Machine Translation Marathon of the Americas, College Park, MD</li> <li>• <i>Using Translations and NLP for Documentation of Endangered Languages</i> AUEB NLP Seminar, Athens, Greece</li> <li>• <i>Multitask Models for Speech Translation</i> post-JSALT meetup, Grenoble, France</li> <li>• <i>Speech translation for documentation of endangered languages</i> Georgetown NLP, DC</li> <li>• <i>Speech translation for documentation of endangered languages</i> Babelfish Research Meeting, Google, Mountainview, CA</li> </ul>	Sep 2020 May 2019 Apr 2018 Jan 2018 Oct 2017 Jul 2017

- *Speech to translation alignment for documentation of endangered languages*  
Probabilistic Models for Language Seminar, Edinburgh, UK      Feb 2017

## SOFTWARE

### LANG2VEC

- A python package that provides typological features for more than 6000 languages, as well as pre-computed distance vectors between them.  
<https://pypi.org/project/lang2vec/>

### MULTITASK AND MULTISOURCE NEURAL MODELS

- Sequence-to-sequence neural multi-task and multi-source models implemented in DyNet. Extensions of the seq2seq model that improve speech transcription and translation quality in low-resource settings.  
<https://bitbucket.org/antonis/dynet-multitask-models>,  
<https://bitbucket.org/antonis/dynet-multisource-models>

### INFLECTION NEURAL MODELS

- Our winning system of the SIGMORPHON 2019 Shared Task on Morphological Inflection, implemented in DyNet, for state-of-the-art results in low-resource languages.  
<https://github.com/antonisa/inflection>

### SPEECH2TRANSLATION

- A tool for directly aligning speech to its translation text without access to transcriptions. It extends the IBM model 2 combining it with a clustering model based on Dynamic Time Warping as a distance measure.  
<https://bitbucket.org/ndnlp/speech2translation>

### AQET

- An open-source package for performing online Quality Estimation for Machine Translation, based on Online Support Vector Regression and Online Gaussian Processes. Developed at Fondazione Bruno Kessler. The current release was made possible through the [MateCat](#) project.  
<https://bitbucket.org/antonis/adaptiveqe>

### CONTRIBUTIONS TO OPEN-SOURCE SOFTWARE

- CNN (now DYNET) : a C++/Python neural networks library. Worked on using grammar-like statements for more user-friendly development at the early stage of the library's development (Spring 2015).  
<https://dynt.readthedocs.io>
- PIALIGN : a C++ library for unsupervised learning to jointly align and extracting phrase table using bayesian inference. Expanded the model to include new tuning parameters, improving performance by up to 5%.  
<http://www.phontron.com/pialign/>

## DATASETS

### MAPUDUNGUN CORPUS

- A corpus of spoken Mapudungun (an endangered language in Chile) transcribed and translated in Spanish, originally collected by the CMU-AVENUE project.  
<https://github.com/mingjund/mapudungun-corpus>

#### JFLEG-ES

- Spanish translations of the JHU Fluency-Extended GUG corpus, in order to facilitate research on Machine Translation for non-native English speakers.  
<https://bitbucket.org/antonis/nmt-grammar-noise/src/master/>

#### GRIKO NARRATIVES

- A digitization of 113 narratives in Griko, with translation in Italian, and annotated with Part-Of-Speech tags, in order to facilitate computational research in under-represented endangered languages.  
<https://bitbucket.org/antonis/grikoresource>

#### GRIKO DATABASE

- An online corpus of Griko, a greek dialect spoken in southern Italy. The corpus contains transcriptions of sentences, along with the corresponding sound files, recorded during oral interviews made by our team in four villages in Grecía Salentina, in 2013. The project was financially supported by the John S. Latsis Public Benefit Foundation.  
<http://griko.project.uoi.gr>

#### A SMALL SPEECH-TO-TRANSLATION GRIKO-ITALIAN CORPUS

- A small speech corpus in Griko with translations in Italian, in order to facilitate reproducible computational language documentation experiments and evaluation for spoken term discovery and speech-to-translation alignment.  
<https://github.com/antonisa/griko-italian-parallel-corpus>