

LIN 6932 Advances in Computational Methods for Low-Resource Languages

Spring 2024

Mon 12:50am-1:40pm & Wed 11:45am-1:40pm

Room: Turlington 2349



Course description

The amount of available data remains a key factor determining the performance of natural language processing (NLP) systems. Most languages have only limited data available. This includes endangered languages such as Arapaho as well as major languages such as Persian. In this seminar, we will explore together intriguing directions in current research that address this “low resource” issue. We will look at proven methods such as cross-lingual transfer data augmentation. We will also explore innovative methods such as human-in-the-loop strategies and the possibilities of large language models (LLMs) such as ChatGPT. No previous experience in computational linguistics or NLP is required but will be helpful. Some grasp of linguistics concepts and terminology is recommended.

Instructor: Sarah Moeller
Email: smoeller@ufl.org
Tel: (352) 294-7449
Office: Turlington 4017
Office hours: Wednesdays
10:30-11:30am & 3-4pm

Goals

After the course, students will:

- ...have engaged with significant recent research in low-resource NLP.
- ...be familiar with trends and advances in low-resource NLP.
- ...have undertaken their own research in low-resource NLP.

Grade breakdown

Assignments	15%
Readings & Participation	20%
Paper Presentations	15%
Research Project	50%

Assignments

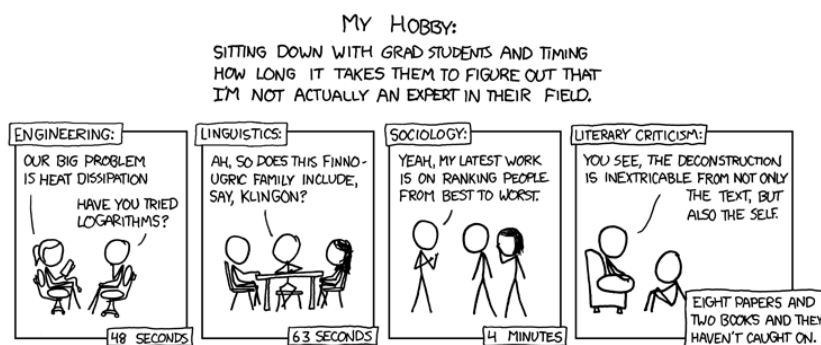
Short assignments are intended to enhance students’ engagement with the readings and guide their professional development as researchers.

Readings and participation in class discussion.

This course centers on discussing recent peer-reviewed papers. Students should attend every class having already read assigned papers. They should contribute ideas, comments, and questions to discussions and demonstrate critical engagement on the readings. Everyone should ask themselves as they read:

- What in linguistics could contribute to these tools, algorithms, methods?
- What in NLP could advance these resources, technology, methods, or linguistic descriptions?

Students will choose reading roles for each paper. Students must come to class prepared to fulfill their assigned reading roles. Everyone must serve as “Academic Peer Reviewer” at least 6 times during the semester. This person starts the discussion.



Paper Presentations

Students read 1-2 papers individually and then present them to the class. They will give a 10-20-minute overview of the papers which their classmates have not read. Presentations should include slides and handouts as necessary to aid audience's understanding.

Research Project

Students will conduct and present their own research project that will be assessed in three parts: research proposal, research presentation, and research paper. Students will submit a proposal for a research project they wish to conduct related to the topic of this course. During the last weeks of classes, students will present their own research in a clear and engaging way and in a style appropriate for an international academic conference. By the end of the semester, students will submit a paper reporting the motivation, related literature, methods, and outcome of their research.

The fine print

ACADEMIC INTEGRITY. UF students are bound by The Honor Pledge. On all work submitted for credit at the University of Florida, the following pledge is either required or implied: “On my honor, I have neither given nor received unauthorized aid in doing this assignment.” The Honor Code (<https://www.dso.ufl.edu/sccr/process/student-conduct-honor-code/>) specifies a number of behaviors that are in violation of this code and the possible sanctions. Furthermore, you are obligated to report any condition that facilitates academic misconduct to appropriate personnel. If you have any questions or concerns, please consult with the instructor or TAs.

DIGITAL DISTRACTION. When your device is displaying things irrelevant to class, you distract not only yourself but other students to whom the screen is visible. Research demonstrates that such distraction is detrimental to learning. I expect your full attention and presence, and I expect you to allow the same to others. Violation may result in public reprimand.

CLASSROOM CONDUCT. Students and faculty each have responsibility for maintaining an appropriate learning environment. Those who fail to adhere to such behavioral standards may be subject to discipline. I pledge to treat each of you with dignity, respect, and professional courtesy; I expect you to do the same for me and for each other.

ACCOMMODATION POLICIES. If you qualify for accommodations because of a disability, please submit your accommodation letter from the Disability Resource Center to me in a timely manner so that your needs can be addressed. Get started with the Disability Resource Center <https://disability.ufl.edu/students/get-started/>.

RELIGIOUS OBSERVANCES A student should inform the instructor of observances of their faith that will conflict with class attendance, tests or examinations, or other class activities *prior* to the class or occurrence of that test or activity. Faculty is obligated to accommodate that particular student’s religious observances. See policy details at <https://catalog.ufl.edu/UGRD/academic-regulations/attendance-policies/#religiousholidaystext>.

COURSE EVALUATIONS. Students are expected to provide professional and respectful feedback on the quality of instruction in this course by completing course evaluations. Students will be notified when the evaluation period opens, and can complete evaluations through the email they receive from GatorEvals, in their Canvas course menu or via ufl.bluera.com/ufl/.

COLLEGIAL ENVIRONMENT. Students are encouraged to employ critical thinking and to rely on data and verifiable sources to interrogate all assigned readings and subject matter in this course as a way of determining whether they agree with their classmates and/or their instructor. No lesson is intended to espouse, promote, advance, inculcate, or compel a particular feeling, perception, viewpoint or belief. Students are encouraged to share their viewpoints, data, and sources in class and to speak with the instructor or classmates, in class or privately, about any perceived violation of this policy.

GRADING SCHEME. This course follows UF grades and grading policy: <https://catalog.ufl.edu/UGRD/academic-regulations/grades-grading-policies/>

	B+	87-89%	C+	77-79%	D+	67-69%			
A	94-100%	B	83-86%	C	73-76%	D	63-66%	F	0-59%
A-	90-93%	B-	80-82%	C-	70-72%	D-	60-62%		

Assignments

5% Citation manager. Choose, set up, and use a citation manager tool such as those listed here: <https://research.com/software/best-academic-writing-software>

10% Annotated Bibliography. A PDF with consistent formatting that contains at least 10 peer-reviewed items with brief annotations (different from your reading notes). The annotations should be relevant to your research interests. Consider: How is this helpful to me? What sections did I read/skim/not understand? What research questions/ideas came to me?

Research Project Breakdown

10% Research proposal with a bibliography of references related to the research proposal. The proposal must be approved to be your term research project. The research must be related to the topic of this course. The proposal must include a research question, an outline of proposed methods, a description of the data or resources to be used. This should be up to one page. Another page will include a bibliography related to the research proposal.

10% Conference Style Research Presentation. The last weeks are reserved for you to present your research projects. You should prepare a 15-minute conference-quality presentation and be ready for a 5-10-minute Q/A.

30% Research and Paper. You will undertake original research related to the topic of this class and submit a paper. The paper should describe the motivation, related literature, methods, and outcome of your research. It should 6-8 pages long and follow [ACL formatting template, policies, and guidelines](#) (i.e. it should have a "Limitations" section, which *does* count toward your total length). The paper is due on the scheduled finals day.

Schedule

1/8	Introduction		
1/10	AI and Linguistics		
	NLP Evolution: Rule-based, ML, DL	(Johri et al. 2021) Similar but w/pix here Timeline here	
1/15	<i>NO CLASSES</i>		
1/17	NLP Overview	(Khurana et al. 2023)	Citation Manager
	Data Quality	(Hedderich & Klakow 2018) (Xie et al. 2018) <i>or</i> (Al Sharou, Li & Specia 2021)	<i>Start role-playing</i>
1/22	Low-Resource Challenge	(Simons, Thomas & White 2022) <i>or</i> (Joshi et al. 2020) <i>or</i> (Blasi, Anastasopoulos & Neubig 2021)	
1/24	Model Engineering	(Wu, Cotterell & Hulden 2021)	

Spring 2024 Computational Methods for Low-Resource and Endangered Languages

	Low Resource Methods Overview	(Hedderich et al. 2021) Condensed version w/pix here	
1/29	Joint Learning, Transfer Learning, and Multilingual Models	(Hakimi Parizi & Cook 2021) (Abudouwaili et al. 2023)	Choose LLM papers
1/31		(Wang, Ruder & Neubig 2022) (Adelani et al. 2022) or (Kondratyuk & Straka 2019)	
2/5	LLMs & Low-Resource Languages	(Ramponi 2023)	
2/8		<i>Student Presentations. Choose one:</i> (Luukkonen et al. 2023; Purkayastha et al. 2023; Winata et al. 2021; Zhang et al. 2023; Koto et al. 2023; Gehrmann et al. 2022)	
2/12	Data Augmentation	(Feng et al. 2021)	Annotated Bibliography
2/14	Data Augmentation: Synthesis & Hallucination	(Bartelds, Jurafsky & Wieling 2023) (Samir & Silfverberg 2023)	
2/19	Guest: Zoey Liu		
2/21	Leveraging Resources	(McCarthy et al. 2018) or (Beemer et al. 2020) (Christianson, Duncan & Onyshkevych 2018)	
2/26	Human-in-the-Loop & Active Learning	(Wu et al. 2022)	
2/28		(Gessler, Levine & Zeldes 2022) (Dossou et al. 2022)	
3/4	Human-in-the-Loop & Active Learning	(Kartchner et al. 2022)	Research Proposal
3/6		(Liu, Guo & Mahmud 2021) (Ikhwantri et al. 2018)	
<i>BREAK</i>			
3/21	ComputEL-7	<i>Attend online or in person</i>	
3/22			
3/25	Student's Choice	<i>Student Presentations.</i>	
3/27			
4/1	Linguistically Informed NLP	(Eskander, Klavans & Muresan 2019)	
4/3		(Beemer et al. 2020) (Morita & Iwakura 2019)	
4/8	Dealing with Realities	(Millour & Fort 2019)	
4/10	Linguistically Informed NLP	(Silva & Amarathunga 2019) (Gungor, Gungor & Uskudarli 2019)	

4/15	NLP Ethics & Minority Communities	(Caselli et al. 2021) (Liu et al. 2022) or (Schwartz 2022)	
4/17	The Future of Linguistics in the Age of AI	(McShane & Nirenburg 2021) sections 1.1, pages 7-9, skim 1.4.1-4, 1.6.3, 1.6.6, 1.6.7, 1.6.11, 1.6.12	
4/22	Research Presentations		
4/24			

Readings

- Abudouwaili, Gulinigeer, Kahaerjiang Abiderexiti, Nian Yi & Aishan Wumaier. 2023. Joint Learning Model for Low-Resource Agglutinative Language Morphological Tagging. In Garrett Nicolai, Eleanor Chodroff, Frederic Mailhot & Çağrı Çöltekin (eds.), *Proceedings of the 20th SIGMORPHON workshop on Computational Research in Phonetics, Phonology, and Morphology*, 27–37. Toronto, Canada: Association for Computational Linguistics. <https://doi.org/10.18653/v1/2023.sigmorphon-1.4v>.
- Adelani, David, Graham Neubig, Sebastian Ruder, Shruti Rijhwani, Michael Beukman, Chester Palen-Michel, Constantine Lignos, et al. 2022. MasakhaNER 2.0: Africa-centric Transfer Learning for Named Entity Recognition. In *Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing*, 4488–4508. Abu Dhabi, United Arab Emirates: Association for Computational Linguistics. <https://doi.org/10.18653/v1/2022.emnlp-main.298>.
- Al Sharou, Khetam, Zhenhao Li & Lucia Specia. 2021. Towards a Better Understanding of Noise in Natural Language Processing. In *Proceedings of the Conference Recent Advances in Natural Language Processing - Deep Learning for Natural Language Processing Methods and Applications*, 53–62. INCOMA Ltd. Shoumen, BULGARIA. https://doi.org/10.26615/978-954-452-072-4_007.
- Allman, Tod, Stephen Beale & Richard Denton. 2014. Toward an Optimal Multilingual Natural Language Generator: Deep Source Analysis and Shallow Target Analysis. *Philippine Computing Journal* 9(1). 55–63. <http://www.thebibletranslatorsassistant.com/pdf/Philippine%20Computing%20Journal%202014.pdf>. (8 June, 2017).
- Bartelds, Martijn, Dan Jurafsky & Martijn Wieling. 2023. Making More of Little Data: Improving Low-Resource Automatic Speech Recognition Using Data Augmentation. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics Volume 1: Long Papers*, 715–729. <https://aclanthology.org/2023.acl-long.42.pdf>.
- Beemer, Sarah, Zak Boston, April Bukoski, Daniel Chen, Princess Dickens, Andrew Gerlach, Torin Hopkins, et al. 2020. Linguist vs. Machine: Rapid Development of Finite-State Morphological Grammars. In *Proceedings of the 17th SIGMORPHON Workshop on Computational Research in Phonetics, Phonology, and Morphology*, 162–170. Online: Association for Computational Linguistics. <https://doi.org/10.18653/v1/2020.sigmorphon-1.18>.
- Blasi, Damián, Antonios Anastasopoulos & Graham Neubig. 2021. Systematic Inequalities in Language Technology Performance across the World’s Languages. In. arXiv. <http://arxiv.org/abs/2110.06733>. (8 December, 2023).
- Caselli, Tommaso, Roberto Cibin, Costanza Conforti, Enrique Encinas & Maurizio Teli. 2021. Guiding Principles for Participatory Design-inspired Natural Language Processing. In Anjalie Field, Shrimai Prabhumoye, Maarten Sap, Zhijing Jin, Jieyu Zhao & Chris Brockett (eds.), *Proceedings of the 1st Workshop on NLP for Positive Impact*, 27–35. Online: Association for Computational Linguistics. <https://doi.org/10.18653/v1/2021.nlp4posimpact-1.4>.

- Christianson, Caitlin, Jason Duncan & Boyan Onyshkevych. 2018. Overview of the DARPA LORELEI Program. *Machine Translation* 32(1). 3–9. <https://doi.org/10.1007/s10590-017-9212-4>.
- Dossou, Bonaventure F. P., Atnafu Lambebo Tonja, Oreen Yousuf, Salomey Osei, Abigail Oppong, Iyanuoluwa Shode, Oluwabusayo Olufunke Awoyomi & Chris Emezue. 2022. AfroLM: A Self-Active Learning-based Multilingual Pretrained Language Model for 23 African Languages. In Angela Fan, Iryna Gurevych, Yufang Hou, Zornitsa Kozareva, Sasha Luccioni, Nafise Sadat Moosavi, Sujith Ravi, Gyuwan Kim, Roy Schwartz & Andreas Rücklé (eds.), *Proceedings of The Third Workshop on Simple and Efficient Natural Language Processing (SustainLP)*, 52–64. Abu Dhabi, United Arab Emirates (Hybrid): Association for Computational Linguistics. <https://doi.org/10.18653/v1/2022.sustainlp-1.11>.
- Escribe, Marie. 2019. Human Evaluation of Neural Machine Translation: The Case of Deep Learning. In *Proceedings of the Human-Informed Translation and Interpreting Technology Workshop (HiT-IT 2019)*, 36–46. Varna, Bulgaria: Incoma Ltd., Shoumen, Bulgaria. https://doi.org/10.26615/issn.2683-0078.2019_005.
- Eskander, Ramy, Judith Klavans & Smaranda Muresan. 2019. Unsupervised Morphological Segmentation for Low-Resource Polysynthetic Languages. In *Proceedings of the 16th Workshop on Computational Research in Phonetics, Phonology, and Morphology*, 189–195. Florence, Italy: Association for Computational Linguistics. <https://www.aclweb.org/anthology/W19-4222>. (12 August, 2019).
- Feng, Steven Y., Varun Gangal, Jason Wei, Sarath Chandar, Soroush Vosoughi, Teruko Mitamura & Eduard Hovy. 2021. A Survey of Data Augmentation Approaches for NLP. In Chengqing Zong, Fei Xia, Wenjie Li & Roberto Navigli (eds.), *Findings of the Association for Computational Linguistics: ACL-IJCNLP 2021*, 968–988. Online: Association for Computational Linguistics. <https://doi.org/10.18653/v1/2021.findings-acl84>.
- Gehrmann, Sebastian, Sebastian Ruder, Vitaly Nikolaev, Jan A. Botha, Michael Chavinda, Ankur Parikh & Clara Rivera. 2022. TaTa: A Multilingual Table-to-Text Dataset for African Languages. arXiv. <https://doi.org/10.48550/arXiv.2211.00142>.
- Gessler, Luke, Lauren Levine & Amir Zeldes. 2022. Midas Loop: A Prioritized Human-in-the-Loop Annotation for Large Scale Multilayer Data. In Sameer Pradhan & Sandra Kuebler (eds.), *Proceedings of the 16th Linguistic Annotation Workshop (LAW-XVI) within LREC2022*, 103–110. Marseille, France: European Language Resources Association. <https://aclanthology.org/2022.law-1.13>. (5 December, 2023).
- Gungor, Onur, Tunga Gungor & Suzan Uskudarli. 2019. The effect of morphology in named entity recognition with sequence tagging. *Natural Language Engineering* 1(1). 1–23. https://www.researchgate.net/publication/326657401_The_effect_of_morphology_in_named_entity_recognition_with_sequence_tagging. (21 February, 2019).
- Hakimi Parizi, Ali & Paul Cook. 2021. Evaluating a Joint Training Approach for Learning Cross-lingual Embeddings with Sub-word Information without Parallel Corpora on Lower-resource Languages. In *Proceedings of *SEM 2021: The Tenth Joint Conference on Lexical and Computational Semantics*, 302–307. Online: Association for Computational Linguistics. <https://doi.org/10.18653/v1/2021.starsem-1.29>.
- Hedderich, Michael A. & Dietrich Klakow. 2018. Training a Neural Network in a Low-Resource Setting on Automatically Annotated Noisy Data. In *Proceedings of the Workshop on Deep Learning Approaches for Low-Resource NLP*, 12–18. Melbourne: Association for Computational Linguistics. <http://www.aclweb.org/anthology/W18-3402>. (23 October, 2018).
- Hedderich, Michael A., Lukas Lange, Heike Adel, Jannik Strötgen & Dietrich Klakow. 2021. A Survey on Recent Approaches for Natural Language Processing in Low-Resource Scenarios. In *Proceedings of the 2021 Conference of the North American Chapter of the Association for*

- Computational Linguistics: Human Language Technologies*, 2545–2568. Online: Association for Computational Linguistics. <https://doi.org/10.18653/v1/2021.naacl-main.201>.
- Ikhwantri, Fariz, Samuel Louvan, Kemal Kurniawan, Bagas Abisena, Valdi Rachman, Alfian Farizki Wicaksono & Rahmad Mahendra. 2018. Multi-Task Active Learning for Neural Semantic Role Labeling on Low Resource Conversational Corpus. In *Proceedings of the Workshop on Deep Learning Approaches for Low-Resource NLP*, 43–50. Melbourne: Association for Computational Linguistics. <https://doi.org/10.18653/v1/W18-3406>.
- Johri, Prashant, Sunil K. Khatri, Ahmad T. Al-Taani, Munish Sabharwal, Shakhzod Suvanov & Avneesh Kumar. 2021. Natural Language Processing: History, Evolution, Application, and Future Work. In Ajith Abraham, Oscar Castillo & Deepali Virmani (eds.), *Proceedings of 3rd International Conference on Computing Informatics and Networks*, vol. 167, 365–375. Singapore: Springer Singapore. http://link.springer.com/10.1007/978-981-15-9712-1_31. (27 November, 2023).
- Joshi, Pratik, Sebastin Santy, Amar Budhiraja, Kalika Bali & Monojit Choudhury. 2020. The State and Fate of Linguistic Diversity and Inclusion in the NLP World. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, 6282–6293. Online: Association for Computational Linguistics. <https://doi.org/10.18653/v1/2020.acl-main.560>.
- Kartchner, David, Davi Nakajima An, Wendi Ren, Chao Zhang & Cassie S. Mitchell. 2022. Rule-Enhanced Active Learning for Semi-Automated Weak Supervision. *AI. Multidisciplinary Digital Publishing Institute* 3(1). 211–228. <https://doi.org/10.3390/ai3010013>.
- Khurana, Diksha, Aditya Koli, Kiran Khatter & Sukhdev Singh. 2023. Natural language processing: state of the art, current trends and challenges. *Multimedia Tools and Applications* 82(3). 3713–3744. <https://doi.org/10.1007/s11042-022-13428-4>.
- Kondratyuk, Dan & Milan Straka. 2019. 75 Languages, 1 Model: Parsing Universal Dependencies Universally. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, 2779–2795. Hong Kong, China: Association for Computational Linguistics. <https://doi.org/10.18653/v1/D19-1279>.
- Koto, Fajri, Nurul Aisyah, Haonan Li & Timothy Baldwin. 2023. Large Language Models Only Pass Primary School Exams in Indonesia: A Comprehensive Test on IndoMMLU. arXiv. <https://doi.org/10.48550/arXiv.2310.04928>.
- Liu, Zhe, Yufan Guo & Jalal Mahmud. 2021. When and Why a Model Fails? A Human-in-the-loop Error Detection Framework for Sentiment Analysis. In *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies: Industry Papers*, 170–177. Online: Association for Computational Linguistics. <https://doi.org/10.18653/v1/2021.naacl-industry.22>.
- Liu, Zoey, Crystal Richardson, Richard Hatcher & Emily Prud'hommeaux. 2022. Not always about you: Prioritizing community needs when developing endangered language technology. In Smaranda Muresan, Preslav Nakov & Aline Villavicencio (eds.), *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, 3933–3944. Dublin, Ireland: Association for Computational Linguistics. <https://doi.org/10.18653/v1/2022.acl-long.272>.
- Luukkonen, Risto, Ville Komulainen, Jouni Luoma, Anni Eskelinen, Jenna Kanerva, Hanna-Mari Kupari, Filip Ginter, et al. 2023. FinGPT: Large Generative Models for a Small Language. arXiv. <https://doi.org/10.48550/arXiv.2311.05640>.
- McCarthy, Arya D., Miikka Silfverberg, Ryan Cotterell, Mans Hulden & David Yarowsky. 2018. Marrying Universal Dependencies and Universal Morphology. In *Proceedings of the Second Workshop on Universal Dependencies (UDW 2018)*, 91–101. Brussels, Belgium: Association

- for Computational Linguistics. <http://www.aclweb.org/anthology/W18-6011>. (5 February, 2019).
- McShane, Marjorie & Sergei Nirenburg. 2021. Our Vision of Linguistics for the Age of AI. In *Linguistics for the Age of AI* (CogNet), 448. The MIT Press. <https://direct.mit.edu/books/oa-monograph/5042/chapter/2977361/Our-Vision-of-Linguistics-for-the-Age-of-AI>. (7 December, 2023).
- Millour, Alice & Karën Fort. 2019. Unsupervised Data Augmentation for Less-Resourced Languages with no Standardized Spelling. In *Proceedings of the International Conference on Recent Advances in Natural Language Processing (RANLP 2019)*, 776–784. Varna, Bulgaria: INCOMA Ltd. https://doi.org/10.26615/978-954-452-056-4_090.
- Morita, Hajime & Tomoya Iwakura. 2019. A Fast and Accurate Partially Deterministic Morphological Analysis. In *Proceedings of the International Conference on Recent Advances in Natural Language Processing (RANLP 2019)*, 804–809. Varna, Bulgaria: INCOMA Ltd. https://doi.org/10.26615/978-954-452-056-4_093.
- Purkayastha, Sukannya, Sebastian Ruder, Jonas Pfeiffer, Iryna Gurevych & Ivan Vulić. 2023. Romanization-based Large-scale Adaptation of Multilingual Language Models. arXiv. <https://doi.org/10.48550/arXiv.2304.08865>.
- Ramponi, Marco. 2023. The Full Story of Large Language Models and RLHF. *News, Tutorials, AI Research*. <https://www.assemblyai.com/blog/the-full-story-of-large-language-models-and-rlhf/>. (27 November, 2023).
- Samir, Farhan & Miikka Silfverberg. 2023. Understanding Compositional Data Augmentation in Typologically Diverse Morphological Inflection. arXiv. <https://doi.org/10.48550/arXiv.2305.13658>.
- Schwartz, Lane. 2022. Primum Non Nocere: Before working with Indigenous data, the ACL must confront ongoing colonialism. In Smaranda Muresan, Preslav Nakov & Aline Villavicencio (eds.), *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, 724–731. Dublin, Ireland: Association for Computational Linguistics. <https://doi.org/10.18653/v1/2022.acl-short.82>.
- Silva, Amila & Chaturika Amarathunga. 2019. On Learning Word Embeddings From Linguistically Augmented Text Corpora. In *Proceedings of the 13th International Conference on Computational Semantics - Short Papers*, 52–58. Gothenburg, Sweden: Association for Computational Linguistics. <https://www.aclweb.org/anthology/W19-0508>. (24 May, 2019).
- Simons, Gary F, Abbey L L Thomas & Chad K K White. 2022. Assessing Digital Language Support on a Global Scale. In *Proceedings of the 29th International Conference on Computational Linguistics*, 4299–4305. <https://aclanthology.org/2022.coling-1.379.pdf>.
- Wang, Xinyi, Sebastian Ruder & Graham Neubig. 2022. Expanding Pretrained Models to Thousands More Languages via Lexicon-based Adaptation. In Smaranda Muresan, Preslav Nakov & Aline Villavicencio (eds.), *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, 863–877. Dublin, Ireland: Association for Computational Linguistics. <https://doi.org/10.18653/v1/2022.acl-long.61>.
- Winata, Genta Indra, Samuel Cahyawijaya, Zihan Liu, Zhaojiang Lin, Andrea Madotto & Pascale Fung. 2021. Are Multilingual Models Effective in Code-Switching? In *Proceedings of the Fifth Workshop on Computational Approaches to Linguistic Code-Switching*, 142–153. Online: Association for Computational Linguistics. <https://doi.org/10.18653/v1/2021.calcs-1.20>.
- Wu, Shijie, Ryan Cotterell & Mans Hulden. 2021. Applying the Transformer to Character-level Transduction. In *Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: Main Volume, 1901--1907*. Association for Computational Linguistics. <https://aclanthology.org/2021.eacl-main.163>.

- Wu, Xingjiao, Luwei Xiao, Yixuan Sun, Junhang Zhang, Tianlong Ma & Liang He. 2022. A survey of human-in-the-loop for machine learning. *Future Generation Computer Systems* 135. 364–381. <https://doi.org/10.1016/j.future.2022.05.014>.
- Xie, Ziang, Guillaume Genthial, Stanley Xie, Andrew Ng & Dan Jurafsky. 2018. Noising and Denoising Natural Language: Diverse Backtranslation for Grammar Correction. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers)*, 619–628. New Orleans, Louisiana: Association for Computational Linguistics. <https://doi.org/10.18653/v1/N18-1057>.
- Zhang, Chiyu, Khai Duy Doan, Qisheng Liao & Muhammad Abdul-Mageed. 2023. The Skipped Beat: A Study of Sociopragmatic Understanding in LLMs for 64 Languages. arXiv. <https://doi.org/10.48550/arXiv.2310.14557>.