CHALLENGES IN DETECTING NULL RELATIVIZERS IN AFRICAN AMERICAN LANGUAGE FOR SOCIOLINGUISTIC AND PSYCHOLINGUISTIC APPLICATIONS

Extended abstract presented at Unimplicit: The First Workshop on Understanding Implicit and Underspecified Language at ACL-IJCNLP 2021

Anissa Neal (anneal@umass.edu), Brendan O'Connor, Lisa Green
University of Massachusetts Amherst

Introduction. African American Language (AAL) is a variety of English spoken primarily, but not exclusively, by African Americans. AAL and Mainstream American English (MAE) overlap, but there are notable differences between the two varieties [1]. One difference is in subject-extracted relative clauses (SRCs), where the subject has been extracted from a subject position inside the embedded clause. An overt relativizer (e.g., "that," "which," "who") is optional in AAL but obligatory in MAE. Null relativizers are grammatical in both varieties for object-extracted relative clauses (ORCs), where an object has been extracted from inside the embedded clause.

RC Type	Implicitness	AAL	MAE	Example
SRC	Overt	✓	✓	The banker$_{SUBJ}$ [that __$_{SUBJ}$ irritated the lawyer] *played* tennis every Saturday.
SRC	Null	✓	**X**	The banker$_{SUBJ}$ [__$_{SUBJ}$ irritated the lawyer] *played* tennis every Saturday.
ORC	Overt	✓	✓	The banker$_{OBJ}$ [that the lawyer irritated __$_{OBJ}$] *played* tennis every Saturday.
ORC	Null	✓	✓	The banker$_{OBJ}$ [the lawyer irritated __$_{OBJ}$] *played* tennis every Saturday.

Table 1. Distribution of grammaticality of relative clause structures in AAL and MAE.

This project seeks to automatically detect null SRCs as a computational sociolinguistic tool to measure dialect density [3,4], the frequency of dialect features in a collection of speech or text. Dialect density helps quantify how representative a given text is of a certain variety. Sociolinguistic and psycholinguistic work incorporating structure frequency would benefit from an automated process that determines representativeness of a sample and identifies implicit structures. Further, these constructions are an essential aspect of the AAL grammar and identifying them, and other covert linguistic forms, is crucial to developing comprehensive linguistic understanding.

Background Current work on AAL consistently finds that off-the-shelf NLP models are underequipped to handle the variety [3,5,6,7]. Furthermore, there are different methods to determine what type of language is representative of a dialect. Blodgett et al. [5,6] use geolocation features on Twitter to identify AAL tweets compared to MAE tweets, and Koenecke et al. [3] calculate a dialect density measure for their data's three locations, using the null relativizer as a representative feature which they hand coded. Null SRCs also raise interesting questions for work focused on modeling psycholinguistic behavior. There is a rich psycholinguistic literature investigating SRCs and ORCs [2,8,9] in MAE that finds ORCs are more difficult to process than SRCs. The null SRCs introduce another level of complexity because they are locally ambiguous, leading to possible garden-path interpretations. The embedded verb, "irritated", could be locally interpreted as the main verb of the utterance:

(1) The banker irritated the lawyer…
  a. … by asking extremely detailed questions. **Main Verb continuation**
  b. … played tennis every Saturday. **Null SRC continuation**

Garden-path constructions have also received attention in the computational literature [10], but the focus is on modeling the psycholinguistic effects of these sentences [11]. These garden-path utterances are notoriously difficult for human readers, but these null SRCs in AAL also present difficulties for parsers as well. Transition-based architectures commonly use a popular shift-reduce parsing method [12] to successively move through input from left-to-right and assign arcs between the words, making it susceptible to garden-path effects as it builds structure left-to-right. Alternatively, graph-based parsing infers a globally optimal structure; McDonald & Nivre [13] find this architecture works relatively well for longer distance dependencies. While dependency parsers do not mark null elements, we hypothesize they could be useful for detecting null RCs by detecting RCs (via a Universal Dependencies *acl:relcl* edge) whose subtree head does not have an overt relativizer (*mark* child). Dependency parsing has enjoyed considerable research, software, and data resources; in preliminary work, we evaluate their effectiveness on these constructions and propose to improve their accuracy.

Preliminary Work. We compare graph-based (Stanza [14]) and transition-based (CoreNLP [15]) Universal Dependencies v2.0 [17] parsers' abilities to correctly parse a sentence in the presence of the four different RC types. We construct a test set of sentences derived from the overt relative SRC and ORC examples in Traxler et al.'s human eye-tracking experiments [2], and create null versions by removing the relativizer, resulting in 159 sentences for each of the 2x2 conditions, all of which are grammatical in AAL [1]. Table 1 presents example sentences; the main verb is italicized. We parse all sentences, and evaluate a parser's frequency of correctly identifying the sentence root as the main verb, as a minimal-annotation proxy for overall parse quality. This was done since the crucial interest is the analysis of the two verbs. When a parser fails to identify the main verb as root, often the root is erroneously assigned to either (1) the embedded verb, indicating no relative clause present, or (2) the matrix clause subject ("banker" in Table 1), indicating a noun phrase reading of the entire utterance.

	Stanza (Graph-Based)			CoreNLP (Transition-Based)		
	SRC	ORC	Overall	SRC	ORC	Overall
Overt	94% 155/159	65% 103/159	80% 253/318	81% 129/159	29% 46/159	55% 175/318
Null	**3%** 5/159	67% 107/159	35% 112/318	**19%** 31/159	51% 81/159	35% 112/318
Overall	49% 155/318	66% 210/318	57% 365/636	50% 160/318	40% 127/318	45% 287/636

Table 2. Accuracy of root-attachment for the different parsing architectures.

Note accuracy is sometimes quite low, indicating substantial challenges for current parsers. Stanza's overall accuracy is somewhat better (57% vs. 45%), but it performs very poorly on null SRCs, and is outperformed by CoreNLP (3% vs. 19%). This contradicts our hypothesis that a transition-based parser should perform relatively worse with null SRCs due to their potential to be locally ambiguous garden paths (though other factors confound the comparison between these two different software packages). It is expected that the parsers should be worse on the AAL-specific construction of null SRCs, since their training data presumably includes primarily MAE and little AAL from the UD Treebank; for either parser, accuracy seems too low, and biased against AAL, to be usable for dialect density measures or for psycholinguistic corpus research. Another finding is that Stanza performs better with ORCs, a result that contrasts with human performance as ORCs are considered more difficult to process.

Conclusions and Future Work. In future work, we aim to retrain a parser to improve accuracy on the null RCs for use on large, raw text corpora—ideally, augmenting it with AAL manually annotated data [6] or examples with deleted relativizers—to eventually parse the Corpus for Regional African American Language [18], a large corpus consisting of recorded and transcribed utterances from speakers of AAL from a variety of different locations. Such a parsed corpus could support finding null SRCs as well as other syntactic constructions for AAL research. An alternative approach is to use BERT fine-tuning with constructed minimal pairs, which Demszky et al. use to help measure dialect density in Indian English [4]; we hope to test this approach to detect null SRCs. In conclusion, investigation of null SRCs offers opportunities to further understanding of implicit phenomena and also incorporates practical applications for sociolinguistic and psycholinguistic research.

References.

[1] Green, L. (2002). *African American English: A Linguistic Introduction*. Cambridge: Cambridge University Press. doi:10.1017/CBO9780511800306.

[2] Traxler, M. & Morris, R. & Seely, R. (2002). Processing Subject and Object Relative Clauses: Evidence from Eye Movements. *Journal of Memory and Language*. 47. 69-90. 10.1006/jmla.2001.2836.

[3] Koenecke, A., Nam, A., Lake, E., Nudell, J., Quartey, M., Mengesha, Z., Toups, C., Rickford, J.R., Jurafsky, D., Goel, S. *Proceedings of the National Academy of Sciences Apr 2020*, 117 (14) 7684-7689; DOI: 10.1073/pnas.1915768117.

[4] Demszky, D., Sharma, D., Clark, J.H., Prabhakaran,V., & Eisenstein, J. (forthcoming). Learning to Recognize Dialect Features. In *NAACL* 2021.

[5] Blodgett, S., Wei, J., & O'Connor, B. (2018). Twitter Universal Dependency Parsing for African-American and Mainstream American English. In *Proceedings of the 56th ACL (Volume 1: Long Papers)* (pp. 1415–1425). Association for Computational Linguistics

[6] Blodgett, S., Green, L., & O'Connor, B. (2016). Demographic Dialectal Variation in Social Media: A Case Study of African-American English. In *Proceedings of the EMNLP 2016 (pp. 1119–1130)*. Association for Computational Linguistics.

[7] Jørgensen, A., Hovy, D., & Søgaard, A. (2015). Challenges of studying and processing dialects in social media. In *Proceedings of the Workshop on Noisy User-generated Text (pp. 9–18)*. Association for Computational Linguistics.

[8] Gordon, P.C., Hendrick, R., Johnson, M. Memory interference during language processing. *Journal of Experimental Psychology: Learning, Memory, and Cognition*. 2001 Nov;27(6):1411-23. doi: 10.1037//0278-7393.27.6.1411. PMID: 11713876.

[9] Wanner, H. E., & Maratsos, M. P. (1978). An ATN approach to comprehension. In M. A. Halle, J. Bresnan, & G. A. Miller (Eds.), *Linguistic theory and psychological reality*. M.I.T. Press.

[10] Schijndel, M., & Linzen, T. (2018). A Neural Model of Adaptation in Reading. *In Proceedings of the 2018 EMNLP (pp. 4704–4710).* Association for Computational Linguistics.

[11] Grodner, D., Gibson, E., Argaman, V. Against Repair-Based Reanalysis in Sentence Comprehension. *Journal of Psycholinguistic Research* 32, 141–166 (2003). https://doi.org/10.1023/A:1022496223965.

[12] Aho, A.V. & Ullman, J.D. 1972. *The theory of parsing, translation, and compiling*. Prentice-Hall, Inc., USA.

[13] McDonald, R., & Nivre, J. (2011). Analyzing and Integrating Dependency Parsers. *Computational Linguistics, 37(1)*, 197–230.

[14] Qi, P., Zhang, Y., Zhang, Y., Bolton, J., & Manning, C. (2020). Stanza: A Python Natural Language Processing Toolkit for Many Human Languages. In *Proceedings of the 58th ACL: System Demonstrations*.

[15] Chen, D. & Manning, C. (2014). A fast and accurate dependency parser using neural networks. In *Proceedings of the 2014 EMNLP*. (pp. 740-750).

[16] Manning, C., Surdeanu, M., Bauer, J., Finkel, J., Bethard, S., & McClosky, D. (2014). The Stanford CoreNLP Natural Language Processing Toolkit. In *ACL: System Demonstrations (pp. 55–60)*.

[17] Nivre, J., Marneffe, M.C., Ginter, F., Hajic, J., Manning, C., Pyysalo, S., Schuster, S., Tyers, F., & Zeman, D. (2020). Universal Dependencies v2: An Evergrowing Multilingual Treebank Collection. In *Proceedings of The 12th Language Resources and Evaluation Conference (pp. 4034–4043)*. European Language Resources Association.
[18] Kendall, Tyler and Charlie Farrington. 2020. The Corpus of Regional African American Language. Version 2020.05. Eugene, OR: The Online Resources for African American Language Project. http://oraal.uoregon.edu/coraal.