Inversion in Samaná English Question Formation

Gerard Van Herk
University of Ottawa

Proponents of the creole-origins hypothesis observe similarities in question formation between African American Vernacular English (AAVE) and creoles. According to Dillard (1972), contemporary AAVE questions may resemble Standard English (with inversion) or creoles (with no inversion). If the noninverted forms are a creole legacy, they must have been more common in earlier AAVE. This hypothesis is tested by examining diaspora AAVE in Samana (DR). A variationist analysis of Samana English questions shows noninversion favoured by constraints resembling neither creoles (requiring categorical noninversion) nor contemporary English (requiring inversion). They do, however, match constraints on noninversion of lexical verbs (do-support) in Early Modern English. There, processing and semantic factors discourage inversion with negation, yes-no questions, causatives, and transititves. Samana modals and copulas also resist inversion, like the lexical verbs they once were. These results suggest the observed variability is a remnant of the system originally acquired along with the remainder of the English language.

1.0 Introduction

Similarities in question formation between creoles and African-American Vernacular English (AAVE) have been noted by proponents of the hypothesis that AAVE is descended from a prior creole. Creoles are claimed to exhibit no "difference in syntactic structure between questions and statements (Bickerton, 1981:70). According to Dillard (1972), contemporary AAVE displays both creole (non-inverted) and Standard English (inverted) forms, as in examples (1) and (2).

(1) Why she ain' over here? (Dillard 1972:63) - .

(2) Can he go? (Dillard 1972:63)

If non-inverted forms are a creole legacy, they must have been more common in earlier AAVE. This hypothesis can be tested by examining speech patterns in African-American linguistic enclaves, such as that found in the Samaná peninsula of the Dominican Republic. Descendants of freed American slaves have lived there since 1824. Isolated from contact with any other variety of English, their speech can be considered a precursor of contemporary AAVE.

And in fact, in this connection, DeBose (1996) used 100 questions culled from his corpus of Samaná English to claim that non-inversion in Samaná, as in (3a)-(3d), reflects its prior-creole status:

(3a) We ain' got no coffee fo share wit that man?

(3b) Why I didn't see you?

(3c) From where you is?

(3d) Where you was? (DeBose 1996)

2.0 Methods and issues

The present study attempted to replicate DeBose's findings in the University of Ottawa Corpus of Samaná English (Poplack & Sankoff 1987). However, this replication ran into a number of theoretical and methodological problems. These problems raise enough relevant issues that they will be dealt with here in some detail.

2.1 Defining the variable context

It was easy enough to extract all 803 questions from the University of Ottawa corpus. However, as the creolist diagnostic proposed is lack of subject-auxiliary inversion, it was first necessary to define the variable context -- that is, to restrict the data to questions that involved a subject and an (optionally deleted) auxiliary. A first interesting finding is that more than three-quarters of all questions in the corpus qualify as what Rickford and associates have termed "non-count," that is, they are outside the variable context.

These consist of repetition requests, as in (4); tag questions, as in (5), fixed expressions, as in (6), fragments, as in (7), echo questions, as in (8), and comprehension checks, as in (9).

(4) How? Hmm? Huh? Eh? (SA 021/523,530,537,555)

(5) They fought for Jiminez, eh? (SA 001/369)

(6a) Ain't true? (SA 005/283)

(6b) You don't believe? (SA 003/394)

(7) The hotel? (SA 001/939)

(8) Interviewer: Who was the last one, uh ...
Informant: Who was the last one I knew? (SA 0161697)

(9) Interviewer: Do you think that's bad?
Informant: If I think it's bad? (SA 016/355)

Question (9) is a shorter form of "(Are you asking me) if I think it's bad?" Note that another, similar question type could easily be mistaken for a non-inverted ("Creole") form:

(10) Interviewer: What did you learn with Horacio?
Informant: Why I didn't learn with Horacio? (SA 016/552)

As in (9), it is clear from the full context that (10) is a shorter form of "(Are you asking me) why I didn't learn with Horacio?"

Unlike DeBose, who reported only five unsuitable questions in his data, our peeling away of incomplete and anomalous questions reduces the corpus by 75%, (N=606 questions). This is the dauntingly large black portion of Figure 1.

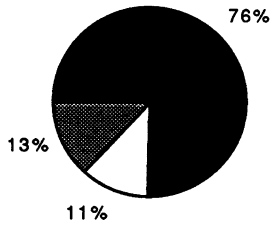

Figure 1. Non-count contexts in Samaná English

In addition, in more than half of the remaining 197 questions, (N=105) the auxiliary is deleted altogether, rendering them ambiguous as to underlying form. This is the shaded portion of Figure 1, as exemplified in (11).

(11) And where you-all come from? (SA 008/359)

Such questions cannot speak to the issue at hand, because it is impossible to determine whether the original auxiliary was inverted or not prior to deletion. Underlying example (11), for instance, could be:

(11a) Where ***did*** you-all come from? (StdE); or

(11b) Where you-all ***did*** come from? (creole).

Some auxiliary deletions are yet more difficult to spot. DeBose (1996) classes as uninverted "auxiliaries" sentences like (12a) - (12b).

(12a) You been to Oregon?

(12b) Where you gon' stay?

Such question forms are also found in the present corpus.

(13) What we going to do? (SA 003/427)

"Going to" here *is* an auxiliary, but not the one that would invert in this sentence in StdE -- witness the ungrammaticality of (13a).

(13a) *What going we to do?

The example in (13) is actually a deleted "are", which might have arisen from

(13b) What ***are*** we going to do? (StdE) or

(13c) What we ***are*** going to do? (creole).

The same is true of DeBose's examples, as evidenced by their ungrammaticality when inverted, as in (14) and (15).

(14) *Been you to Oregon?

(15) *Where gon' you stay?

These "uninvertable" auxiliaries actually make up 60% of DeBose's cited examples of non-inversion. This problem alone seriously damages any argument that Samaná questions pattern like creoles. Once the inapplicable questions exemplified in (12) are removed, fewer than half of DeBose's remaining cited questions are non-inverted.

After we exclude 105 questions with deleted auxiliaries, as in (13) above, from our variable context, we are left with 92 questions that contain an overt auxiliary. This is less than 12% of our original 803. In addition to "pure" verbal auxiliaries (16), this study includes "true copulas" (that is, *be* used other than as a verbal auxiliary) (17) -- in other words, we include any, and only, forms eligible for inversion.

(16) To who ***was*** they going? (SA 002/313)

(17) Where your riches ***is***? (SA 001/814)

2.2 Variation and diagnosticity

At first glance, the research question at issue here is remarkably simple. It is presumed that Standard English always inverts, and creoles never do. If Samaná English shows near-categorical inversion, it is descended from StdE; if non-inversion, from a creole (in the second case, according to some, inversion in contemporary AAVE would be the result of influence from, or "code-switching" to, StdE). In reality, neither the facts of English generally, nor the behaviour of Samaná English, support this simple agenda.

As Crowley & Rigby (1987) point out, yes/no questions, which do not invert in creoles, do not need to invert in English, either. Witness (18), from Ottawa English.

(18) It's near Billings Bridge? (OT 244/A005)

We are faced with a question of *diagnosticity* of the variable (as summarized in Table 1) -- a key, though relatively little documented, problem in the origins debate.

Table 1: Unequal diagnosticity: Acceptability of inversion in question types by language.

	Question Type	StdE	Creole
No inversion	yes/no	✔	✔
	WH	✗	✔
Inversion	yes/no	✔	✗
	WH	✔	✗

Both systems allow non-inverted yes-no questions, so such non-inversion cannot constitute evidence of either a prior creole or of English acquisition. A preponderance of inverted forms, though, *can* count as evidence against a creole -- the creole hypothesis would predict a complete absence of such forms, with exceptions explained as code-switching or decreolization.

The data, too, are unwilling to "behave" -- non-inversion is neither total nor totally absent. As Table 2 shows, of the 92 questions eligible for inversion, 54 (59%) showed no inversion. As both creoles and English permit non-inversion in yes-no questions, we would expect the greatest number of non-inverted forms to show up in this environment. This is, indeed, the case. Yes/no questions showed 69% non-inversion.In WH questions, where the creole diagnostic is non-inversion, we see only 39% non-inversion.

Table 2. Non-inversion in Samaná English by question type.

			Prediction	
Question Type	%	N	StdE	Creole
yes/no	**69**	61	✔	✔
WH	**39**	31	✗	✔
All	**59**	92		

Clearly, if Standard English and creoles both feature categorical question formation, neither can be the model for the system described in Table 2. Variationist studies of English-based creoles or English might reveal that question formation in those systems is actually more variable than the literature would have us believe. For now, however, we must look elsewhere for models that include variation.

The history of English provides just such a model. This is the adoption in Early Modern English of periphrastic *do*, thoroughly quantitatively described in Ellegard (1953), Kroch (1989, 1982/1989), and Stein (1988/1992). The increasing use of *do*-support in this period was a form of non-inversion. Questions moved from Old English lexical verb inversion, as in (19), to dummy auxiliary *do* inversion and a question with Subject-lexical Verb-Object order, as in (20).

(19) How great and greuous tribulations ***suffered*** the Holy Appostyls? (Kroch 1989)

object *lexical verb* *subject*

(20) Where ***doth*** the grene knyght ***holde*** hym? (Kroch 1989)

do *subject* *lexical verb* *object*

This Early Modern English non-inversion follows a clear hierarchy of conditioning factors, as summarized in Table 3: non-inversion is favoured by negatives, yes-no questions, causatives, transitives, NP subjects, and by adverbial over WH-object questions.

Table 3. Factors shown to condition non-inversion in Early Modern English

Variable Context	Factor Group	Favouring Factor	Source
All questions	Negation	negative	Ellegard
Affirmative yes/no & Adverbial questions	Question type	yes/no	Stein,Kroch* Ellegard
Affirmative WH questions	Causativity	causative	Stein, Ellegard
	Transitivity	transitive	Ellegard, Kroch
	Subject type	NP	Kroch
	Auxiliary type	modal, copula	après Kroch
Non-copula WH questions	Question type	adverbial	Stein,Kroch, Ellegard

Sources: Ellegard 1953, Kroch 1989, 1982/1989, Stein 1988, 1992. *Kroch for later EME only

In this paper, I intend to show that the extension of non-inversion from lexical verbs, as in (20), to auxiliaries, as in the Samaná data, is part of the same process of regularization of word order in questions.

Note that in the same way that we earlier peeled away non-application contexts from the full data set, here we have to split out different contexts in which different factors apply. Observe that the first column of Table 3 describes *four* different contexts in which different Early Modern English constraints apply. If these different constraints also account for the Samaná data, they will only apply to like subsets of the data (thus explaining the different Ns in each subsequent run).

Table 4: Factors contributing to non-inversion in Samaná English for each question type.

	All Questions, Overt Aux	Affirmative yes/no, Adverbial	Affirmative WH
Corrected mean			
Total N	**92**	**26**	**28**
Factor-group:			
NEGATION			
Negative	**.794**	[ ]	[ ]
Affirmative	.287	[ ]	[ ]
RANGE:	*51*		
QUESTION TYPE			
yes/no	[ ]	**.715**	[ ]
WH	[ ]	.255	[ ]
RANGE:		*46*	
TRANSITIVITY			
Transitive	[ ]	[ ]	**.729**
Intransitive	[ ]	[ ]	.178
RANGE:			*55*
SUBJECT			
Pronoun			**.914**
NP			.025
			89
RANGE:			
AUXILIARY TYPE			
Modal	[ ]	[ ]	**.895**
Copula	[ ]	[ ]	**.805**
have/be/do	[ ]	[ ]	.008
RANGE:			*88*

3.0 Results

The patient reader will recall that our earlier "peeling away" left us with 92 Samaná and 500 ANSE questions in which auxiliary and subject were present. These were coded for negation and for type of question, subject, object, and auxiliary. This covers all the factors proposed to condition non-inversion in Early Modern English in Table 3.

3.1 Negation favours non-inversion

As Table 3 shows, non-inversion in all Early Modern English questions is most strongly-favoured by negation, as in (21).

(21) He don't know the pastor? (SA 003/965)

The Samaná data show the same result. Table 4 shows us that negation favours non-inversion, with a probability of .79. Negation rendered all other factors statistically non-significant.

3.2 Yes-no questions favour non-inversion

Turning to Early Modern English affirmative questions, we find that non-inversion was favoured more by yes-no than WH adverbial questions (where, when, why, how). Once we remove *do* auxiliaries, which categorically invert in all the corpora, we find the same robust effect to hold for Samaná: Yes-no questions favour non-inversion with a probability of .715 (Table 4).

3.3 Causatives favour non-inversion

The strongest constraints in Early Modern English (and, we are seeing, in Samaná) are linked semantically by Stein (1988). He suggests negative and yes-no questions both focus on the truth of a proposition as a whole, and that this might be why such propositions are considered as a whole -- that is, without inverting. Stein further extends this suggestiion to the WH-questions that he says most closely resemble yes-no questions -- the causatives. In Early Modern English, causative questions -- the "why?" questions, as in (22) -- favoured non-inversion more than other WH questions.

(22) Why you-all ***are*** English-speaking people and you speak Spanish with the children? (SA 003/175)

Example (22) is our only Samaná causative, and matches the Early Modern English prediction. This one example is not conclusiv, but this match is also confirmed by preliminary results from another diaspora variety, African Nova Scotian English (ANSE). Recall that Stein's Early Modern English work links negation and causativity semantically. In ANSE, negation and causativity conspire to create a highly-favourable environment for non-inversion. Table 5 shows that negative

causatives, as in (23) and (24) below, favour non-inversion very strongly, with a probability of .96.

Table 5: Factors contributing to non-inversion in WH-questions in African Nova Scotian English.		
Corrected mean		
Total N	**363**	
Factor-group:		
NEGATION/CAUSATIVITY		
	Causative	Non-Causative
Negative	**.961**	no data
Affirmative	**.902**	.386
RANGE	*57*	
SUBJECT		
First person	**.894**	
Third person	**.698**	
NP	**.630**	
Second Person	.194	
RANGE	*70*	

Factor-groups not selected: Transitivity, auxiliary type.

It is probably no accident, then, that most published examples of contemporary AAVE non-inversion are precisely this question type -- again, as in examples (23) and (24).

(23a) ***Why*** you ***don 't*** like him?

(23b) ***Why*** I ***don 't*** need no grease? (Labov et al 1968:294)

(24) ***Why*** she ***ain'*** over here? (Dillard 1972)

Note also the causative "How come?" form, as in example (25). This is the only question type in vernacular (white and black) English where non-inversion is categorical.

(25) How come the children don't play like that? (ANSE 005/263)

3.4 Transitivity favours non-inversion

Early Modern English non-inversion is further favoured by transitivity and a heavy NP subject, which are linked. To paraphrase Kroch (1989), an object, if present, dislikes a subject of its own weight or heavier intervening between it and the verb. In Samaná English, we find that transitivity favours non-inversion, as predicted, but only in WH questions. Again, see Table 4. However, we find NP subjects disfavour non-inversion, contrary to Early Modern English. This might be due to a parallel-processing effect in the Samaná data, documented here for the first time, which appears not to exist in the Early Modern English data. Here, object and

subject tend to be of the same weight, so subject weight is less of a factor -- subjects are never "heavier" than their object counterparts.

3.5 Lexically-heavy auxiliaries favour non-inversion

For auxiliary-type effects, we have of course no direct comparative base in Early Modern English -- the entire point of *do*-periphrasis is that it involves *do* and a lexical verb. We can, however, extrapolate from theoretical Early Modern English work. Recall that the point of *do* -periphrasis was to maintain Subject - *lexical* Verb - Object order. From this, we would predict that the semantically and lexically heavier modals and copulas should favour non-inversion -- exactly as Table 4 indicates. Modals (at .895) and copulas (.805) strongly favour non-inversion. This may also indicate that modals retain lexical weight in AAVE, and so are farther behind on the grammaticization continuum. This lexical weight is also evidenced by their appearance in AAVE as second verbs, as in example (26):

(26) He ***must could*** go.

The grammaticization hypothesis icould be further confirmed if we fnd that the lexical-weight effect is weaker or non-existent in varieties of AAVE with less non-inversion than Samaná.

3.6 Do adverbials favour non-inversion?

The only apparent disagreement between Early Modern English and Samaná is that Early Modern English is said to have more non-inversion in adverbial questions, as in (27), than in WH-object questions, as in (28).

(27) But ***when*** you***'ll*** be coming back again? (SA 004/701)

(28) ***What does*** she say? (SA 008/177)

Our Samaná numbers are too small for a multivariate analysis, but the raw numbers run counter to EME predictions. This may be due to a shortage of questions, or the problem may lie with the literature on Early Modern English. Ellegard (1953:205, cited in Kroch 1989) combines adverbials with yes-no questions and causatives, which we have already seen favour non-inversion. Stein (1988) appears to combine WH-object *what* with WH-subject questions, which categorically invert (see example 29).

(29) ***Who is*** there, Virgie? (SA 013/825)

In Stein's raw data, *whom*, the only unambiguously WH-object form, actually shows non-inversion in the same high range as causatives... as in the few available EBE questions.

3.7 Summary of results

We have seen a clear hierarchy of non-inversion in this Samaná English data set, with the clearest parallels to Early Modern English supported by robust results in the largest variable contexts. Non-inversion is most likely with negative questions. In affirmative questions, it is more likely in yes-no than in WH questions. In WH-questions, it is more likely with causatives. In the remaining WH-questions, an easily-processed Subject - lexical Verb - Object order is maintained through non-inversion with transitives, and with modals and copulas. These parallels to Early Modern English cannot be coincidental.

4.0 Discussion

The results of the present study are suggestive with respect to both methodological issues and the creole-origins debate.

4.1 Methods revisited

From a methodology standpoint, we must more carefully consider the diagnosticity of the variable used to claim possible creole origins for AAVE. In the case at hand, non-inversion in yes-no questions, despite its inclusion in earlier work, is clearly not a uniquely creole diagnostic, as non-inversion is possible in both Early Modern and contemporary English.

Further, we need a good deal of comparative quantitative work on AAVE question formation before we can say anything about convergence/divergence. We need similar work on a decreolizing language we can all agree springs from a creole, such as a Caribbean upper mesolect. We need to see what constrains variability in those systems, along with explanations of what grammar the constraints derive from, as we were able to provide here thanks to Kroch, Ellegard, and Stein.

We also need to know how widespread non-inversion actually is in varieties of English, to see if the forms in all systems are conditioned by the same factors. My current research leads in this direction. Preliminary results from a small white Ottawa vernacular
English sample suggest some similar conditioning. As in Samaná and ANSE, negation strongly conditions non-inversion, as in (30).

(30) So you ***didn't*** come home stum- stumbling drunk? (OT 244/A084)

4.2 Inversion and the creole-origins debate

A first important finding to emerge from this study is that once the variable context is accurately circumscribed, the inverted forms are seen to be far too

common for each occurrence to be explained away as "code switching" or tape noise. In yes-no questions, where inversion is optional, even in English, inverted questions form a strong minority; in WH questions, where non-inversion is a creole diagnostic, inversions are in the majority. Furthermore, replication with another diaspora variety of early AAVE, African Nova Scotian English, finds non-inverted questions to represent only a tiny minority (3% of eligible WH-questions).

Secondly, Samaná's non-inverted forms are clearly rule-governed. They must be the remnants of some system. This paper has shown that the variable constraints conditioning that system are the same ones that constrained subject-verb inversion in Early Modern English, and that those constraints have been extended to auxiliaries. Here, the standard creole argument will not do -- it requires categorical non-inversion, and even if it admitted variation, certainly makes no provision for a negation or causativity distinction. In contrast, the constraint hierarchy described here has been found to operate not only in Samaná, but also independently in another diaspora variety, ANSE, and this in a much larger corpus of questions (N=363). But the most telling point is that the same hierarchy is found even in DeBose (1996), which advocates a prior-creole origin. In that paper, all four cited examples with non-inverted eligible auxiliaries feature either negation or copulas, as we can see by repeating examples (3a) -(3d).

(3a) We ain' got no coffee fo share wit that man?

(3b) Why I didn't see you?

(3c) From where you is?

(3d) Where you was? (DeBose 1996)

We can only conclude that, contrary to DeBose's claims elsewhere, "his" Samaná English and this Samaná English are actually the same.

What of the possibility that the similar constraint hierarchy in Early Modern English and Samaná reflect some deep, universal tendencies, which operate on Samaná "creole" speakers even as they attempt to acquire an acrolect? Here, what is known of (non) inversion of copulas in acquisition may offer something new to the discussion. SLA work by Pienemann et al. (1988) across a wide range of languages suggests that copulas are the only area of question inversion where second-language learners differ from child language learners. Second language learners are said to invert copulas before other auxiliaries. If the variation found in Samaná English were actually "acrolectal code-switching," i.e. use of a system acquired after acquisition of a creole language, copulas should show more inversion

than other auxiliaries. Instead, as we have seen in Table 4, they invert less. This suggests the observed variation is actually part of a system acquired from birth.

We are still learning about that particular system of question formation, but for now, the parallels between EBE and amply-documented Early Modern English findings cannot be accidental.

References:

Bickerton, D. (1981). *Roots of Language*. Ann Arbor, MI: Karoma.

Crowley, T., & B. Rigby. (1987). Question formation. In Shopen, T., *Languages and their Status*.

DeBose, C.E. (1996). Question formation in Samaná English. Paper presented at NWAVE 25, Las Vegas, NV, Oct 17-20, 1996.

Dillard, J.L. *Black English: Its History and Usage in the United States*. New York: Random House.

Ellegard, A. (1953). *The auxiliary* do*: The establishment and regulation of its use in English* (Gothenburg Studies in English). Stockholm: Almqvist and Wiksell.

Kroch, A.S. (1982/1989). Function and grammar in the history of English: Periphrastic *do*. In Fasold, R.W. & D. Schiffrin (Eds.), *Language Change and Variation.*, pp. 133-172. Amsterdam: Benjamins.

Kroch, A.S. (1989). Reflexes of grammar in patterns of language change. *Language Variation and Change, 1,* 199-244.

Labov, W., P. Cohen, C. Robins, & J. Lewis. (1968). *A Study of the Non-Standard English of Negro and Puerto Rican Speakers in New York City. Vol. I: Phonological and Grammatical Analysis*. Philiadelphia: U.S. Regional Survey.

Pienemann, M., M. Johnston, & G. Brindley.(1988). Constructing an acquisition-based procedure for second language assessment. *Studies in Second Language Acquisition, 10,* 217-243.

Poplack, S., & D. Sankoff. The Philadelphia story in the Spanish Caribbean. *American Speech, 62(4),* 291-314.

Stein, D. (1988). Semantic similarity between categories as a vehicle of linguistic change. *Diachronica 5(1-2),* 1-17.

--- (1992). Semantic aspects of syntactic change. In Gerritsen, M., & D. Stein (Eds.), *Internal and External Factors in Syntactic Change,* pp. 155-166. Berlin: Mouton de Gruyter.