Microvariation in American English Negative Concord: A comparative corpus study

Frances Blanchette & Cynthia Lukyanenko*

Abstract. Negative Concord (NC) (e.g., *I didn't eat nothing*, 'I ate nothing') is often treated as categorical. However, observations of English vernacular varieties illustrate that NC is not the only option for expressing negation in a given variety. We present a comparative analysis of patterns of NC in corpora of two American English varieties, the Washington, DC portion of *The Corpus of Regional African American Language* and *The Audio-Aligned and Parsed Corpus of Appalachian English*. Results demonstrate quantitative differences in the frequency of NC overall, and qualitative differences in the expression of subject negation (e.g., *Didn't nobody eat*, *Nobody didn't eat*, *Nobody ate*). The results highlight similarities and intriguing differences among American English varieties in which NC is instantiated, and raise interesting questions about the source of the observed differences in usage patterns. We also preview an extended analysis that integrates data from a comparable portion of *The Corpus of Contemporary American English* representative of Mainstream US English. The results highlight important and unexpected similarities and differences between English varieties in which NC is and is not systematically employed in usage.

Keywords. negation; Negative Concord; African American Language; Appalachian English; syntax; variation

1. Introduction. Negative Concord (NC), where two or more syntactic negations contribute to a single semantic negation (e.g., *I didn't eat nothing*, "I ate nothing"), has been called an "English vernacular universal" (Nevalainen 2006). It is often treated as categorical: a grammar either has it or it does not (e.g., Tubau 2016; Wolfram & Fasold 1974). However, NC is not the only option for expressing negation in a given variety. In many English varieties, negative phrases occur variably in NC and in sentences with no additional syntactic negation (e.g., *I ate nothing*; Childs 2017; Smith 2001; Tubau 2016). Moreover, NC can appear in a variety of different structure types (e.g., *Nobody didn't eat*, *Didn't nobody eat*; Green 2014), and these structures may occur with varying frequency between varieties and even individual speakers.

The current study explores patterns of negation and NC in two state of the art corpora of vernacular American Englishes: *The Audio-Aligned and Parsed Corpus of Appalachian English* (AAPCAppE; Tortora et al. 2017) and the Washington, DC portion of *The Corpus of Regional African American Language* (CORAAL-DC; Kendall & Farrington 2023). Results demonstrate quantitative differences in the frequency of NC overall, and qualitative differences in the expression of subject negation, suggesting a broader diversity among American English NC varieties than has previously been observed. The results provide a foundation for important questions

* We are grateful to the many research assistants who assisted with the hand coding and checking of the data, including Abigail Salem, Grace Smith, Felicity Sarnoff, Luciano Ragonese, Alyson DeMarco, Lucia Soluri, Malcolm Patton, Madison Gillner, and Maya Asher. Thanks also to Mary Henderson, Benjamin Hunt, Maya Asher, and Rupak Kumar Das for their contributions to data extraction, organization, visualization and analyses. Special thanks to Jessi Grieser, Paul Reed, and the undergraduate research assistants at the University of Alabama who assisted with the early stages of data coding. This work was supported by funds from the Penn State Eberly College of Science and a National Science Foundation collaborative research grant (#2234039). Authors: Frances Blanchette, Pennsylvania State University (fkb1@psu.edu) & Cynthia Lukyanenko, George Mason University (clukyane@gmu.edu).

about the source of the differences observed, which may stem from grammatical, sociolinguistic or other factors.

Extending beyond the negative dependency type found in NC constructions, we also present some initial findings comparing NC usage and negative polarity items (NPIs) under negation (e.g., *anything*, as in *I didn't eat anything*). To do this, we curated a subcorpus of the Corpus of Contemporary American English (COCA; Davies 2008–) to represent Mainstream US English (MUSE). While this variety is not expected to instantiate NC usage, we do expect it to include NPIs under negation. Since NPI constructions are also available in Appalachian English (Blanchette 2015) and African American Language (Green 2002), we were able to compare rates of NC and NPI usage under negation, as well as the use of single negative arguments with no dependency relation as in *I ate nothing*. The results highlight the value of considering MUSE usage patterns in explorations of NC variation, despite the expected lack of NC in this variety.

2. Methods.

2.1. THE CORPORA: AAPCAPPE AND CORAAL. To better understand negative sentence patterning in vernacular varieties of American English, we capitalized on the development of two cutting-edge, freely available corpora: The AAPCAppE and the Washington, DC portion of CORAAL (DCA and DCB). Each of these is approximately one million words, making them readily comparable in terms of their size. They are also comparable on a number of other features, including their interview style as well as the fact that, for the most part, both interviewers and interviewees are from the communities under study, and this is reflected in their speech patterns.

Both AAPCAppE and CORAAL include high-quality, linguistically accurate orthographic transcriptions that are time-aligned with accompanying audio files, and are readily searchable–a helpful feature when transcriptions are ambiguous, since the audio can sometimes serve to disambiguate. The AAPCAppE also provides a syntactic parse for each sentence, and although we did not use this feature for our analysis since we wanted to use the same methods across corpora, we did sometimes reference it during our coding process in cases where the structure was not transparent from the transcription. Further detailed information on these corpora can be found on their respective websites (https://aapcappe.commons.gc.cuny.edu/project-description/, and https://oraal.github.io/coraal).

2.2. EXTRACTION AND CODING. We extracted all utterances with the negative phrases *nobody*, *nothing*, *none*, *no-one*, and *no* X (e.g., *no buses*) from each corpus. We also extracted utterances with noun phrase NPIs as well as the terms *never* and *ever*, though these are not the primary focus of the current study. We provide a preview of the NPI results below in section 3.4.

Following utterance extraction, the sentence containing the target word in each utterance was identified. This was the target domain for the initial round of hand coding reported here. Most extracted utterances were equivalent to sentences, but it was occasionally necessary to trim additional material or to check adjacent lines in the corpus to complete partial sentences.

Once sentences were identified, we proceeded to hand code them for a number of different features, including those shown below:

(i) *negative marker presence*: whether the sentence contains a negative marker (*not*, *-n't*)

(ii) *negative subject*: whether the sentence contains a negative subject

(iii) *subject position*: if a negative subject is present, whether it is preverbal or postverbal

(iv) *syntactic negation count*: the number of syntactic negations in the sentence

(v) *semantic negation count*: the number of semantic negations in the sentence

This coding procedure allowed us to identify and categorize sentences on a number different variables of interest. For example, since negative subjects can occur both pre- and postverbally in NC constructions (e.g., *nobody didn't...* vs. *didn't nobody...*), if a sentence had both a negative marker (i) and a negative subject (ii), then coding the subject position (iii) allowed us to identify which category a sentence fell into.

In addition, and importantly, a sentence with two negations is in fact ambiguous, and can have either a single negation NC interpretation or a double negation interpretation in which each syntactic negation contributes a semantic negation. For example, in a context in which a speaker wishes to deny an interlocutor's assertion that they ate nothing for breakfast, they might respond by saying *I didn't eat nothing for breakfast, I had a piece of toast*. In this case the two negations cancel each other to result in an affirmative. Such double negation constructions are heavily pragmatically conditioned and therefore expected to be relatively infrequent (Horn 2001). Nevertheless, we coded for both the number of syntactic negations (iv) and the number of semantic negations (v). Sentences in which the number of semantic negations was lower than the number of syntactic negations were classified as NC.

3. Results. The results of our coding procedure allowed us to classify negative sentences in each corpus into a number of different types, and compare frequencies for different types across corpora. Table 1 contains examples of the different sentence types we found containing a negative noun phrase, which are the focus of our main analyses here. Sentence types are labeled according to whether the negative phrase is a subject or a non-subject, as well as whether it is the only negative element in the sentence (single negation) or whether it occurs in concord with another negative element (NC). Note that each type is represented in each corpus. This suggests that differences between corpora are more likely to be quantitative than qualitative in nature.

Table 1. Examples of negative sentence types in AAPCAppE and CORAAL-DC

Sentence Type	AAPCAppE	CORAAL
non-subject single negation	*After I met your papaw, I had eyes for nobody else.* (SKCTC-BP-1.122)	*I have no particular favorite.* (DCA_se3_ag1_m_04)
non-subject NC	*I don't know nobody over there.* (ALC-SD-1.130)	*I didn't know nobody.* (DCB_se2_ag4_f_05)
subject single negation	*Nobody had water in the house.* (ALC-FR-2.144)	*Nobody voted for me.* (DCB_se1_ag2_m_01)
subject NC - preverbal	*Nobody didn't have water in the house.* (ALC-RN-1.83)	*Nobody didn't wanna help me...* (DCB_se1_ag4_f_01)
subject NC - postverbal	*Didn't nobody beat them.* (ALC-377-1.77)	*Don't nobody even call it Chocolate City anymore.* (DCB_se1_ag2_f_01)

3.1. NEGATIVE CONCORD OVERALL. Our first quantitative comparison across corpora explored the rates at which negative phrases occurred as part of an NC construction or as the single negative element in a sentence. As shown in Figure 1, NC proved to be a common feature in both corpora, appearing in the majority of sentences with a negative phrase (Figure 1).

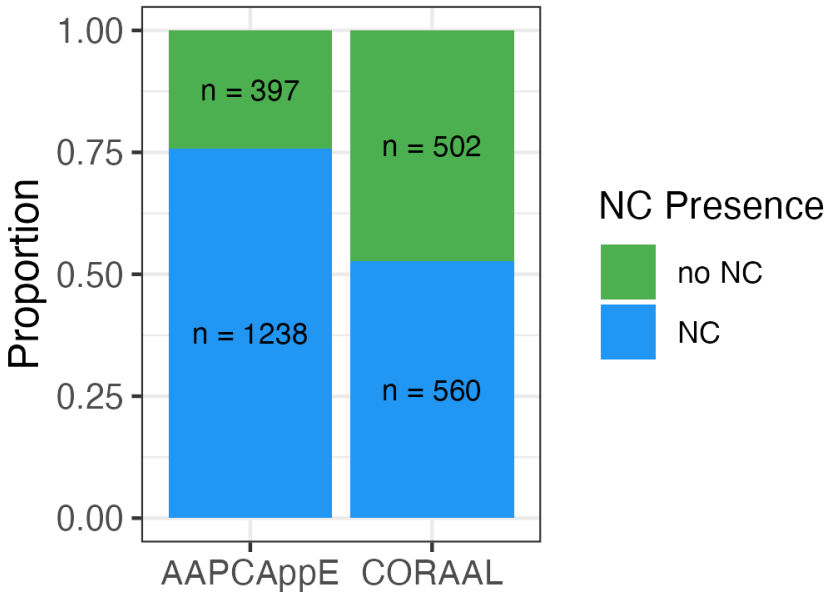

Figure 1. Rates of NC in sentences with a negative noun phrase across corpora.

Though NC was more common than single negation in both corpora, rates differed substantially. As shown in Figure 1, NC was more common in AAPCAppE (76%) than in CORAAL-DC (53%). This highlights the fact that, while negative noun phrases can occur in both NC constructions and as single negations in both corpora, the two varieties are not equivalent: NC appears to be more common in Appalachian English than in African American Language. This difference in overall rate of NC suggests that the grammatical or sociolinguistic status of NC may be distinct in these two "NC varieties" of American English, a possibility we return to below.

3.2. NEGATIVE NON-SUBJECTS VS. NEGATIVE SUBJECTS. Extending beyond overall NC rates, we further explored the extent to which NC occurred in sentences with negative noun phrases in subject as compared with non-subject positions. The results of this comparison are shown in Figure 2.

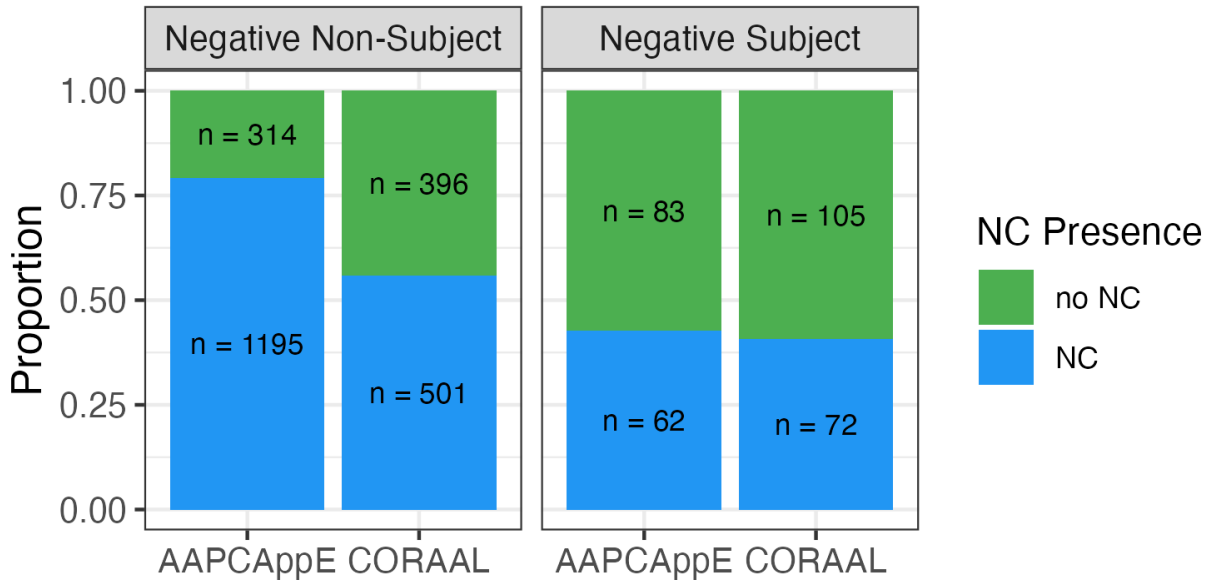

Figure 2. Rates of NC by subject and non-subject structure types.

As Figure 2 shows, across both corpora, NC constructions are less common in sentences with a negative subject than in sentences with a negative non-subject. Fewer than half of sentences with a negative subject are instances of NC, while the majority of sentences with a negative non-subject are. Both corpora clearly allow negative concord with negative subjects, but this structure is much less common than negative concord with non-subjects for two reasons. First, negative subjects are less frequent than negative non-subjects overall, and second, negative concord is less common with negative subjects than negative non-subjects. The combination of these two factors leads to a nearly 20-to-1 prevalence of non-subject NC to subject NC in AAPCAppE, and approximately a 7-to-1 prevalence in CORAAL-DC.

Figure 2 highlights a further difference between AAPCAppE and CORAAL, which helps clarify the source of the differences in overall NC rate seen in Figure 1. Rates of NC with negative subjects are similar, 43% in AAPCAppE and 41% in CORAAL-DC. The differences in overall NC rates come primarily from the negative non-subject sentences, in which NC occurs 79% of the time in AAPCAppE and only 56% of the time in CORAAL-DC. This shows that the substantial difference in overall NC rate across AAPCAppE and CORAAL is driven primarily by negative non-subjects.

3.3. ZOOMING IN ON NEGATIVE SUBJECTS. As shown in Table 1 above, negative subjects occur in three possible structure types: subject single negation (e.g., *nobody had water*), subject NC with a preverbal subject (e.g., *nobody didn't have water*), and subject NC with a postverbal subject (e.g., *didn't nobody have water*). Thus, while negative subjects occur in NC and as the only negation in a sentence at similar rates across corpora, as shown in 2, there are different possibilities for how frequency distributions may occur within subject NC across corpora. We therefore conducted an additional comparison of the rates at which each negative subject sentence type occurs in each corpus, the results of which are shown in Figure 3.

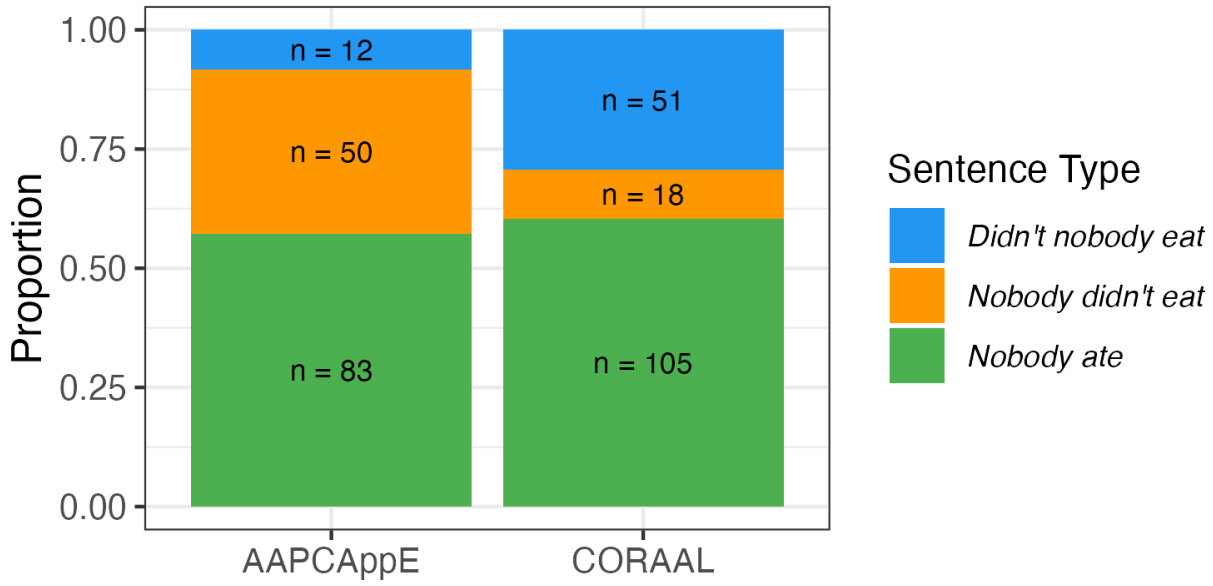

Figure 3. Rates of sentence types with negative subjects across corpora.

As shown in Figure 3, there is an asymmetry in how subject NC occurs across corpora. In particular, while NC with a preverbal negative subject (*nobody didn't eat*) is more common in AAPCAppE, in CORAAL postverbal negative subjects (*didn't nobody eat*) are the more common NC type. The predominance of preverbal subject NC in AAPCAppE and postverbal subject NC in CORAAL corroborates previous observations made in the literature (Green 2002, 2011, 2014; Wolfram & Christian 1976). It is possible that the quantitative difference illustrated here reflects an additional qualitative difference between the two varieties. Before returning to this in the discussion below, we present some preliminary data from an analysis that builds on the work presented above, integrating an additional type of negative sentence and an additional variety of American English.

3.4. Comparison with Negative Polarity Items Under Negation. In addition to extracting sentences with negative words from AAPCAppE and CORAAL, we extracted parallel sentences from a curated portion of the *Corpus of Contemporary American English* (COCA; Davies 2008–). Since this corpus is over 400 million words, we compiled a subcorpus of approximately 1 million words. In order to make it as comparable as possible to the other corpora under study, we selected a subset of the *60 Minutes* transcripts contained in the spoken portion of COCA. Their frequent use of interviews in a more formal television news setting represents a reasonable parallel to the sociolinguistic interview format used in AAPCAppE and CORAAL, while still being a context that should strongly favor the use of mainstream English.

To build this subcorpus, we performed a random sort on the list of available transcripts, and then iteratively included or excluded transcripts until we reached 1 million words. Transcripts were excluded if the interviewee was not a native speaker of American English or if it contained linguistic features that are absent in Mainstream US English (MUSE). Interviewees' status as native speakers of American English was determined to the best of our ability on the basis of descriptions within the interviews, cross-referenced where possible with online bios or other news sources. The presence of linguistic features absent in MUSE was determined by a search for the

forms *hisself*, *ain't*, *y'all*, *youse*, *he/she don't*, *they/we was*, and *I were*. As constructed, we take this COCA subcorpus to be a proxy for a mainstream variety of US English (MUSE).

While NPIs come in a variety of different forms in English, we extracted only those noun phrases containing the form *any*, including *anybody*, *anything*, *anyone*, and *any* X (e.g., *any buses*), generating a set that was parallel to the *no-* forms discussed above. Importantly, while we do not expect NC to occur in the MUSE speech represented in our COCA subcorpus, we do expect both single negative phrases and NPIs under negation. With the addition of NPIs, we then have three possibilities or "variants" for the expression of negative sentences with negative phrases and NPIs in non-subject position, which are illustrated in Table 2. While all three variants in Table 2 are possible in American English, only two of them are expected in all three corpora: single negation and the NPI under negation structure. NC stands out as the only variant that is not expected in MUSE.

Table 2. Three negative sentence variants.

Sentence Type	Example	Expected Distribution
negative concord	I didn't eat nothing.	vernacular, non-standardized (AAPCAppE, CORAAL)
single negation	I ate nothing.	vernacular and mainstream (all corpora)
negative polarity item	I didn't eat anything.	vernacular and mainstream (all corpora)

To understand how these three variants were distributed across the three corpora, we performed a round of semi-automated coding distinct from the coding reported above. For this round of coding we focused specifically on finite clauses as opposed to entire sentences, since finite clauses tend to be a maximal domain within which NC relations and NPI dependencies occur (Déprez 1997; Ladusaw 1979). We used a Python script to automate the coding of numerous features, and performed hand and spot checking of a subset of items. We further compared the results of our semi-automated coding with our previous hand coding of the negative phrase sentences to confirm that the semi-automated procedure was not generating substantially different results.

For the purpose of this paper, we preview the subset of the data containing either a negative phrase or an NPI in a postverbal, non-subject position.[1] For the NPI sentences, we focused specifically on NPIs under negation such as the one in 2, setting aside sentences with NPIs occuring in non-negative contexts (e.g. conditionals, interrogatives). Figure 4 illustrates how the three negative sentence variant types illustrated in 2 were distributed across the three corpora.

[1] We are currently preparing a manuscript with the full methods, results and analysis for our semi-automated coding.

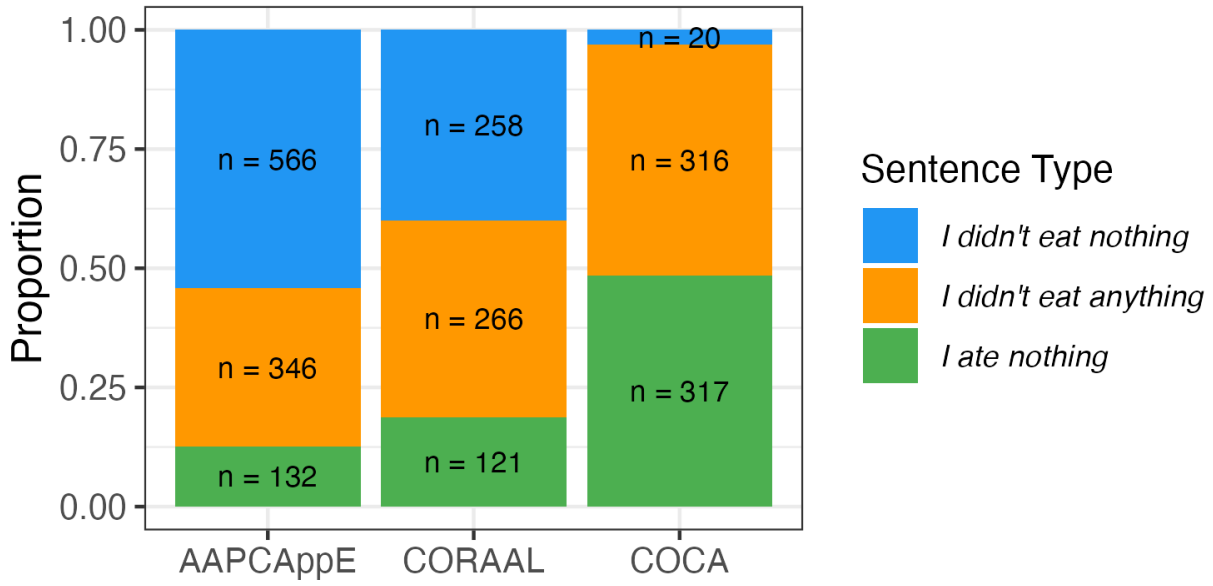

Figure 4. Rates of postverbal phrase types across all corpora.

Figure 4 shows that sentences with a negated auxiliary and a negative postverbal non-subject (e.g., *I didn't eat nothing*) appeared frequently in both AAPCAppE (n=566, 54% of total tokens) and CORAAL (n=258, 40% of total tokens) but were vanishingly rare in COCA (n=20, 3% of total tokens). This difference is driven by the relative lack of NC in MUSE, and its systematic use in Appalachian English and African American Language.

Given that NC is not expected to occur in MUSE, perhaps the more interesting comparison that this analysis affords is between rates of NPIs under negation and single negative noun phrases in a non-subject position. For the two vernacular corpora under study we see similar patterns: NPIs under negation are more frequent, at 33% in AAPCAppE and 41% in CORAAL, while single negation occurs less frequently at 13% in AAPCAppE and 19% in CORAAL.

Since NC is not expected in MUSE, and much research on NC has highlighted its similarity with NPIs under negation and their underlying distinctness from single negation (Burnett et al. 2018), we might predict that NPIs under negation would be the most common form in our COCA data. However, this is not what we observe. Instead, we see NPIs under negation and single negation being used at similar rates: NPIs under negation at 48%, and single negation at 49%. Put differently, if we think about the NC constructions in the vernacular corpora 'trading off' with a construction in the COCA subcorpus, it looks like more of the trade-off is happening with single negation than NPI constructions. There is a much larger difference in the green bars across corpora in Figure 4 than in the orange bars. While further research is needed to understand these patterns, they suggest that new insights might be gained into patterns of variation by comparing across varieties in which NC is and is not used.

4. Summary and Discussion. The primary analyses presented in this paper illustrate the landscape of negative phrases in two state of the art corpora of vernacular American English: the AAPCAppE and CORAAL. We observed high rates of NC use in both corpora with these negative phrases, with a strong preference for NC emerging in AAPCAppE but not CORAAL (Figure 1). There are a number of possible explanations for this difference, which might link it to social

or linguistic factors or some combination of these. Leaving this matter for future research, we note that this initial observation opens up a range of possibilities that could lead to further insight into the nature of variation in negation.

Finer grained analyses of a range of different structures revealed an even more complex and interesting picture. One novel finding that emerged from our analyses was the fact that negative non-subjects—that is, negative phrases in a non-subject position—were driving the differences observed in overall NC rates. This is because, while speakers in both corpora used NC with negative subjects at similar rates, in non-subject positions AAPCAppE speakers were more likely to choose the NC variant than their CORAAL counterparts (Figure 2). The fact that this global difference in NC rates is ascribed to postverbal positions makes sense, since there is another commonly employed option—NPIs under negation, that can also be used. Speakers thus have more options in non-subject positions than in subject positions (cf. **anybody didn't eat*, with a preverbal NPI subject). The question of why AAPCAppE speakers more frequently chose the NC variant in postverbal position remains, however. We are currently conducting additional analyses with the data presented here to explore possible causes of this variation.

Further systematic variation was revealed through exploration of the quality and quantity of structures with a negative subject in the AAPCAppE and CORAAL. We found that negative subjects occurred in NC constructions slightly less than half the time in both corpora. Intriguingly, despite this quantitative similarity, there were qualitative differences: negative subjects in AAPCAppE tended to occur preverbally, whereas in CORAAL-DC they tended to occur after a negated auxiliary (Figure 3), a pattern known as Negative Auxiliary Inversion (Green 2014). This difference in the quality of NC sentences with a negative subject across corpora might represent a grammatical divide across these two varieties, a possibility we are investigating using experimental methods (for a preview, see Lee et al. 2026: this volume).

In a final set of preliminary analyses of variation between NC, sentences with a single negative phrase, and NPIs under negation, we illustrated the interesting patterns that emerge when a mainstream variety of US English is brought into the comparison (Figure 4). In current work we are further exploring these patterns, which we hope will provide new insight into both social as well as grammatical factors shaping speakers' choice among these possible variants.

The results presented here contribute to a growing body of research demonstrating variety-internal variation in the expression of English negative sentences (Childs 2017; Tubau 2016; Burnett et al. 2018). These findings open the way for further inquiry into whether frequency differences in the expression of negative sentences with a negative subject reflect grammatical distinctions, as well as how and why social factors may condition the variation. More broadly, the results highlight the rich grammatical diversity that exists within American English varieties, and the various ways in which speakers may employ their broad linguistic repertoire to express negation.

References

Blanchette, Frances. 2015. *English negative concord, negative polarity, and double negation*. New York, NY, USA: CUNY Graduate Center Ph.d. dissertation.

Burnett, Heather, Hilda Koopman & Sali A. Tagliamonte. 2018. Structural explanations in syntactic variation: The evolution of English negative and polarity indefinites. *Language Variation and Change* 30(1). 83–107. https://doi.org/10.1017/S0954394517000266.

Childs, Claire. 2017. Integrating syntactic theory and variationist analysis: The structure of neg-

ative indefinites in regional dialects of British English. *Glossa: a journal of general linguistics* 2(1). 106. https://doi.org/10.5334/gjgl.287.
Davies, Mark. 2008–. The Corpus of Contemporary American English (COCA): 520 million words. 1990–present. https://www.english-corpora.org/coca/. Accessed: 2020-01-30.
Déprez, Viviane. 1997. Two types of negative concord. *Probus* 9(2). 103–143.
Green, Lisa. 2014. Force, focus and negation in African American English. In Raffaella Zanuttini & Laurence R. Horn (eds.), *Micro-Syntactic Variation in North American English*, 115–142. New York: Oxford University Press.
Green, Lisa J. 2002. *African American English: A linguistic introduction*. Cambridge, UK: Cambridge University Press.
Green, Lisa J. 2011. *Language and the African American child*. Cambridge: Cambridge University Press.
Horn, Laurence R. 2001. *A natural history of negation* The David Hume Series: Philosophy and Cognitive Science Reissues. Stanford, CA: CSLI Publications.
Kendall, Tyler & Charlie Farrington. 2023. The Corpus of Regional African American Language (CORAAL). https://oraal.github.io/coraal. https://doi.org/10.7264/1ad5-6t35. Version 2023.06, The Online Resources for African American Language Project, Eugene, OR. Accessed: 2026-03-04.
Ladusaw, William A. 1979. *Polarity sensitivity as inherent scope relations*: University of Texas at Austin dissertation.
Lee, Brittany, Cynthia Lukyanenko, Frances Blanchette, Jessica A. Grieser & Paul E. Reed. 2026. Microvariation in the interpretation of English negative sentences. In *Proceedings of the linguistic society of america annual meeting*, New Orleans, LA. Presented at the Annual Meeting of the Linguistic Society of America, January 8–11, 2026.
Nevalainen, Terttu. 2006. Negative concord as an English "vernacular universal": Social history and linguistic typology. *Journal of English Linguistics* 34(3). 257–278. https://doi.org/10.1177/0075424206293144.
Smith, Jennifer. 2001. Negative concord in the old and new world: Evidence from Scotland. *Language Variation and Change* 13(2). 109–134. https://doi.org/10.1017/S0954394501132011.
Tortora, Christina, Beatrice Santorini, Frances Blanchette & CEA Diertani. 2017. The Audio-Aligned and Parsed Corpus of Appalachian English (AAPCAppE). http://csivc.csi.cuny.edu/aapcappe/.
Tubau, Susagna. 2016. Lexical variation and Negative Concord in traditional dialects of British English. *The Journal of Comparative Germanic Linguistics* 19(2). 143–177.
Wolfram, Walt & Donna Christian. 1976. *Appalachian speech*. Washington, D.C.: Center for Applied Linguistics.
Wolfram, Walt & Ralph W Fasold. 1974. *The study of social dialects in American English*. Prentice Hall.