A variationist approach to subject-aux question inversion in Bajan and other Caribbean creole Englishes, AAVE and Appalachian

JOHN R. RICKFORD AND ROBIN MELNICK
Stanford University

ABSTRACT

Our quantitative, variationist (Rbrul) analyses of subject-auxiliary inversion in question formation across three Caribbean creole Englishes (Guyanese, Jamaican, and Bajan, the mesolectal creole English of Barbados) and two North American vernacular Englishes (AAVE and Appalachian) combine with prior comparable (Varbrul) analyses of this variable in Samaná and African Nova Scotia English to illustrate the patterning of its variable constraints. Previously, canonical descriptions held that English-based creoles differ from standard and metropolitan colloquial Englishes by forming questions with rising intonation alone (e.g., Guyanese: Jaan de hoom “John is at home”), without subject-auxiliary inversion (Is John at home?). In practice, however, we find considerable, and substantially similar, constraint-based variation across the varieties studied, with YES/NO questions and auxiliary do favoring non-inversion. These and other cross-varietal similarities are significant for understanding this variable but suggest that question formation may not be fruitful for investigating the question of prior creolization in AAVE.

1. INTRODUCTION

Our goal in this paper is to provide a variationist account of subject-auxiliary inversion in question formation in several creole and vernacular varieties of English in which this variable has not previously been studied, beginning with Bajan, the English-based mesolectal creole of Barbados. In this introductory section we contrast the canonical characterization of question formation in English-based creoles with the variation we encounter in recorded data. Next, we review prior studies of this variable in British and New World Black Englishes. After introducing our data and methods we present a multivariate analysis, first addressing Bajan then extending it to Jamaican, Guyanese, African American Vernacular English, and Appalachian.

In theory, the contrast in question-formation between standard English and English-based creoles is absolute. Standard English forms its questions by inverting the subject NP and the auxiliary verb (Huddleston & Pullum 2002: 95), both in YES/NO questions (1b) and WH-questions (1c). If the corresponding declarative statement has a regular main verb rather than an auxiliary, as in (1d), do is inserted to facilitate question inversion, as in (1e):

(1) a. [John]_{NP} [is]_{AUX} at home.
b. \([Is]_{\text{aux}} [\text{John}]_{\text{NP}} \text{ at home?}\)

c. Where \([is]_{\text{aux}} [\text{John}]_{\text{NP}}\)?

d. \([\text{Robin}]_{\text{NP}} \text{ likes Barbados.}\)

e. \([Does]_{\text{aux}} [\text{Robin}]_{\text{NP}} \text{ like Barbados?}\)

By contrast, in English-based creoles, as in creoles more generally, “there is no inversion of the subject and auxiliary … to form questions” (Holm 1988: 212). Statements and YES/NO questions are said, in fact, to show no “difference in syntactic structure” (Bickerton 1981: 70),\(^2\) distinguishable only by rising intonation, as in (2a/b). WH-questions are marked by the presence of a WH-word at the beginning, optionally preceded by a copula-like focus marker, like \(da\) or \(a\) (cf. Winford 2008: 29), and by a corresponding gap in the question (2c):\(^3\)

\[(2)\]
\[
a. \text{Jaan de}_{[V-\text{LOC}]} \text{ hoom.} \\
\text{‘John is at home.’}
\]
\[
b. \text{Jaan de}_{[V-\text{LOC}]} \text{ hoom?} \\
\text{‘Is John at home?’}
\]
\[
c. (a) \text{we}_{1} \text{ Jaan de } \underline{____1}? \\
\text{‘Where is John?’}
\]

However, in practice, the contrast between English and Creole question formation is not that clear-cut. On the one hand, mainstream English in the UK and US sometimes forms questions without subject-aux inversion, often, but not always, with rising intonation (Bolinger 1957; Visser 1969):

\[(3)\] Your aim that evening, then, was to go to the discotheque? (Huddleston & Pullum 2002: 868, 6.1)

On the other hand, we sometimes get subject-aux inversion in Caribbean questions whose syntax, phonology, and lexicon would lead us to classify them as “creole,” as in this Jamaican example:

\[(4)\] \(Iz \text{ it dat } \underline{\text{dem}} \text{ laik unu av waar wid di ada komyuuniti } \underline{\text{dem}}?\) (JA22.315)\(^4\)
\text{‘Is it that they like you all to have war with other communities?’}

\(^2\) “No creole shows any difference in syntactic structure between questions and statements.” (Bickerton 1981:70)

\(^3\) Holm (1988: 213) summarizes our discussion so far succinctly and forcefully: “while the creoles’ lexical source languages often require the inversion of the subject and the verb (or auxiliary) to transform a statement into a question, this is not a part of creole syntax.”

\(^4\) Notation for referencing examples from our data set follows the format \(xx##.##\), where “\(xx\)” is a two-letter identifier for the language variety (“AA”=African American; “AP”=Appalachian; “BA”=Bajan; “GC”=Guyanese; “JA”=Jamaican), the number before the period identifies the recording, and the number after indicates line number(s) in the transcript.
Moreover, in mesolectal or intermediate creole-speaking communities like Barbados (see 5) and among speakers of semi-creole or African American Vernacular English (see 6), this kind of variation occurs even more often, as in:

(5)  
   a.  Ma, *could* I leff de corner now? (BA29.959, inverted)  
   b.  So you *don’t* like it stir up? (BA26.1059, uninverted)  

(6)  
   a.  *Is* Patricia, Ladon sister? (EPA8.293, inverted)  
   b.  You *don’t* believe me? (EPA7.2100, uninverted)  

Quantitative studies indicating how often question inversion occurs in standard, vernacular, and creole Englishes, and what factors condition it, are either rare or non-existent. The purpose of the present study is to remedy this situation, concentrating first on Bajan (Barbadian creole English), but then considering Jamaican, Guyanese, Appalachian and African American Vernacular English, as part of a larger study of grammatical variation and change.5

2. PRIOR QUANTITATIVE STUDIES OF SUBJECT-AUX INVERSION IN ENGLISH QUESTIONS

Especially relevant is Ellegård’s (1953) detailed study of the rise of auxiliary *do* between 1400 and 1700, including the process by which inversion of the main verb in questions like (7) gave way to inversion with auxiliary *do*, as in (8):

(7)  By what adventure *com* ye hidir? (Malory 1480: 75-10, as cited in Ellegård 1953: 205)  
(8)  ... *doe* you speake this seriously? (Johnson, ca. 1599: 436, as cited in Ellegård 1953: 205)

The Ellegård data reveal that auxiliary *do* was established earliest and was most frequent in YES-NO questions (“*Went he?*” > “*Did he go?*”); was established somewhat later and was less frequent in adverbial WH questions (“*When came he?*” > “*When *did* he come?”); and arrived latest and was least frequent in WH-object questions (“*What said he?*” > ”*What *did* he say?”).

Ellegård’s study provided the conditioning factors and data for several subsequent variationist studies, including Stein (1988), Kroch (1989), and Van Herk (2000). However, the “inversion” that Ellegård examined was the archaic form seen above in (7), with main verb

---

5 This research was facilitated by National Science Foundation Grant No. 0545424 to PI John R. Rickford from 2006-2010, which is gratefully acknowledged. Robin Melnick was the leading graduate Research Assistant for the question inversion portion of the project. We thank the anonymous referees for this paper and, at Stanford, Professors Tom Wasow and Ewart Thomas and many students, including Rachel Christy, Rebecca Greene, Catherine Howard, Zoe Lidstrom, Kimberley McKinson, Mackenzie Price, and Lisa Young. In Barbados, thanks are due to Professors Jeannette Allsopp and Hazel Simmons-McDonald of UWI Cave Hill, Paulette and Junior Gooding, and UWI students Clement “Tony” DePeiza, Jennifer Critchlow, Ann Fergusson (especially), Undine Shorey, and Romel Springer, among others.
before subject. Ellegård made no suggestion that the factors conditioning do-support as an alternative to now-archaic main-verb inversion might apply to copular or modal inversion.

Table 1 summarizes the findings of four more-recent quantitative works on question inversion in varieties of New World Black English, from Labov et al.'s (1968) study of African American Vernacular English in NYC to Van Herk's (2000) study of Samaná English and African Nova Scotia English, which he and other contributors to Poplack (2000) treat (along with the US Ex-Slave Recordings) as representative of “Early African American English.” Of these studies, only Van Herk provides a multivariate, variationist account of question inversion, complete with variable-rule factor weights, so we will give it primary coverage in our discussion. Note, though, that the relative frequencies of inversion in Table 1 vary widely (e.g., 8% for WH-questions in Samaná according to DeBose (1996) vs. 61% according to Van Herk) and that WH-questions show more inversion than YES/NO questions in some varieties and less in others. These fluctuations depend to some extent on methodology, to which we now turn.

### Table 1. Question inversion in New World Black Englishes

<table>
<thead>
<tr>
<th></th>
<th>AAVE</th>
<th>AEBE</th>
<th>Samaná_a</th>
<th>Samaná_b</th>
<th>ANSE</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>YES-NO questions</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Inverted - <em>Is he here?</em></td>
<td>no data</td>
<td>91</td>
<td>88</td>
<td>29</td>
<td>14</td>
</tr>
<tr>
<td>Non-inverted - <em>I can play?</em></td>
<td>no data</td>
<td>9</td>
<td>88</td>
<td>71</td>
<td>14</td>
</tr>
<tr>
<td>Ambiguous - (*Ø?) He (<em>Ø?) going?</em></td>
<td>no data</td>
<td>7</td>
<td>37</td>
<td>70?</td>
<td>100?</td>
</tr>
<tr>
<td><strong>WH-questions</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Inverted - <em>Why is he here?</em></td>
<td>80</td>
<td>52</td>
<td>73</td>
<td>52</td>
<td>8</td>
</tr>
<tr>
<td>Non-inverted - <em>Why I can’t play?</em></td>
<td>20</td>
<td>52</td>
<td>27</td>
<td>52</td>
<td>92</td>
</tr>
<tr>
<td>Ambiguous - *Where (*Ø?) he (<em>Ø?) at?</em></td>
<td>45</td>
<td>18</td>
<td>31</td>
<td>35?</td>
<td>220?</td>
</tr>
</tbody>
</table>

**Note:** AAVE = African American Vernacular English (Labov et al. 1968: 293-6); AEBE = American Earlier Black English (ex-slave narratives) (Schneider 1989: 205-7); Samaná = Samaná English (Dominican Republic, descendants of African American who emigrated in the 19th century; a = DeBose [1996]; b = Van Herk [2000: 177, 181]; ANSE = African Nova Scotian English (Canada, descendants of African Americans who emigrated in the late 18th and 19th centuries) (Van Herk 2000: 177, 181). Samaná and ANSE studies provided totals for Ambiguous (zero copula/auxiliary), but their distribution across YES/NO and WH are estimated here from overall proportions of these question types in the respective studies, hence the corresponding question marks.

---

6 Indeed, Ellegård (1953: 201) says that “the commonest explanation of the establishment of the do form refers to the tendency to avoid inversion” (i.e., of the main verb and subject). And Van Herk (2000), noting that the main verb retains its SVO word order in questions with auxiliary do, calls them “uninverted” and compares them with Samaná and ANSE non-inversion, but as Rickford (2006: 136) notes, this is problematic, since “do insertion questions … are, from the modern/current perspective, and from the perspective of the entire sentence, inverted forms.”
3. METHOD

3.1. Data

The data in this section are from a series of interviews with native speakers of Bajan (Barbadian creole English), recorded in Barbados between 1987 and 2007.

3.2. Exclusions

As in any variation study, we exclude a number of forms (as “Don’t Count” cases), either because they are indeterminate or because they are categorical (in this case allowing only the uninverted variant). Following the only other multivariate, variationist account of question inversion (Van Herk 2000), we excluded most of the types that Van Herk did -- fragments (missing subject or verb), as in (9), fixed expressions (10), questions with tags (11), clarification checks (12) -- and some others: scripted (interviewer) questions (13), *what if/what about* questions (14), and questions with *wh*-subject and full verb (which retain uninverted word order of statements, cf. Huddleston and Pullum [2002]), as in (15).

(9) Five o’clock a morning? (BA31.186)

(10) a. Ya know? (BA31.348)
     b. You understand? (BA31.770)

(11) You said secondary, right? (BA33.131)

(12) If I know anybody that ever get rob? (BA33.922)

(13) What is your name? (BA33.8)

(14) Wuh if I ain’t know? (BA33.138)
     ‘What if I don’t know?’

(15) And wha happen? (BA31.248)

One type that we followed Van Herk (rather than DeBose) in excluding is “ambiguous questions” with a zero copula or auxiliary, “in which it is impossible to determine whether the original auxiliary [or copula] was inverted … prior to deletion” (Van Herk 2000:178):

(16) a. [Ø?] It [Ø?] good? (BA29.880)
     b. Where [Ø?] she [Ø?] livin now? (BA27a.347)

We excluded 84 Bajan questions as “ambiguous” by this criterion and used it to classify tokens in earlier studies as “ambiguous,” too, as in Table 1.
We did not follow Van Herk in excluding questions with regular main verbs, as in (17a). These only surface with \textit{do} support when inversion occurs, as in (17b). There are no non-emphatic questions with uninverted \textit{do} after the subject (cf. the ungrammatical 17c):

(17) a. You \textit{think} dat I could go pon the man work today and talk like this? (BA26.1155)

b. \textit{Do} you \textit{think} that I could go…?

c. *You \textit{do} think that I could go …?

Since (17a) \textit{is} the uninverted version of questions with regular main verbs, there is no reason to set aside sentences such as (17a) as “ambiguous”. Only sentences like (17b) will count as inversion.

Our Bajan question data set totaled 1003 questions after Don’t Count cases and ambiguous copula tokens were excluded. We also found other “Knockout” types that allow only the non-inverted variant, including a number of vernacular or creole verbs or aspect markers:

(18) a. \textit{Got}: Why yuh \textit{got} to run home? (BA27b.486)

b. \textit{Ain’t}: You ain’t see them? (BA11/12.1147)

c. \textit{Does} (unstressed, habitual): You \textit{does} use slang? (BA33.1034b)

d. \textit{Did} (unstressed, anterior/past): You \textit{did} ever sick bad anytime? (BA26.919)

Setting these aside, we retained 840 tokens for our variable analysis, substantially more than earlier variationist studies of question formation in the African American diaspora.

3.3. Modeling variation

The variable models of Samaná and African Nova Scotia English in Van Herk (2000) employed factors identified by Ellegård (1953) for the rather different phenomenon of the rise of Early

\footnote{Van Herk (2000: 178) treats sentences like (a) “Where you-all come from?” as resulting from \textit{did} deletion in (b) “Where \textit{did} you-all come from?” or (c) “Where you all \textit{did} come from?” We disagree for the Bajan case, and perhaps more generally. The (c) type, with an anterior \textit{did} that does not derive from \textit{do-support} but has an independent anterior meaning, is not semantically equivalent to the (a) type. Moreover, the \textit{did} in the (b) type, from \textit{do-support}, is more likely to be stressed and less likely to undergo deletion. Van Herk’s argument in his 2007 presentation for the Society for Pidgin Creole Linguistics (we thank him for generously sharing his slides with us) that zero past marking in sentences like (a) is higher than past marking outside of questions does not seem to apply to Bajan and does not take into account that non-inversion, like non-past marking, marks register differences. That is, sentences without inversion are also more likely not to be past-marked, other things being equal.}

\footnote{Compare 92 tokens for Samaná and 500 tokens for ANSE (Van Herk 2000). Since interviewers ask many more questions in the typical sociolinguistic interview than interviewees do, this particular variable is plagued by problems of limited data. In our data, the problem was alleviated to a large extent by the fact that most of the interviewers (in the case of Bajan, Guyanese, Jamaican, and AAVE) were members of the target speech community and were speaking informally. This may also have been true of the ANSE data used in Van Herk (2000).}

\footnote{See Rickford (2006: 133 including footnote 39), for more discussion.}
Modern English *do*, but these can provide at least a starting point for our analysis of present-day Anglophone Caribbean varieties.

We considered the effect of polarity, opposing negative (19a) and affirmative (19b) questions:

(19)  a. You *don’t* help you self? (BA31.975)

   b. So, *do* you remember at school you friends and ting you had? (BA26.361)

We coded for the type of question (YES/NO [20a] vs. WH [20b]), and added to the prior work’s scheme by further distinguishing WH-objects (20b) from *what for* and *why* causatives (20c) and other adverbials such as *where* and *when* (20d):

(20)  a. *Were* you good at them games? (BA27b.163)

   b. *What* do you feel the fellas doing? (BA29.214)

   c. *What* you did frighten *for*? (BA26.353)

   d. *Where* that did? (BA26.94)

We examined subject type, contrasting pronoun (21a) vs. full NP subjects (21b):

(21)  a. So what did *they* use to do then? (BA29.772)

   b. So how old *de children* is? (BA31.265)

Following Rickford, Wasow, Mendoza-Denton & Espinosa (1995), we also further distinguished extended NPs -- conjoined or with a prepositional phrase (22a) -- and *sentential* NP subjects (22b):

(22)  a. *Any of them* did any bad fight? (BA26.583)

   b. *The one who love you* will make you happy? (BA31.834)

We also added a factor group for subject length (number of words), since significant length (or “weight”) effects have been found in a number of syntactic studies (Bresnan 2007; Jaeger 2006, Wasow, Jaegar & Orr 2011).

Next, our coding of auxiliary included the following distinctions:

(23)  a. Modal: *Could* I leff the corner now? (BA29.959)

   b. Copular *be*: That *is* you favorite? (BA27b.1480)

   c. Verbal aux *be*: So he *is* going school still? (BA31.275)
d. Auxiliary have: (Ø) You ever had a real scary dream then? (BA27b.1151)

e. Auxiliary do: Do you think that they should teach it to children in schools? (BA29.292)

Finally, we controlled for demographic factors including gender, role (interviewer vs. non-interviewer), approximate socio-economic status (working class vs. middle class), and age.

4. RESULTS

4.1. Inversion rate

After removing ambiguous and other Don’t Count tokens, our Bajan speakers produced non-inverted questions 95% of the time. While this is far from the 0% that stipulative descriptions assert for standard English, neither is it the categorical 100% that stipulative descriptions of creole would lead us to expect. Still, Barbados English is popularly considered to be closer to standard English than the more deeply basilectal Jamaican or Guyanese creoles, so 95% non-inversion is higher than we might have expected. Even if we look at the non-inversion rate for standard English more realistically as the non-categorical 14% found in the Switchboard corpus of colloquial American English (Rickford 2006: 134), 95% non-inversion in Bajan appears comparatively close to categorically non-inverted.

A further consideration is that we counted many questions as non-inverted where we had a lexical main verb without auxiliary do, neither explicitly inverted nor explicitly non-inverted since non-inverted do appears only in emphatic forms. If we remove these, the non-inversion rate for Bajan would be slightly lower: 91%. Additional support for keeping these in, however, comes from the fact that many of these questions include modals that are unambiguously non-inverted:

(24) a. But you could remember Hutchinson? (BA11/12.61a)

b. What you would do different? (BA26.786)

c. I say, “Yuh kyaan do dat?” (BA28b.11a)

4.2. Regression analysis

We fit a generalized mixed-effects linear regression model to the data using the Rbrul package11 (Johnson 2008, 2013), with speaker as a random variable to control for individual variation. Table 2 displays our results.

---

TABLE 2. Variable rule (Rbrul) analysis of subject-aux non-inversion in Bajan questions

<table>
<thead>
<tr>
<th>Corrected mean/Total N</th>
<th>0.63$^{12}$</th>
<th>95</th>
<th>840</th>
</tr>
</thead>
<tbody>
<tr>
<td>POLARITY</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Negative</td>
<td>[.50]</td>
<td>99</td>
<td>83</td>
</tr>
<tr>
<td>Affirmative</td>
<td>[.50]</td>
<td>95</td>
<td>757</td>
</tr>
<tr>
<td>QUESTION TYPE</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>YES/NO</td>
<td>.80</td>
<td>99</td>
<td>560</td>
</tr>
<tr>
<td>WH-adverbial</td>
<td>.65</td>
<td>95</td>
<td>81</td>
</tr>
<tr>
<td>WH-object</td>
<td>.32</td>
<td>86</td>
<td>193</td>
</tr>
<tr>
<td>Causatives</td>
<td>.23</td>
<td>83</td>
<td>6</td>
</tr>
<tr>
<td>Range</td>
<td></td>
<td>57</td>
<td></td>
</tr>
<tr>
<td>SUBJECT LENGTH</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(logodds)</td>
<td>-0.90</td>
<td>96</td>
<td>840$^{13}$</td>
</tr>
<tr>
<td>AUXILIARY</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Do</td>
<td>.77</td>
<td>99</td>
<td>632</td>
</tr>
<tr>
<td>Modals</td>
<td>.48</td>
<td>96</td>
<td>94</td>
</tr>
<tr>
<td>Copula be</td>
<td>.25</td>
<td>76</td>
<td>114</td>
</tr>
<tr>
<td>Verbal aux be</td>
<td>K/O</td>
<td>100</td>
<td>5</td>
</tr>
<tr>
<td>Ain’t</td>
<td>K/O</td>
<td>100</td>
<td>11</td>
</tr>
<tr>
<td>Range</td>
<td></td>
<td>52</td>
<td></td>
</tr>
</tbody>
</table>

Note: Shaded areas indicate factors found significant using stepwise regression. Factor weights in square brackets represent non-significant groups, using values from the first step-down run, where all factors are forced into the regression. Factor weights favoring non-inversion are indicated in boldface. Demographic categories revealed no significant variation.

Three factors significantly influence variation: question type, subject length, and auxiliary type. YES/NO and WH adverbial questions favor non-inversion; increasing subject length has a very strong effect for inversion; and copular forms favor inversion, while do forms favor non-inversion.

With respect to question type, Kroch (1989) similarly found YES/NO questions most strongly favoring non-inversion in Early Modern English data from Ellegård (1953), though as discussed earlier, Early Modern English represents a very different kind of inversion. Van Herk (1998: 78) relates a suggestion from Stein (1988: 14) that non-inversion in YES/NO questions is promoted by the entire proposition being in question, as opposed to WH questions, where the WH element yields a query of narrower scope. When the entire proposition is in question, speakers may be more likely to resist any kind of word reordering.

---

$^{12}$ Rbrul does not produce an input probability (i.e., corrected mean) for models with any continuous variables, as input probabilities and logodds results are incompatible (Daniel Ezra Johnson, personal communication). The corrected mean here (and on Table 3) was generated by an alternate run with subject length “binned.”

$^{13}$ We use a continuous-variable, logodds treatment of subject length as it proved a better fit for the present data than a binned, factor-weight approach. For reference, if subject length is instead treated as binned, non-inversion percentages and total Ns are as follows: for subject length 1, 98% of 780 tokens; length 2, 88% of 25; length 3, 54% of 14; length 4, 71% of 7; and for lengths 5+, 27% of 15.
What is most striking about the auxiliaries is that *do*, modals, and copular *be* are the only types that are *not* knockouts for non-inversion, though token counts for aux *be* and *ain’t* are very low.

Polarity is non-significant, but note that contracted negatives -- like *ain’t, don’t, can’t, won’t, haven’t* -- are all categorically non-inverted in our Bajan sample. As knockouts, these auxiliary exclusions reduce what is left to examine as a separate negation effect.

Subject type does not appear because subject type and subject length are highly correlated in our data: pronouns (25) were almost always a single word (subject length = 1); simple NPs showed a mix of lengths, extended NPs (26) were somewhat longer on average, and sentential subjects were overwhelmingly five or more words long. If we remove subject length from the model, subject type emerges as significant, which suggests that these variables reflect the same underlying effect. Intuitively, this might represent an audience-design motivation, in terms of a hearer’s need to recognize an interrogative as early as possible. With inverted questions, word order signals up-front that the sentence is an interrogative. With non-inverted rising-intonation questions, however, as subject length increases, the hearer must wait deeper and deeper into the sentence before an end-rising intonation becomes noticeable:

(25) That is you favorite? (BA27b.1480)

(26) The part of the enjoyment of the cricket is you friends? (BA29.123)

5. COMPARING VARIETIES

In this section, we explore which of the systemic constraints on question non-inversion reported for Bajan apply to other Anglophone Caribbean varieties, and to other English varieties like AAVE, Appalachian, Samaná and ANSE.

5.1. Other Caribbean creoles

We begin with a pair of other Caribbean English-based creoles: Jamaican and Guyanese, each known to be closer to the basilectal end of the continuum than Barbadian (Alleyne 1980). Table 3 summarizes the results of our analyses, displayed alongside the analysis of Bajan from Table 2. Table 4 details application percentages and token counts corresponding to the factor weights in Table 3.

---

14 In the Bajan data the only longer pronoun subjects are the plural pronoun *you all* (length two) and compounds such as *you and he* (length three), for example in “But *you and he* did close?” (BA28a.1783).

15 For recordings and/or transcripts of the Jamaican data, we thank the following faculty members and students of UWI, Mona: Lisa Monique Barker, Annife Campbell, Tusheney Francis, Audene Henry, Trecel Messam, Velma Pollard, Jodian A. Scott, Andre Sherriah, Kathryn Shields-Broder, Kadian Walters, and Kedisha Williams. For transcription assistance with the Guyanese data, most of which was recorded between 1975 and 1982 by John Rickford and Daizal Samad, we thank these faculty members and students of the University of Guyana: Andrea Ally, Alim Hosein, and Daizal Samad.

16 Tables 3 and 4 are separated for space considerations, enabling side-by-side analysis on Table 3 of all seven varieties discussed here.
TABLE 3. Variable-rule factor-weight analyses of subject-aux non-inversion in questions in various Caribbean and North American Englishes

<table>
<thead>
<tr>
<th></th>
<th>Caribbean creole and semi-creole varieties</th>
<th>US varieties</th>
<th>African-American diaspora</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Bajan</td>
<td>Jamaican</td>
<td>Guyanese</td>
</tr>
<tr>
<td>Corrected mean</td>
<td>0.63</td>
<td>0.94</td>
<td>0.78</td>
</tr>
<tr>
<td>Overall rate</td>
<td>95%</td>
<td>96%</td>
<td>92%</td>
</tr>
<tr>
<td>Total N(^{17})</td>
<td>840</td>
<td>229</td>
<td>201</td>
</tr>
</tbody>
</table>

**POLARITY**

<table>
<thead>
<tr>
<th></th>
<th>Negative</th>
<th>Affirmative</th>
<th>Range</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>[.50]</td>
<td>[.50]</td>
<td>38</td>
</tr>
<tr>
<td>K/O</td>
<td>[.39]</td>
<td>[.61]</td>
<td>50</td>
</tr>
<tr>
<td></td>
<td>[.56]</td>
<td>[.44]</td>
<td>49</td>
</tr>
</tbody>
</table>

**QUESTION TYPE**

<table>
<thead>
<tr>
<th></th>
<th>YES-NO</th>
<th>WH-adverbial</th>
<th>WH-object</th>
<th>Causative</th>
<th>Range</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>.80</td>
<td>.65</td>
<td>.32</td>
<td>.23</td>
<td>57</td>
</tr>
<tr>
<td></td>
<td>[.70]</td>
<td>[.42]</td>
<td>[.38]</td>
<td>[.20]</td>
<td>70</td>
</tr>
<tr>
<td></td>
<td>.90</td>
<td>K/O</td>
<td>.30</td>
<td>K/O</td>
<td>24</td>
</tr>
<tr>
<td></td>
<td>.66</td>
<td>K/O</td>
<td>.42</td>
<td>K/O</td>
<td></td>
</tr>
<tr>
<td></td>
<td>.82</td>
<td>.49</td>
<td>.36</td>
<td>.36</td>
<td>51</td>
</tr>
<tr>
<td></td>
<td></td>
<td>.47</td>
<td>.36</td>
<td>.45</td>
<td>45</td>
</tr>
</tbody>
</table>

**SUBJECT LENGTH** (logodds)

<table>
<thead>
<tr>
<th></th>
<th>-0.90</th>
<th>-1.09</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>[-0.54]</td>
<td>[-0.61]</td>
</tr>
<tr>
<td></td>
<td>[-1.09]</td>
<td>[-0.55]</td>
</tr>
</tbody>
</table>

**AUXILIARY**

<table>
<thead>
<tr>
<th></th>
<th>Do</th>
<th>Modals</th>
<th>Copula be</th>
<th>Verbal aux be</th>
<th>Have</th>
<th>Ain’t</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>.77</td>
<td>.48</td>
<td>.25</td>
<td>K/O</td>
<td>N/A</td>
<td>K/O</td>
</tr>
<tr>
<td></td>
<td>.80</td>
<td>.60</td>
<td>.14</td>
<td>N/A</td>
<td>K/O</td>
<td>K/O</td>
</tr>
<tr>
<td></td>
<td>.93</td>
<td>.12</td>
<td>.37</td>
<td>K/O</td>
<td>K/O</td>
<td>K/O</td>
</tr>
<tr>
<td></td>
<td>.79</td>
<td>.22</td>
<td>.48</td>
<td>K/O</td>
<td>K/O</td>
<td>K/O</td>
</tr>
<tr>
<td></td>
<td>.73</td>
<td>.38</td>
<td>.38</td>
<td>K/O</td>
<td>K/O</td>
<td>.21</td>
</tr>
</tbody>
</table>

**Range**

|                      | 52     | 66     | 81       | 57       | 60    |

**Note.** See Table 2 notes for general form. Data to the left of the vertical line here are original to this paper, analyzed via Rbrul, with speaker as a random effect. Data on Samaná and ANSE to the right of this line are from Van Herk (2000: 186, Table 6.6), analyzed via Varbrul. (Non-significant factor weights were not available for the latter.)

\(^{17}\) After Dont Counts and knockouts.
Looking first at Jamaican, the overall rate of non-inversion is similar to Bajan. Our Jamaican sample was smaller than for Bajan, however, and is further limited by additional categorically non-inverted auxiliaries such as Jamaican *bin* and *a*. With smaller N, otherwise similar effects may not rise to significance. This may partially explain why neither question type nor subject length is retained as significant. We still find auxiliary type significant, with the effects of *do*, modals, and copula *be* in the same order.

The sample size for Guyanese is smaller still, but the constraint effects are actually very similar to those in Bajan. Polarity is non-significant, while subject length is significant. YES/NO questions strongly favor non-inversion in both varieties, and WH-objects and causatives are comparably non-favorable. *Do* is again the most favourable auxiliary, though the relative order of modals and copula *be* is reversed.

Note that while the constraint rankings are comparable across the three Caribbean varieties, the corrected means (or input probabilities) are more sharply differentiated (Bajan .63, Guyanese .78, Jamaican .94) than the corresponding non-inversion percentages (Bajan 95%, Guyanese 92%, Jamaican 96%). The corrected mean orderings also match earlier research and public perception that Jamaican is the most basilectal variety and Bajan the least.\(^{18}\)

\(^{18}\) The biggest gaps between the percentages and corrected means are in Bajan and Guyanese and, on further analysis, are attributable in part to controlling for individual speaker variability.

---

**TABLE 4. Application percentages and token counts (Ns) corresponding to factor weights for non-continuous variables in Table 3**

<table>
<thead>
<tr>
<th>POLARITY</th>
<th>Bajan</th>
<th>Jamaican</th>
<th>Guyanese</th>
<th>AAVE</th>
<th>Appalachian</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>%</td>
<td>N</td>
<td>%</td>
<td>N</td>
<td>%</td>
</tr>
<tr>
<td>Negative</td>
<td>99</td>
<td>83</td>
<td>K/O 55</td>
<td>97</td>
<td>37</td>
</tr>
<tr>
<td>Affirmative</td>
<td>95</td>
<td>757</td>
<td>96 228</td>
<td>90</td>
<td>164</td>
</tr>
<tr>
<td>QUESTION TYPE</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>YES-NO</td>
<td>99</td>
<td>560</td>
<td>97 167</td>
<td>99</td>
<td>141</td>
</tr>
<tr>
<td>wh-adverbial</td>
<td>95</td>
<td>81</td>
<td>94 17</td>
<td>100</td>
<td>71</td>
</tr>
<tr>
<td>wh-object</td>
<td>86</td>
<td>193</td>
<td>91 45</td>
<td>75</td>
<td>51</td>
</tr>
<tr>
<td>Causative</td>
<td>83</td>
<td>6</td>
<td>K/O 4</td>
<td>78</td>
<td>9</td>
</tr>
<tr>
<td>AUXILIARY</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Do</em></td>
<td>99</td>
<td>632</td>
<td>98 187</td>
<td>99</td>
<td>155</td>
</tr>
<tr>
<td>Modals</td>
<td>96</td>
<td>94</td>
<td>96 24</td>
<td>57</td>
<td>7</td>
</tr>
<tr>
<td>Copula <em>be</em></td>
<td>76</td>
<td>114</td>
<td>72 18</td>
<td>69</td>
<td>39</td>
</tr>
<tr>
<td>Verbal aux <em>be</em></td>
<td>100</td>
<td>5</td>
<td>N/A 0</td>
<td>100</td>
<td>1</td>
</tr>
<tr>
<td><em>Have</em></td>
<td>N/A</td>
<td>0</td>
<td>K/O 1</td>
<td>100</td>
<td>1</td>
</tr>
<tr>
<td><em>Ain’t</em></td>
<td>100</td>
<td>11</td>
<td>K/O 32</td>
<td>100</td>
<td>8</td>
</tr>
</tbody>
</table>

*Note:* Data presented here are for the varieties on Table 3 original to the present study. Percentages and Ns were not available for Samaná and ANSE studies from Van Herk (2000). Regarding token counts for continuous variables (i.e., subject length), see footnote 13.
5.2. U.S. varieties

The last sets of data to be introduced comprise a pair of vernacular U.S. dialects -- African American Vernacular English and Appalachian English,\(^{19}\) representing Black and White speakers, respectively -- enabling a structural comparison across a range of Englishes, both Caribbean and North American, beyond mainstream American English.

Our AAVE data display substantial non-inversion (35\%), though considerably less than in the three Caribbean varieties. More striking are the similarities across the varieties in terms of constraint effects. Polarity is non-significant in AAVE, as it is in Bajan and Guyanese. The constraint patterning for Question Type is comparable among all varieties insofar as YES/NO questions strongly favor non-inversion, and WH-objects and causatives disfavor it (although wh-adverbials pattern with WH-objects in AAVE rather than with YES/NO questions). Subject length is non-significant, but this is also true of Jamaican. *Do* is the auxiliary most favorable to non-inversion, with modals and copula *be* considerably less so, as in Bajan and Guyanese.

Finally, Appalachian English, a southern U.S. white variety, shows the lowest rate of non-inversion among the varieties explored in this study: 15\%, in line with the 14\% found in the Switchboard corpus of colloquial American English (Rickford 2006: 134).

Looking at systemic constraints for Appalachian, we observe that unlike any of the other varieties newly coded for this study, polarity is selected as significant, with negation favoring non-inversion. As with Jamaican and AAVE, subject length is not retained. We also find both auxiliary and question type significant for Appalachian as we did with Bajan, Guyanese, and AAVE, with similar alignment of effects within each factor group. The only exception in this respect is in the relative effects of verbal aux *be, have,* and *ain’t.*

5.3. Comparison to prior studies

As a final point of comparison, we revisit the Van Herk (2000) results, represented in the rightmost columns of Table 3.\(^{20}\) A key premise of Van Herk’s studies was that finding similarities between factors in the rise of *do* support (vs. main-verb/subject inversion) in Early Modern English (EME) and constraints on subject-aux inversion (vs. rising-intonation non-inversion) in Early African American English (EAAE) suggests a systemic relationship between earlier English and EAAE. Further, since Caribbean Englishes were believed to be categorically non-inverted, the constraint-based variation in EAAE was portrayed as evidence against a creole-origin account for modern AAVE.

Establishing the relationship between EME and EAAE, however, has always seemed to us a moot point; there is little question that Early Modern English has played a role to some degree in each of the English-based varieties under discussion here. The critical issue then is that as of

---

\(^{19}\) The AAVE data come from recordings made in East Palo Alto, California between 1986 and 2008 by Faye McNair-Knox and her daughter Rashida Knox (but also by John Rickford and his Stanford students). The Appalachian data are from West Virginia recordings made in the 1970s by Walt Wolfram, Donna Christian and their associates, and Beech Bottom, NC recordings made by Christine Mallinson, Becky Childs, Daniel Schreier and their associates in 2001. Thanks to them all, and to Clare Dannenberg and Tyler Kendall for facilitating access through the North Carolina Sociolinguistic Archive and Analysis project (NCSLAAP).

\(^{20}\) Note that these were computed by Varbrul and thus without control for individual speaker variation (random effect) as in the present study (computed by R/Rbrul).
these earlier studies, the same systemic relationship had not been shown between Anglophone Caribbean varieties and AAVE. Our results suggest exactly such a relationship.

The effect of question type that Van Herk found in Samaná and African Nova Scotian English (ANSE) we also find in Bajan. Indeed, the same constraint appears significant for Guyanese, AAVE, and Appalachian, as well, suggestive of an effect underlying many if not all English-based varieties.

Meanwhile, Van Herk also found that negation favored non-inversion in Samaná and ANSE. However, in our data, this polarity effect only retains significance in Appalachian. At the same time, each factor that came up as significant for AAVE showed a similarly significant effect in both Bajan and Guyanese, suggestive of a potentially systemic relationship between modern African American English and Anglophone Creole varieties.

6. CONCLUSIONS

In this, the first accountable, variationist study of non-inversion in questions in Caribbean English-based creoles, AAVE and Appalachian, we have several interesting findings to report.

Perhaps the most general is that despite differences in overall rates of non-inversion, the varieties show more similarities than differences. It is at first striking that Samaná, ANSE, and Appalachian show a comparable, variable effect of polarity (with negative favoring and affirmatives disfavoring), but Jamaican is somewhat similar in its categorical favoring negative effect. And when the effect of question type is considered, ANSE, Appalachian, AAVE, Guyanese, and Bajan are similar, with YES/NO questions most favorable to non-inversion, and WH-object questions disfavorable.

Subject length is significant in Bajan and Guyanese but not the other Caribbean and US varieties we examined, and it was apparently not considered in the analysis of Samaná and ANSE questions. But while auxiliary type was not found to be significant in the latter two, representing the African American diaspora (Van Herk 2000: 186), auxiliary type was significant in all of the other five varieties, with do most favorable to non-inversion in questions, and modals and copula be much less so. Indeed, in a subsequent analysis in which we combined data from all five varieties in a large, cross-variety regression (Melnick & Rickford in prep), we found no significant interactions (therefore no difference) between variety and the effects of negation, causativity, question type, and subject length. This suggests, as do comparable results found by Rickford (2013) for relativization, that question formation may not be a fruitful site for the investigation of the question of prior creolization of AAVE. Beyond this, both variables are rich sites for unraveling their constraints, and their connections to larger issues in the study of variation and change.

REFERENCES


---

21 Van Herk’s YES/NO analysis was limited to affirmative questions. However, when we make the same exclusions and re-run our statistics, we find the same essential result: the YES/NO question type continues to significantly favor non-inversion.


Johnson, D. E. (personal communication).


Melnick, R., & Rickford, J. R. (In prep). Improving analysis of varietal similarity in English question formation.


Rickford, J. R. (2013). Relativizer omission, the independence of linguistic and social constraints, and variationist Comparative Reconstruction. Paper presented to the Department of Linguistics, University of Edinburgh.


