Automatic Detection of Morphosyntactic Dialect Features
in African American English Oral Histories

Kevin Tang[1,2], Sarah Moeller[2], Alice Rozet[2]
[1]Heinrich Heine University Düsseldorf, Faculty of Arts and Humanities, Institute of English and American Studies, Department of English Language and Linguistics
[2]University of Florida, College of Liberal Arts and Sciences, Department of Linguistics
kevin.tang@hhu.de, smoeller@ufl.edu, arozet@ufl.edu

African American English (AAE) has received recent attention in the field of natural language processing (NLP). Efforts to address bias against AAE in NLP systems tend to focus on lexical differences (Barikeri et al., 2021; Cheng et al., 2022; Garimella et al., 2022; Hwang et al., 2020; Kiritchenko and Mohammad, 2018; Maronikolakis et al., 2022; Silva et al., 2021). Whenever the structure of AAE is considered, the approach is often to remove or neutralize the unique features (Tan et al., 2020).

Our work focuses on automatic feature detection of three morphosyntactic features of AAE – habitual meanings of "be", negative concord, and person-number disagreement. While their frequency varies across speakers, in all three cases we were challenged by a class imbalance because AAE features are infrequently attested. Therefore, we did not opt for a direct data-driven approach, but rather a probabilistic approach based on patterns in a new spoken corpus of AAE that was compiled using naturalistic oral history interviews. First, we leverage published linguistic descriptions and then analyze a subset of annotated data to identify previously undescribed morphosyntactic patterns then use both to generate a set of rule-based pattern identifiers. Second, we apply data augmentation techniques to reduce the class-imbalance issue. Third, we train an ensemble of classical machine learning models and a transformer deep learning model on a representation of both our hand-crafted feature rules (Part-of-Speech and syntactic dependency patterns) and supplemented data-driven features such as word and POS n-grams.

For habitual be, integrating syntactic information improves the identification of habitual uses of "be" by about 65 F1 points over a simple baseline model of n-grams. For negative concord, a statistical model is unnecessary as the rule-based model achieved nearly perfect precision and recall, struggling only where clause boundaries are unclear in the transcriptions. For person-number disagreement, we achieved a macro averaged F1 of 95 and will discuss the challenges in detecting this complex feature.

The success of our approach demonstrates the potential impact to improve NLP when we embrace, rather than neutralize, the structural uniqueness of dialects such as African American English. Our approach can also enrich existing linguistic descriptions through model error analyses. We note in passing that while automatic feature detection can improve the fairness issues in NLP systems, automatic feature detection itself can also suffer from fairness issues (Villarreal 2024) if models are not transparent and interpretable, like ours are.

References

Soumya Barikeri, Anne Lauscher, Ivan Vulic, and Goran Glavas. 2021. Redditbias: A real-world resource for bias evaluation and debiasing of conversational language models. In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing, ACL/IJCNLP 2021,* (Volume 1: Long Papers), Virtual Event, August 1-6, 2021, pages 1941–1955. Association for Computational Linguistics.

Lu Cheng, Nayoung Kim, and Huan Liu. 2022. Debiasing word embeddings with nonlinear geometry. In *Proceedings of the 29th International Conference on Computational Linguistics, COLING 2022*, Gyeongju, Republic of Korea, October 12-17, 2022, pages 1286–1298. International Committee on Computational Linguistics.

Aparna Garimella, Rada Mihalcea, and Akhash Amarnath. 2022. Demographic-aware language model fine-tuning as a bias mitigation technique. In *Proceedings of the 2nd Conference of the Asia-Pacific Chapter of the Association for Computational Linguistics and the 12th International Joint Conference on Natural Language Processing, AACL/IJCNLP 2022 - Volume 2: Short Papers, Online only, November 20-23, 2022, pages 311–319.* Association for ComputationalLinguistics.

Alyssa Hwang, William R. Frey, and Kathleen McKeown. 2020. Towards augmenting lexical resources for slang and African American English. In *Proceedings of the 7th Workshop on NLP for Similar Languages, Varieties and Dialects*, pages 160–172, Barcelona, Spain (Online). International Committee on Computational Linguistics (ICCL).

Svetlana Kiritchenko and Saif M. Mohammad. 2018. Examining gender and race bias in two hundred sentiment analysis systems. In *Proceedings of the Seventh Joint Conference on Lexical and Computational Semantics, *SEM@NAACL-HLT 2018*, New Orleans, Louisiana, USA, June 5-6, 2018, pages 43–53. Association for Computational Linguistics.

Antonis Maronikolakis, Philip Baader, and Hinrich Schütze. 2022. Analyzing hate speech data along racial, gender and intersectional axes. In *Proceedings of the 4th Workshop on Gender Bias in Natural Language Processing (GeBNLP)*, pages 1–7, Seattle, Washington. Association for Computational Linguistics.

Andrew Silva, Pradyumna Tambwekar, and Matthew Gombolay. 2021. Towards a comprehensive understanding and accurate evaluation of societal biases in pre-trained transformers. In *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 2383–2389, Online. Association for Computational Linguistics.

Samson Tan, Shafiq Joty, Lav Varshney, and MinYen Kan. 2020. Mind your inflections! Improving NLP for non-standard Englishes with Base-Inflection Encoding. In Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP), pages 5647–5663, Online. Association for Computational Linguistics.

Dan Villarreal. 2024. "Sociolinguistic auto-coding has fairness problems too: measuring and mitigating bias" *Linguistics Vanguard*. https://doi.org/10.1515/lingvan-2022-0114