

AAVE/Creole copula absence

A critique of the imperfect learning hypothesis*

Devyani Sharma and John R. Rickford

Queen Mary University of London / Stanford University

This study confirms the robustness of the finding in the literature on African American Vernacular English [AAVE] and creole English (especially in the Caribbean) that omission of copular and auxiliary *be* varies systematically according to predicate type. Verbal predicates are associated with the highest rates of copula absence and following NPs with the lowest rates; following adjectives or locatives show intermediate rates (see Rickford 1998:190). Although this pattern is highly consistent, convincing explanations for it remain elusive. A recurrent suggestion (McWhorter 2000; Winford 1998, 2004; Wolfram 2000:54) is that the AAVE and creole English pattern is inherited independently from general processes of imperfect second language learning (simplification, generalization) that operated as the African ancestors of today's speakers acquired English. In this paper, we pursue this possibility, but discover that the grammatical conditioning of copula absence in AAVE and creole varieties is distinct from the patterns found in second language learning data. We examine four sets of data on English acquired as a second language (Indian English, South African Indian English, Singaporean English, Spanish English) and show, using two statistical measures, that conditioning of copula absence in the second language data does not resemble the AAVE and creole pattern. (One possible exception is the high rates of omitted *be* with verbal predicates, for which we explore possible explanations.) We show further that typological diversity in copula systems also militates against a universal markedness-based pattern. The findings reduce the possibility that the overall AAVE/creole pattern derives from a general tendency in second language acquisition and increase the possibility that the pattern reflects a shared

* This paper represents the culmination of several years of joint work on this topic, from LSA 2002 to UKLVC 2007. We would like to thank audiences at these and other forums (e.g. the University of Canterbury, New Zealand) for their feedback, as well as John Singler, Don Winford, Misha Becker, Anthea Fraser Gupta, Bill Haddican, Kate Kearns, Erez Levon, and two anonymous reviewers for many helpful suggestions. We also thank Ewart Thomas and Sara Binzer Hobolt for generous discussions of statistical measures. We alone bear responsibility for the final version.

substrate influence from West African languages or other historical contact factors.

Keywords: copula absence, quantitative measures, African-American Vernacular English, (Caribbean) English creoles, creole origins controversy, typology, second language acquisition

Introduction

A solid finding in the study both of African American Vernacular English (AAVE) and of English-based creoles, particularly in the Caribbean, is the quantitative patterning of English copula and auxiliary absence according to predicate type. Predication without a copula, or zero copula, is most frequent with a *gon(na)* future or a progressive (*she Ø gon tell him; she Ø walking*), least frequent before a noun phrase (*he Ø a man*), and of intermediate frequency before a locative or adjective (*he Ø in the car; she Ø happy*).¹ The general pattern for AAVE is given in (1), with predicate types listed according to increasing rates of copula absence:

- (1) NP < Loc < Adj < V-*ing* < *gonna*

Some scholars have contended that this predicate-based copula absence pattern in AAVE reflects prior creolization (Baugh 1980, 1983; Holm 1976, 1984; Rickford 1998), and that the creole pattern in turn represents potential influence from West African languages (Holm, *ibid*). An alternative proposal is that the AAVE pattern corresponds to general patterns of imperfect second language acquisition (SLA) or ‘creative second language acquisition’ involving universal strategies of simplification alongside additional creolization influences (McWhorter 2000:419–20; Winford 1998:111, 2004:3; Wolfram 2000:54).² In this paper, we evaluate the ‘imperfect learning’ hypothesis by comparing AAVE/creole copula absence to second language (L2) learning data. We discover that the precise conditioning of copula

1. The use of *be* with verbal (V-*ing*, *gonna*) and non-verbal (NP, Adj, Loc) predicates is generally referred to as auxiliary and copula *be*, respectively; in the present discussion we generally subsume both types under the term ‘copula’, but distinguish them where necessary.

2. The treatment of ‘simplification’ and ‘creolization’ as distinct is problematic, given that creolization itself, especially its antecedent pidginization, is commonly thought to itself involve simplification (Hymes 1971; Whinnom 1971). It may be more accurate to say that the difference is one of degree rather than kind, with pidgin/creole formation occurring in circumstances where the L2 learners are more distant from the L2 norm, and where the ecological and etiological/emotional conditions make it less likely that L2 learning will proceed successfully and more likely that L2 learners will have to depend on L1 transfer, simplification, and so on.

absence according to predicate type shared by AAVE and creole varieties is in fact distinct from the patterns found in L2 acquisition data, in particular data from Indian English, South African Indian English, Singaporean English, and Spanish English. These findings challenge the assumption that there is a general order of English copula acquisition by predicate type and that this order could be the source of the robust creole and AAVE patterns.

In the first section we present two hypotheses relating to the source of the AAVE copula omission pattern. The second section discusses the methodology used, and the third and fourth sections review AAVE, creole, and L2 data on copula use. The discussion and conclusion sections evaluate the two hypotheses in relation to the range of empirical data, making further reference to cross-linguistic typology and first language (L1) acquisition.

The imperfect learning argument

Winford (1998:111), elaborating on his earlier idea (referred to in Rickford 1998:180), proposes a dual hypothesis to account for the AAVE copula pattern, suggesting it may be ‘best explained as the result of imperfect second language learning, with transfer from creolized or restructured varieties playing a significant role.’ In support of the first part of this proposal, Winford notes that ‘some studies of Spanish-speaking adults acquiring English as a second language (Butterworth & Hatch 1978; Shapira 1978) also show that variable absence of copula and auxiliary *be* is typical, and is due primarily to simplification’ (p. 114). He concludes:

It would be interesting to do a full quantitative study of the interlanguage of groups of such learners using methodologies similar to those employed in the analysis of copula variability in AAVE and other situations. Such studies may provide further support for the view that both strategies of simplification and leveling in second language acquisition, and creole substratum influence under shift, contributed to the characteristic pattern of copula use in AAVE.

The present paper proposes to conduct precisely such a comparison between L2 systems and AAVE. To begin with, it is important to note that Winford’s observations combine two distinct aspects of second language acquisition (SLA) — universal strategies of simplification and specific language transfer — and in order to be assessed empirically they must be distinguished clearly.

HYPOTHESIS I: Certain tendencies are universally characteristic of second language learning, such that all adult learners of English will follow a shared learning trajectory, regardless of their first language. Such a trajectory may arise either from relationships among structures in the particular language being acquired

(e.g. English) or from universal markedness or simplification tendencies. The simplification view (Ferguson 1971; Mühlhäusler 1977; Pit Corder and Roulet 1977) encompasses 'regularization of irregularities, loss of redundancy (such as grammatical gender), and an increase in transparency' (Trudgill 2004:307). Other manifestations of simplification include phonetic erosion, use of less marked forms, fewer rules, fewer semantic distinctions, and less inflectional morphology (McWhorter 2001:154), as well as a smaller lexicon, a reduced phonemic inventory, fewer grammatical categories, less bound morphemes, and fewer exceptions to rules (Veenstra 2007).

HYPOTHESIS II: Transfer from the first language of the speaker plays a significant role in restructuring the second language variety. Unlike the first hypothesis, this explanation relies on specific effects deriving from the grammar(s) of the first language(s) of the speaker in question and hence predicts significantly different outcomes in different L2 learning situations depending on the grammars of the first languages involved. Hypothesis II has been discussed widely in creole studies (Lefebvre 1998; Migge and Smith 2007; Siegel 2003; Veenstra 2007; Winford 2003) and we do not review it in detail here. An example is Siegel's (2003) review, in which he observes that similarities between creoles and SLA are most striking when the substrate and first languages share traits. He cites several studies (Mather 2000 and Véronique 1994, among others) that have found parallel features in the interlanguage systems of L2 French learners and in French creoles specifically when the L1 and substrate languages share those features.³ This view suggests that when features are shared between creoles and SLA, parallel substrate effects rather than universal pressures may be the cause.

Whereas Hypothesis II has been explored in detail in creole and dialect studies, there has been considerable recent interest in Hypothesis I, and this hypothesis forms the core of the present comparative study.

A number of studies examining AAVE have invoked general traits of imperfect learning (Hypothesis I) as a possible source of copula absence. McWhorter (2000:419–21), after noting that copula absence in AAVE and African American diaspora varieties like Samaná English cannot reasonably be attributed to 'English inheritance' or 'historical accident,' suggests that it could be attributed to the kind

3. Mather (2000, cited in Siegel 2003:192) gives the example of postposed determiners, which exist in substrate languages such as Ewe and thus arise frequently in French creoles, but do not occur in the L2 French of Arabic speakers, as this is not a feature of the L1 Arabic system. Siegel also highlights further limitations of generalized SLA-creole comparisons, noting that none of the characteristic TMA features of creoles are found with any consistency in the interlanguage systems of L2 learners.

of ellipsis or overgeneralization characteristic of second language acquisition more generally. He observes: 'Black English was born as a koiné of white nonstandard English modified slightly by cross-linguistically typical results of widespread second language acquisition. This is also the position of Winford (1997, 1998)' (p. 424). He contrasts this koineization and general SLA basis of AAVE against the additional characteristics of substrate transfer and internal innovation found in creoles (p. 420).

Wolfram (2000:54) similarly suggests that copula absence in AAVE may reflect imperfect language learning rather than creolization:

[C]opula absence among African Americans is probably traceable to a contact-based case of fossilized, 'imperfect language learning' found among groups in African diaspora. This feature was then transmitted throughout the African-American population as a distinctive dialect feature. Support for copula absence may have come from creole speakers, but this does not necessarily mean it has an exclusive creole origin since copula absence can arise in a variety of language-contact and language-learning situations.

Noting that there is 'no reasonable evidence of copula absence in the British donor dialects' (p. 54), Wolfram, like McWhorter, appeals to the widespread occurrence of copula omission in second language learning as an explanation for the AAVE system.

Ferguson (1971) also argued for a 'universal' explanation of copula absence, suggesting that it is the 'unmarked' variant of the copula and that 'the simpler of two comparable features is likely to be the more widespread among languages of the world' and would also tend 'to be used for both [variants] in simpler speech' (p. 145). He argued that these universal processes of simplification 'would help to explain some of the otherwise surprising similarities among distant creoles by setting the starting point in a universal simplification process' (p. 148).

Several further studies (Andersen 1983; Siegel 2003; Veenstra 2003, 2007) have also explored generalized parallels between prototypical pidgin features and characteristic features of reduction, regularization, and transparency shared across adult second language speech, amid continuing debates over the definition and scope of the notion of 'simplification' (see McWhorter 2001 for a summary). Winford (2004:4) notes that Klein and Perdue's (1997) 'Basic Variety' features of adult learner speech (and also Felix's (1978) order of learning for copular and auxiliary *be*) 'shares many characteristics with 'prototypical' pidgins' (p. 4). He concludes that 'all cases of creole formation are in varying ways similar to cases of second language acquisition in 'natural' settings' (p. 24), although he does cite the perseverance of L1 features (again invoking both Hypothesis I and II as sources of creole features).

All of these studies appeal to Hypothesis I (generalized traits of SLA or imperfect learning) to account for AAVE copula absence. However, none address the specific order in (1) — neither the question of whether this order is indeed repeated across L2 data, nor, if it is, why this would be. True support for Hypothesis I should ideally take the form of: (i) parallel claims within SLA theory, (ii) quantitative evidence of the AAVE ordering in L2 data, and (iii) an intuitive semantic or syntactic basis for such an order to persist cross-linguistically. Let us briefly consider each of these more detailed expectations:

(i) SLA literature: If universal SLA tendencies are being invoked to account for AAVE or creole systems, there should be evidence within the SLA literature that such general tendencies do in fact exist in SLA. We have found no specific proposal that a ‘universal’ SLA order of acquisition exists across all copular and auxiliary uses of *be*. The most detailed research on implicational ordering of copula acquisition has been in studies of L2 learners of German. Nicholas (1984:303–6) summarizes this research, which includes the three proposed orderings in Table 1.

Table 1. Proposals for universal SLA order of acquisition for verbal morphology. (Based on Nicholas 1984:303–6.)

Study	Order of acquisition			
a. Dittmar (1980)	simple verbs	< modals	< copula	< auxiliary
b. Felix (1978: 50, 113)	equative sentences without copula	< equative sentences with copula	< auxiliaries	< full verbs < complex verbal groups
c. Pienemann (1981: 37)	full verbs	< copula	< auxiliaries and modals	

These proposals make no claim about distinctions in copula acquisition with different non-verbal predicates (NP, Adj, Loc). They do, however, agree that the copular use of *be* is acquired before the auxiliary use, a binary sequence of acquisition that also forms part of Krashen’s (1982) ‘natural order hypothesis’ for English and one that has been observed by Stauble (1984:324) and Hawkins (2001).⁴ We return to an assessment of the differential acquisition of copular and auxiliary constructions in the discussion section. Aside from this binary ordering, however, our research has found no claim in the SLA literature in favor of a universal SLA sequence that resembles (1). This is confirmed by the L2 datasets we examine in the present study.

4. Stauble (1984:351) does note that Spanish learners of English are consistently closer to the target language copula usage as compared to Japanese learners of English, suggesting that the first language still does play a role in learning this part of the grammatical system.

(ii) Quantitative evidence: Many of the comparisons to SLA reviewed above remain at the coarse level of merely identifying the fact of copula absence, rather than contrasting the fine-grained grammatical conditioning of copula absence in different systems. The latter approach is more precise and falsifiable, and has other theoretical advantages.⁵ We refine our comparison in the present study to copula absence patterning by predicate type (the most commonly noted constraint) rather than the mere occurrence of copula absence alone. This finer analysis shows that while AAVE, creoles, and some L2 varieties of English are alike at the coarse level of displaying copula absence (differing from Standard English and most other L1 varieties of English in this respect), there is no single ordering of copula use by predicate type across all such datasets. Under a finer analysis, only AAVE and the English-based creoles appear to share a similar system, ultimately lending support to Hypothesis II and the possibility of shared origins for AAVE and creole varieties.

(iii) Explanations: In our discussion we also explore potential semantic and syntactic motivations for the existence of a universal order, as predicted by Hypothesis I. Although systematic structural properties are associated with certain copular constructions in different languages, we find no clear motivation for the existence of a single universal ordering across all contexts.

Data and methods

The data examined for AAVE and English-based creoles are compiled from existing studies, summarized in Rickford (1998).⁶ These datasets are compared to four

5. See Poplack (2000) and Rickford (2006) for the method and value of using quantitative evidence of constraint patterning for comparing language varieties and making inferences about their historical connections.

6. A note on the comparability of subject, tense, and data type in the present datasets: Some AAVE studies have restricted their focus to 3rd person singular subjects while others have not, and Rickford et al. (1991:112) conclude, after a detailed analysis of variable constraints on each, that 'is and are behave similarly enough to be treated together, as they were in Poplack and Sankoff (1987), making the data pool larger and more robust and ensuring that their similarities in constraint effects need be stated only once.' Similarly, some studies examined here exclude past tense tokens while others do not. Finally, the type of data we compare here is raw percentages and straight deletion (not VARBRUL weights or Labov deletion), even though studies have shown VARBRUL weights to be a finer measure of predicate effects, as they take into account the relative effect of other factors. Our reliance on raw data avoids inappropriate statistical operations and comparisons over dissimilar data types. Table (i) lists source details and tense/subject information for the AAVE and creole datasets used here; wherever second language data are presented later in the paper, we provide footnotes containing similar information.

different sets of L2 English data: Indian English, South African Indian English, Singaporean English, and Spanish English. The Indian English data come from sociolinguistic interviews conducted by Devyani Sharma with Indian speakers of English, all of whom have Indo-Aryan L1s with similar copula systems. The other L2 datasets, and the additional L1 learning data considered briefly later in the paper, come from previously published studies, all of which are based on conversational speech recordings.

Although multivariate regression (especially as part of the VARBRUL program) has been used repeatedly in the study of copula absence, there has been no attempt so far to introduce statistical measures of conformity to assess the extent to which the constraint orderings in two or more datasets are 'similar' or 'different' to each other. The use of such measures allows us to move from examining broad competition among factors (or reaching premature conclusions based on our ideological positions) towards precise measurements of the consistency of an ordered pattern. The two measures we use for the data compared in the present study are a nonparametric test (Spearman's rho) and a reliability test (Cronbach's alpha).⁷

Table (i). Tense, subject type, and source for AAVE and creole studies in Tables 1 and 2.

Study	Tense	Copula forms (subject types)
East Palo Alto (Rickford et al. 1991:117, Table 6)	present	<i>is, are</i>
Texas older folk-speakers and kids (Bailey & Maynor 1987:457)	present	<i>is, are</i>
New York City (Baugh 1979:180)	present	<i>is</i>
Detroit Working Class (Wolfram 1969:172, Fig. 49)	present	<i>is, are</i>
New York City (Labov 1969:732)	present	<i>is</i>
Jamaica 1960 (Rickford 1996:363, Table 3)	present and past	<i>am, is, are</i>
Jamaica 1991 (Rickford 1999:151, Table 4)	present and past	<i>am, is, are</i>
Barbados (Rickford & Blake 1990:268)	present	<i>am, is, are</i>
Barbados (Rickford 1992:191)	present and past	<i>am, is, are was, were</i>
Trinidad (Winford 1992:34, Table 5)	present	<i>am, is, are</i>
Hawaii (Day 1973:111, Table 9)	present	<i>is</i>

7. As with much linguistic data, the current datasets do not always satisfy certain underlying assumptions for parametric tests: (i) their size is not always as robust as the large datasets used in quantitative studies in other social sciences (at least when group averages rather than individual tokens are being compared), and (ii) we cannot assume a normal distribution for the data across the population in question, or even a continuous variable along the x-axis (unlike, for instance, a distribution of exam results or income across a population). Non-parametric tests make fewer assumptions about the data and distribution and are thus more appropriate, if somewhat less powerful, for the present data (Agresti and Finlay 1997:232).

The Spearman rank-order correlation coefficient (Spearman's rho, R_s) is a common nonparametric measure of the strength and direction of correspondence between two sets of ranked data (Siegel and Castellan 1988) and has been used recently in the study of L2 English variation (e.g. Rasinger 2005). By examining the degree to which the rank-order of values in a distribution directly or inversely mirrors the rank-order of values in a second distribution, this measure allows us to compare the extent to which two distributions resemble one another. Note that the Spearman rank-order correlation only attends to the ordinal scale of the data; it discards any further details of the actual values (e.g. a linear progression — 1, 2, 3 — and an exponential progression — 3, 9, 27 — have the same rank-order: they both differ in the same way from the orderings 2, 1, 3 or 9, 3, 27). Its use, therefore, is in evaluating the degree to which a set of data conforms to the ordered distribution of another set of data or to a hypothetical order. All Spearman coefficients reported in this paper compare an individual data range to the ordering in (1), in order to examine the degree to which that dataset conforms to the hypothesized order. The Spearman coefficient ranges between -1 (perfect negative correlation) and 1 (perfect positive correlation).⁸

This measure is attractive for natural language data from speech communities, as we may want to identify an underlying construct without predicting a perfect mirroring of actual rates across speakers. However, we may wish not only to understand how closely data conform to a particular order, but also to assess the relative closeness of actual values among a set of speakers of a variety. For instance, two sets of data may have the same ordering, but one slope may be shallow and one steep. To measure this finer conformity within datasets we turn to Cronbach's alpha (Cronbach 1951, Miller 1995), a statistic used to judge the reliability of tests by quantifying the extent to which they provide the same results on repeated trials.⁹ (For example, on a test measuring an aspect of personality, individual X may

8. A Spearman coefficient value of -1 indicates a perfect negative correlation; if the value falls between -1 and -0.5 , there is a strong negative correlation; if it falls between -0.5 and 0 , there is a weak negative correlation; if it is 0 , there is no correlation (the null hypothesis is confirmed); if it falls between 0 and 0.5 , there is a weak positive correlation; if it falls between 0.5 and 1 , there is a strong positive correlation; and if it is 1 , there is a perfect positive correlation between the two sets of data. Significance at the $p < .05$ and $p < .01$ levels is determined by the statistical package (SPSS) based on the dataset in question; for the present data a value of $.9$ corresponded to $p < .05$ and a value of 1 corresponded to $p < .01$.

9. Cronbach's alpha can be defined as $\alpha = \frac{n}{n-1} \left(1 - \frac{1}{\sigma_T^2} \sum_{i=1}^n \sigma_i^2 \right)$, where n is the number of trials (here, the number of speaker groups), σ_i^2 is the variance of the i th trial scores across individuals (here, the variance of the i th speaker's rates of omission across copula contexts), and

score n units higher than individual Y . If the test is reliable, X should again score n units higher than Y on a second trial.) In our present use, we adapt this measure to assess how similar rates of omission are for each grammatical context across different groups of speakers of a language variety. In other words, the alpha coefficient allows us to evaluate precisely how similar linear plots are to each other and therefore how closely a group conforms to a single pattern. Cronbach's alpha is not strictly a significance test; it provides a coefficient between 0 and 1 that is a relative measure of consistency or homogeneity in the dataset. A coefficient of .80 or higher is commonly treated as a threshold above which the set of items correspond very closely to the same pattern.

In the discussion that follows, the Spearman coefficient will measure the conformity of a given dataset to the AAVE ranking in (1), while the alpha coefficient for each set of speakers will measure the internal consistency among copula omission patterns for that variety. Next, we present the different sets of data along with these measures of conformity; this is followed by a discussion of similarities and differences across the datasets and possible explanations for both.

AAVE and creole English data

Let us first establish the robustness of the AAVE pattern of zero copula by predicate type. From Labov (1969) to present day research, the overall pattern given earlier in (1) has been substantially confirmed in numerous studies. Figure 1, based on data summarized in Rickford (1998:190, Table 6.16) shows copula absence by following grammatical environment in eight groups of AAVE speakers, in five geographical locales: East Palo Alto (California), Texas, Los Angeles, Detroit, and New York City (see fn. 7 for sources).¹⁰

Minor deviations from the overall pattern are observable at certain points in Figure 1: The New York City Jets have lower rates of zero use before adjectives than before locatives and, as Rickford et al. (1991) have shown, the relative ordering of these two environments is often variable. The greater avoidance of zero copula with nominal predicates, by contrast, is almost exceptionless (Labov 1995:43, fn.3; Winford 1992:35), as is the preference for zero copula with verbal predicates and *gonna*.

σ_T^2 is the variance of the sums of the individual scores over the n trials (here, the variance of the sums of the copula context omission rates over the n speakers).

10. Note that although we have presented the data as line graphs in order to demonstrate the patterns in the data more clearly, the points are of course discrete and there is no implied relationship in the intervening space.

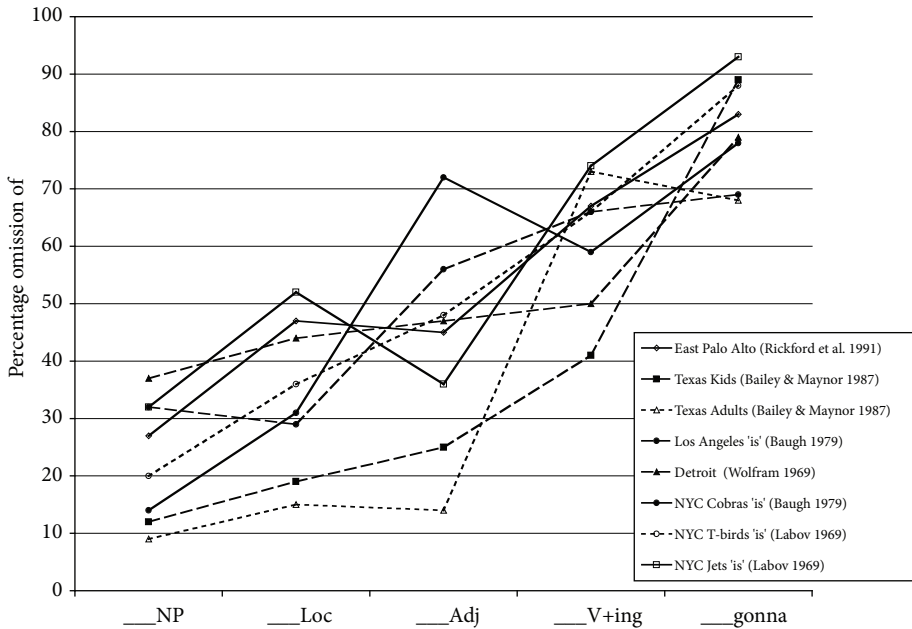


Figure 1. Copula Absence in Contemporary AAVE (based on Table 6.16 in Rickford 1998:190).

Average R_s (correspondence to AAVE order): .93 ($p < .05$, nearly perfect correlation); Cronbach's α (internal consistency): 0.969.

The characterization of this robust distribution as 'notoriously inconsistent' (Walker 2000:49) is a consequence of the reliance on broad assessments of similarity or difference between datasets without full application of appropriate measures to compare findings. This is of course not characteristic of Walker alone, nor of 'Anglicists' or 'dialectologists'; those who favor a 'creole origins' hypothesis are just as likely to claim overall similarity in the constraint patterning across AAVE and creole datasets without supporting their claims with statistical evidence of the type employed here.

The conformity of the disparate datasets to the basic order in (1) can be confirmed by the two statistical measures introduced earlier. In a pairwise Spearman rank-order correlation of each of the 8 datasets with the order in (1), 3 datasets showed perfect correlation ($R_s = 1$, $p < .01$), 4 datasets showed very strong positive correlation ($R_s = .9$, $p < .05$), and one showed strong positive correlation ($R_s = .8$). The average coefficient across the 8 datasets was significant at .93. Similarly, Cronbach's alpha is 0.969, showing that not just the ordering but the actual values of each group pattern are extremely consistent with other groups.

Figure 2, based again on Table 6.16 in Rickford (1998:190), shows the data by predicate type for 7 groups of creole English speakers — from Jamaica, Barbados,

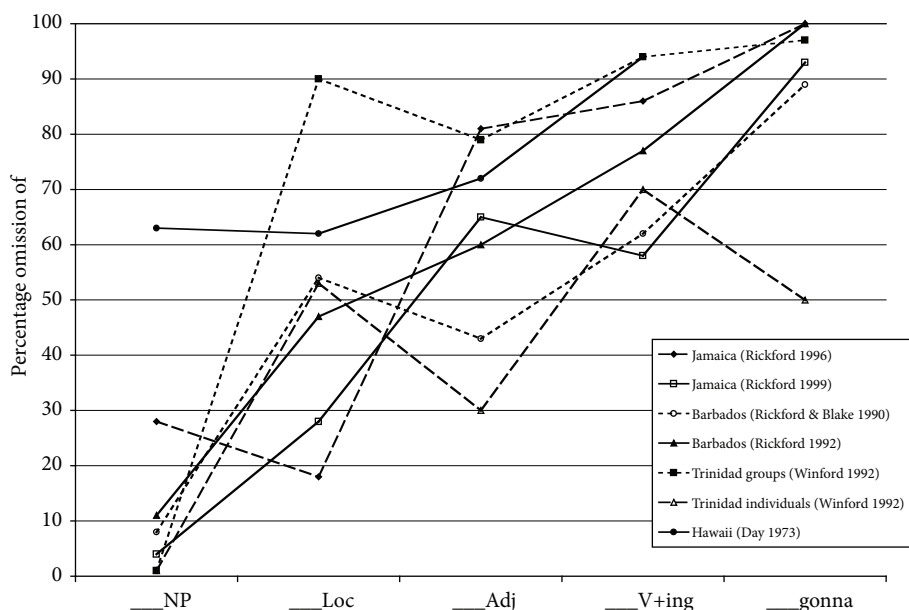


Figure 2. Copula Absence in creoles (based on Table 6.16 in Rickford 1998: 190).

Average R_s (correspondence to AAVE order): .84 (very strong correlation);

Cronbach's α (internal consistency): 0.951.

Trinidad, and Hawaii (see fn. 7 for sources). Once again, there are a few notable differences among the datasets: The major source of variability is that in three datasets (Barbados 1990 and the Trinidad group and individual data), copula absence before adjectives is lower than it is before locatives, while the reverse is true in the other four datasets. Beyond this common source of variability (noted earlier), the Trinidadian individuals show a decline in copula absence from the *V-ing* to *gonna* contexts, rather than an increase, and neither the 1996 Jamaican nor the Hawaiian Creole datasets show the least copula absence before NP predicates, the majority pattern.

Despite these fluctuations, there is, as in the case of the AAVE data, strong conformity to a common ordering of copula absence according to following grammatical context. In the pairwise Spearman rank-order correlation of each of the 7 datasets with the order in (1), one dataset showed perfect correlation ($R_s = 1$, $p < .01$), four datasets showed very strong positive correlation ($R_s = .9$, $p < .05$), and two showed strong positive correlation ($R_s = .8$ and $.5$). The average coefficient across the 7 datasets was very high, at .84. The actual group rates (not just orders) are also highly consistent with each other, as can be seen by the extremely high Cronbach's alpha coefficient (.951).

The similarity evident between the AAVE and creole patterns in Figure 1 and Figure 2 can be further confirmed by measuring the value of Cronbach's alpha for both sets of data combined. The value for conformity to a single pattern when the AAVE and creole datasets are combined goes up to 0.977, suggesting that a combination of both sets of data in fact reinforces the underlying pattern and results in even greater homogeneity.

L2 acquisition data

The data for the four L2 English acquisition cases examined in this section come from Indian English, South African Indian English, Singaporean English, and Spanish English. Winford (1998:114) specifically cites the L2 South African Indian English data and both sets of L2 Spanish data discussed here as parallels to AAVE copula absence. We pursue his claim here with a fine-grained comparison of these data to AAVE and creoles, and conclude that the actual orderings of copula absence by grammatical context do not suggest comparable systems at all. The two types of results we find across all of the L2 data are either: (i) no significant patterning according to predicate type, or (ii) a patterning quite different from that found in AAVE and creoles.

Indian English

The first set of data comes from Sharma's sociolinguistic interviews with eight Indian speakers of English, all of whom speak English as a second language. All individuals have had some English-medium education and use English on a daily basis, in conjunction with their first language, which is used in more intimate or familiar registers. While there are slight differences in proficiency, all eight speakers fall within the lower English proficiency end of a larger Indian English bilingual dataset; copula absence is not a robust feature among more proficient speakers of Indian English. The first languages of all the individuals are Indo-Aryan (2 Gujarati, 3 Hindi, 3 Punjabi) and all three languages require an overt copula in all predicate contexts (Masica 1991:336).

Figure 3 lists the four main grammatical contexts for copula use. *Gonna* is not separated as a predicate type due to insufficient tokens, but the summary of rates in (2) below shows that, contrary to AAVE, not a single token of *going to* omitted the copula. The 8 speakers are listed in the key in decreasing order of proficiency and, as in all figures in this paper, the x-axis repeats the hypothesized ordering in (1). The graph shows clearly that it is difficult to identify a generalized pattern for these speakers, and certainly not the AAVE pattern. This intuition is supported

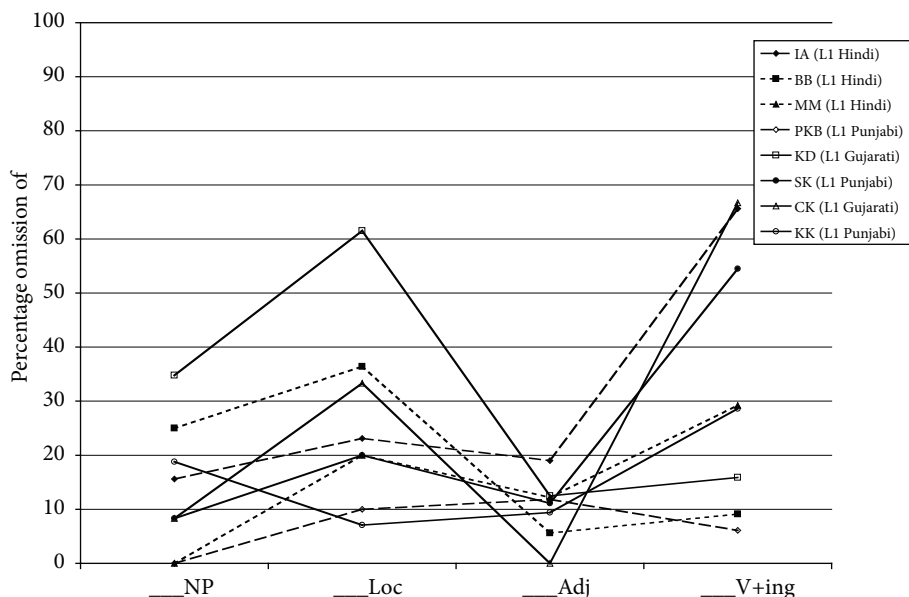


Figure 3. Copula Absence in Indian English.

N = 743; see (2) for details.

Average R_s (correspondence to AAVE order): .3 (very weak correlation);

Cronbach's α (internal consistency): 0.734.

by both statistical tests. Unlike the AAVE and creole data, the pairwise Spearman rank-order correlation of each of the 8 individual patterns with the order in (1) produces not a single significant correlation (i.e. no coefficient is .9 or higher). Three individuals showed strong positive correlation ($R_s = .8$), four showed weak positive correlation ($R_s = .4$), and two showed strong negative correlation ($R_s = -.6$). The average coefficient across the 8 speakers was very weak at .3, reflecting the wide disparity in orderings in relation to (1). Similarly, Cronbach's alpha is 0.734, showing some conformity to a pattern (mainly due to consistently low rates of omission with adjectival predicates), though one that is clearly distinct from the AAVE ordering and not reliable (below the .8 level).

These data diverge from the AAVE pattern in a few ways. First, no speaker has the AAVE order for copula absence. For four of the eight speakers, the rate of copula omission with an NP predicate is higher than that of adjectival predicates, quite unlike the general AAVE pattern. Second, locative predicates are the highest or second highest context for copula omission for seven of the eight speakers, also very unlike the low rates of omission in locative contexts in AAVE. Third, adjectival predicates, which are rather variable contexts for copula omission in AAVE and the creoles examined, although invariably higher than nominal predicates (see Figures 1 and 2), show consistently low copula absence in L2 Indian English.

Finally, although AAVE and creoles certainly exhibit considerable individual variability in copula use, the statistical measures show that this variability is on a significantly smaller scale than in the Indian data. The only partial resemblance to the AAVE pattern is that for five of the eight speakers the *V-ing* predicate shows the highest rate of copula absence.

Since the first languages of these speakers all require a copula in all contexts, there is no direct L1 source for their use of zero copula. In fact, the lack of copula omission in the L1s may cause more individual variation in the order of learning the English copula, as compared to a language that provides a template for contextually sensitive copula omission. Certainly it is clear that the AAVE pattern is not strongly replicated for Indo-Aryan learners of English. The average ordering across the group is given in (2), showing relatively weak differences among contexts and fairly low percentages.¹¹

(2) Indo-Aryan speakers:

___ <i>gonna</i>	< ___ Adj	< ___ NP	< ___ Loc	< ___ <i>V-ing</i>
0% (6)	11.6% (276)	16.5% (164)	26.4% (87)	30.5% (210)

A note on competing factors: Given the diversity of data sources in this paper, we are not able to conduct multivariate analyses of competing factors for each dataset, and we restrict ourselves to factors relating to predicate type. The Indian English data were nevertheless subjected to a preliminary VARBRUL analysis, which found predicate type and subject type to be significant factors, although not surprisingly, given the findings outlined above, variants within predicate type are ordered differently to the ordering found in VARBRUL analyses of the copula in AAVE and creoles. Tense and number were not significant, suggesting further differences with AAVE and creole copula absence patterns. A summary of the VARBRUL findings for these data is given in Table 2; these findings are briefly revisited in our later consideration of typological universals.¹²

11. The N values listed in all examples represent the total N; thus, for instance, in example (2), 11.6% of 276 adjectival predicate tokens omitted the copula.

12. As the Indian English data presented here (and new Englishes in general) cannot be assumed to share any particular constraints found to hold in AAVE or creoles, we included all subject types and tenses. The frequent, reduced, and fixed forms of *it's*, *that's* and *there's* were excluded from coding. Phonetic factors were coded but are not presented in Table 2 as conditioning effects for different verb forms would have to be presented separately. Interestingly, some commonly anticipated effects are not found in the Indian English data due to variation in number agreement, low rates of contraction, and prosody. For instance, frequent omission of *is* may be predicted after *this*, but 11/13 tokens of *this* are followed by overt *is*.

Table 2. VARBRUL results for Indian English copula omission

Factor	N	Percentage	Weight	
Predicate type:				
<i>Adjectival</i>	276	11.6	0.356	
<i>Nominal</i>	164	16.5	0.454	
<i>Locative</i>	87	26.4	0.586	
<i>Verbal</i>	216	29.6	0.681	<u>Range</u> = .325
Subject type:				
<i>1st person pronoun</i>	191	12.0	0.325	
<i>Demonstrative or expletive</i>	69	15.9	0.515	
<i>3rd person pronoun</i>	129	24.8	0.546	
<i>Noun</i>	339	22.1	0.577	
<i>2nd person pronoun</i>	15	33.3	0.615	<u>Range</u> = .290
[Tense:]				
<i>Present</i>	646	19.5	[0.498]	
<i>Past</i>	97	20.6	[0.511]	
[Subject Number:]				
<i>Singular</i>	510	16.9	[0.458]	
<i>Plural</i>	233	25.8	[0.591]	

N = 743; Input value: 0.177; Log Likelihood: -345.054; Significance: 0.002; X-square(13) = 19.548; factors in brackets not found to be significant

South African Indian English

The second set of data, originally presented in Mesthrie (1992:158), shows a different pattern for copula use in South African Indian English. The two sets of data presented by Mesthrie, and repeated in Figure 4, combine data from five speakers of Dravidian languages (Tamil and Telegu) and five speakers of Indo-Aryan languages (Bhojpuri and Urdu), and constitute the basilectal range of South African Indian English. Because the speakers are grouped it is not possible to establish whether the variance in these data is as great as in the Indian English data.¹³

Figure 4 again shows a patterning by predicate type quite different from the AAVE ordering. The main difference here is the high ranking of NP predicates as a context for copula omission. For Indo-Aryan speakers, copula omission with NP predicates is second only to verbal predicates, and for Dravidian speakers, a following NP context has the highest overall rate of zero copula. This is again

13. The South African Indian English data cover all subject person and number types but only include present tense copula use. This contrasts with the Indian English data, which included past and present tense; however, Table 2 showed tense to be a non-significant factor in the Indian English dataset.

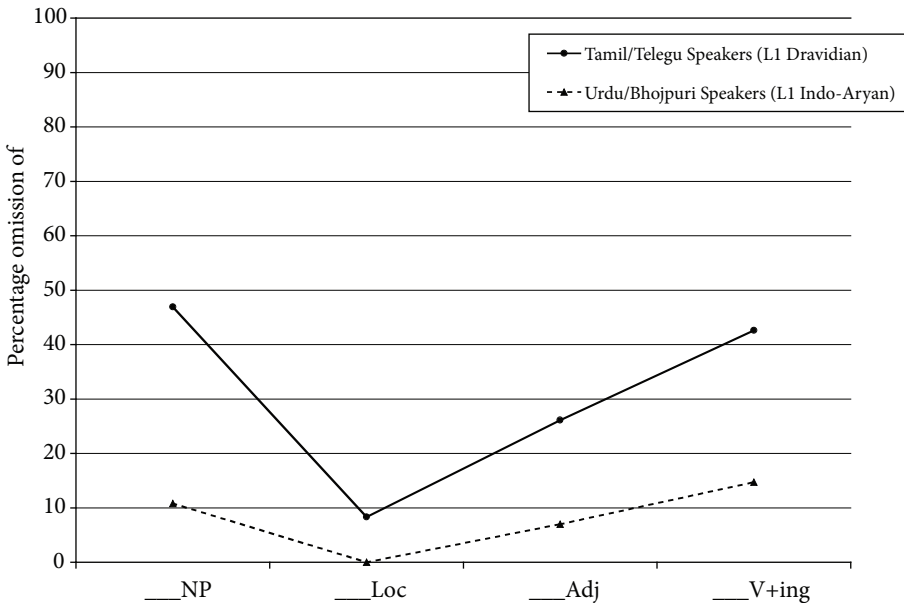


Figure 4. Copula Absence in South African Indian English (drawn from Mesthrie 1992: 158).

N = 332; see (3) and (4) for details.

Average R_s (correspondence to AAVE order): .2 (very weak correlation);

Cronbach's α (internal consistency): 0.744.

very different from the situation in AAVE and creoles, where copula absence is consistently lowest before an NP.

The pairwise Spearman correlation confirms the clear non-conformity of these two orders to the AAVE order in (1). The Indo-Aryan group shows a weak negative correlation with the AAVE ordering, with a coefficient value of $-.2$, and the Dravidian group shows a weak positive correlation with a coefficient value of $.4$, averaging to $.2$ overall and no significant correlations. As with the Indian English data, Cronbach's alpha falls just short of the 0.80 threshold for significance at 0.74, suggesting some degree of internal conformity, but again, as is clear from Figure 4, to a pattern quite distinct from the ordering in (1). Details of the two patterns apparent in the South African Indian English data are given in (3) and (4).

(3) Indo-Aryan speakers: ___ Loc < ___ Adj < ___ NP < ___ V-ing
 0% (19) 7% (43) 10.8% (37) 14.7% (34)

(4) Dravidian speakers: ___ Loc < ___ Adj < ___ V-ing < ___ NP
 8.3% (36) 26.1% (46) 42.6% (68) 46.9% (49)

Mesthrie attributes the Dravidian distribution to first language influence, as Dravidian languages (Tamil and Telegu) allow copula omission. In Tamil, NP-NP constructions are categorically zero copula constructions whereas predicate adjectives must either be nominalized or must appear in adverbial constructions with a copula (Pillai 1992:15; Schiffman 1999:141).¹⁴ Progressive constructions express tense on the verb form in Tamil, not a copula, so the high rate of omission with verbal predicates can also be linked to L1 transfer. In Telegu, similarly, zero copula is standard for NP-NP constructions (a specific copular form exists for negation), and optional for adjectival predicate constructions (Krishnamurti 1998:228).

The rates for the Indo-Aryan group are relatively low, as with the previous Indo-Aryan group listed in (2), possibly because the Indo-Aryan L1s do not permit copula absence. As with the Indian English data, it is more challenging to explain the Indo-Aryan distribution, given the lack of an L1 model. However, it is interesting to note the overall similarity of the two distributions in Mesthrie's data, which points to the likelihood that contact may have led to focusing of the variety (LePage and Tabouret-Keller 1985), perhaps towards the Dravidian-influenced model in the South African case.

An interesting parallel study to Sharma's and Mesthrie's data is Herat's (2005) examination of *be* variation in Sri Lankan English, as Sinhala, the primary substrate for this variety, is an Indo-Aryan language with extensive Dravidian influence. Herat's study is difficult to compare directly to the present data as N values are not provided in her discussion of different predicate types. Nevertheless, we present some key details here for purposes of comparison.

Herat separates *are* and *is* absence in her analysis due to the marked infrequency of the latter (74/429 instances of *are* absence; only 13/541 instances of *is* absence). For *are* absence, Herat finds the following ordering by predicate type: __ Loc (0%) ≈ __ NP (1%) < __ V-ing (15%) ≈ __ Adj (17%) < __ *going to* (54%). The apparent use of *going to* as a grammaticalized auxiliary and the low rates of omission with NP predicates resemble AAVE; however, other contexts show divergent distributions, particularly the low rates of omission with verbal predicates and with locatives. She argues that Sinhala cannot account for the specifics of Sri Lankan English copula omission because it lacks a copula in all contexts. However, she does not appear to take into account the use of copula-like existential verbs with locatives in Sinhala (Gutierrez Morales 2005), which could at least account for the low rates of omission in locative contexts. Sri Lankan English thus represents an additional divergence from the hypothesized AAVE pattern, bearing a few similarities but also a number of differences, some of which potentially arise from substrate effects.

14. Pillai (1992:15) suggests that copula absence in Tamil be treated as deletion, as the copula (*akku*) can occur sometimes in positive, negative, and relative participle constructions.

Singaporean English

The third set of data is from Platt (1979), a study of copula use among 59 Singaporeans with different language and educational backgrounds.¹⁵ Platt groups the speakers according to language-specific educational background, so for instance members of the group English School I had had tertiary (college-level) education in English, while those in English School III had had less than four years of English-medium secondary school. 41 of the 59 participants in the study had had English-medium education; the remaining individuals belonged to groups with Chinese- and Malay-medium education. The data extrapolated from this study are summarized in Figure 5 and details are provided in (5)–(7) below.

(5) Malay-medium	___ V-ing ≈	___ Nom <	___ Adj <	___ Loc
Education:	50% (20)	52.9% (17)	65.2% (46)	100% (5)

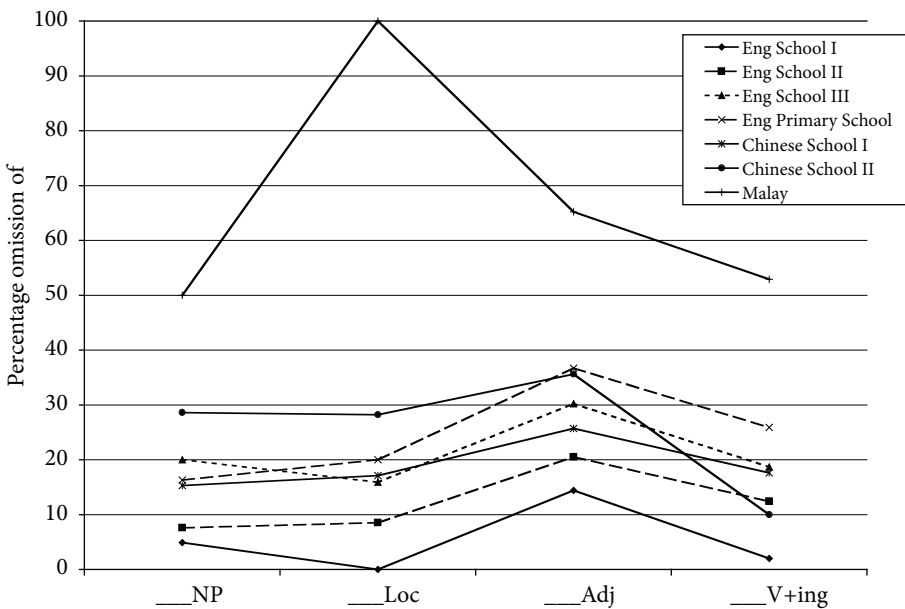


Figure 5. Copula Absence in Singapore English (drawn from Platt 1979).

N = 2440; see (5), (6), and (7) for details.

Average R_s (correspondence to AAVE order): .31 (weak correlation);

Cronbach's α (internal consistency): 0.604.

15. This dataset, like the Indian English data, includes past and present tense as well as all subject number and person types; in a later analysis of the data, Ho and Platt (1993:53) find subject type (1st and 3rd singular pronoun) to exert an influence but neither Platt (1979) nor Ho and Platt (1993) present clear results for tense.

(6) Chinese-medium	___ <i>V-ing</i> \approx	___ Nom \approx	___ Loc <	___ Adj
Education:	15% (73)	16.1% (161)	17.4% (46)	27.7% (224)
(7) English-medium	___ Loc \approx	___ Nom <	___ <i>V-ing</i> <	___ Adj
Education:	9.8% (164)	9.8% (612)	13.1% (237)	22.5% (835)

Whereas the previous two sets of data had some subset of speakers for whom the following *V-ing* context showed the highest rate of zero copula, Platt's study has NO group of speakers for whom this is the case. Instead, a following adjectival predicate environment is consistently the most favorable for zero copula, except for the Malay speakers, among whom a following locative occupies this position. More importantly, the distribution of data for all the groups except the Malay speakers in Platt's study is very flat, suggesting that predicate type may not be a strong determining factor in copula absence among L2 speakers of Singapore English.

In statistical terms, the data in Figure 5 clearly do not show any notable correlation to the original hypothesized order in (1). The Spearman correlation coefficient averages .31, a weak positive correlation, and the range of individual correspondences to the ordering in (1) included three strong positive correlations ($R_s = .8$), one weak positive correlation ($R_s = .2$), one confirmation of the null hypothesis ($R_s = .0$), and one negative correlation ($R_s = -.4$). This inconsistent range derives from the relatively flat distribution of values, causing some orders to switch. The overall Cronbach alpha value is .604, indicating little significant conformity to a pattern; the value excluding the Malay group is .622, indicating marginally greater cohesion among the Chinese- and English-educated groups.

Platt links aspects of the distributions in Figure 5, which are unrelated to other orderings we have seen so far, to the speakers' first language systems, Chinese and Malay. His description is summarized and extended in Table 3, in which '+' indicates use of some overt existential marker and '-' indicates absence; certain discourse exceptions are noted. The greater absence of copulas in Malay may explain the higher frequencies of zero copula among Malay-medium students.¹⁶ Similarly, the higher rate of omission with adjectival predicates among Mandarin and Cantonese speakers learning English may derive from its parallel absence in their first languages. The slightly lower rates of omission with verbal predicates among

16. Platt (1979) does not discuss verbal predication in these languages. We thank Shiao-Wei Tham and Andrew Wong for additional detail on the Malay and Chinese systems, respectively. Note also, in (5), that the exceptionally high (100%) rate of zero copula with locative predicates among the Malay speakers is based on only five tokens.

Table 3. Copula use according to non-verbal predicate type in L1s of Singaporeans. (Compiled from Platt (1979:5–9) and consultations with native speakers. ‘–’ indicates absence of an overt existential marker, ‘+’ indicates presence of one, and parenthetical comments list details of variation.)

	Adjectival	Nominal	Locative	Verbal
Cantonese/ Mandarin	– (exceptions: overt marker for contrast in Mandarin and for em- phasis in Cantonese)	+ (overt marker more likely for assertion than contrast or narrative)	+ (overt markers take the form of locational verbs of existence)	–
Malay	–	– (exception: overt marker for emphasis)	– (exception: option- al use of locational verbs of existence)	–

Chinese L1 students does not appear to derive as clearly from Table 3, which predicts that verbal predicates would strongly favor omission. However, later research by Ho (1986, cited in Ho and Platt 1993:13), with 100 ethnically Chinese, English-medium-educated Singaporeans, showed that the overall rate of omission with *V-ing* was comparable to high rates of omission with adjectival predicates.

Finally, it is interesting to note informally that the different Singaporean English groups potentially show less variance as compared to Indian English. As with Mesthrie’s data, this may mean that Singaporean English as a variety is more focused.

Spanish learners of English

To conclude the L2 data comparison, we examine two additional studies — Butterworth and Hatch (1978) and Shapira (1978) — each a longitudinal study of the early stages of English use by an adult Spanish-speaking learner. As mentioned earlier, these are two of the studies mentioned by Winford (1998:114) as potential parallels to AAVE. The studies are not subjected to the statistical measures of consistency and conformity used above because of their distinct methodology and because data on the full set of predicate types is not available. However, as these studies are longitudinal, unlike any of the studies discussed above, they allow us to explore whether the hypothesized order for copula omission according to predicate type surfaces in the order of acquisition of *be* in individual learning too. This might be anticipated if the hypothesized order is a universal or widespread effect, as has been suggested, as copula-demanding contexts may favor copula use before other contexts.

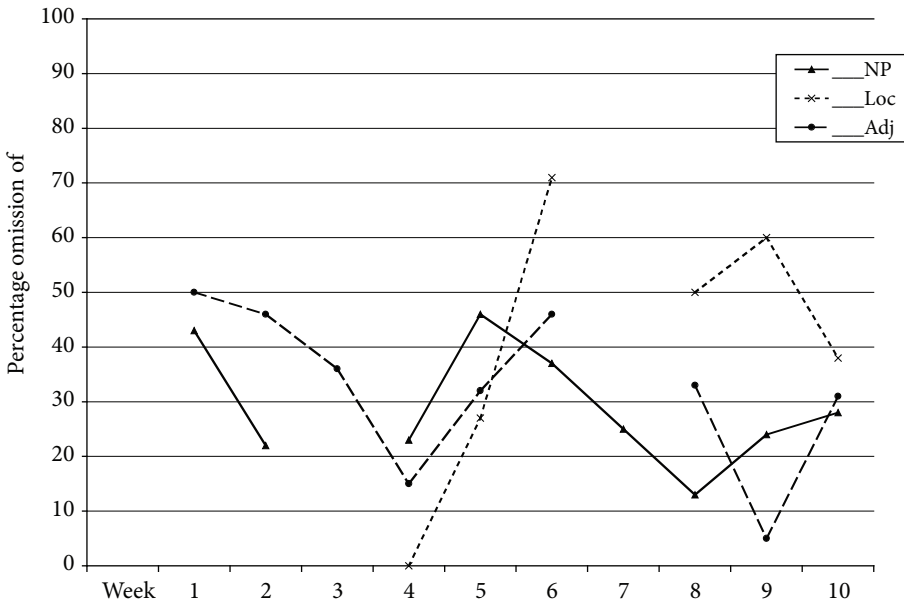


Figure 6. Longitudinal Copula Absence in Spanish English.

(single speaker, drawn from Butterworth and Hatch 1978).

N values not reported in original study. No statistical tests applied.

Butterworth and Hatch (1978) examine only nominal, adjectival, and locative predicates (copula environments in the narrow sense). The subject is a 13-year-old Colombian boy who was studied over a period of 10 weeks; his exposure to English prior to the study was very limited. The authors mention the use of auxiliaries with a progressive verb form but only to note that their subject had not really acquired the construction yet: only 1% of his attested verb forms had the *-ing* inflection with an auxiliary verb, and another 1% had the *-ing* inflection with no auxiliary. This supports the agreement across L2 studies in Table 1 that auxiliary constructions are acquired later than copular constructions.

Figure 6 shows what the authors describe as essentially ‘free variation’ (p. 236), namely no systematic correlation of copula absence with grammatical context. NP predicates sometimes have the lowest rate of copula omission (weeks 1 and 2) and sometimes the highest (weeks 4 and 5); similarly, locative predicates sometimes have the lowest rates (weeks 4 and 5) and sometimes the highest (weeks 6, 8, and 9); and finally adjectival predicates also sometimes have the lowest rates of copula omission (week 9) and sometimes the highest (weeks 1 and 2). The only potentially observable pattern, despite the data gaps at various points, is the high rate of copula absence with locative predicates once such constructions begin to appear in week 4. There is also an overall decline in copula omission

over time, although the process is erratic. The data bear no resemblance to the ordering in (1).

Shapira (1978) also examines longitudinal data for one Spanish speaker, a 25-year-old Guatemalan woman, recorded over a period of three years and with little prior exposure to English. Shapira's data are also incomplete for our present purposes, as she only contrasts verbal and non-verbal predicates, the latter being grouped as a single category (PredP). Differences between nominal, adjectival, and locative predicates are not presented. At each longitudinal stage, Shapira (p. 250) reports a rate of 31% (N = 52), 31% (N = 80), and 27% (N = 147) absence of *be* with a following PredP as against 100% (N = 12), 81% (N = 16), and 95% (N = 61) absence of *be* with a following progressive verb form, suggesting a strong resemblance to the high rate of zero copula with *V-ing* in other datasets; however, without the details of the three types of PredP contexts, we cannot tell whether this pattern resembles the overall AAVE pattern.

The two studies of English used by Spanish learners serve to corroborate (i) the lack of an early L2 preference for a particular learning order within non-verbal predicates, and (ii) the tendency towards higher rates of omission with verbal as compared to non-verbal predicates.

This section has presented six primary sets of data — AAVE, creoles, and four L2 varieties. The Indian English data showed a high degree of variability and the slight trends identifiable did not resemble the AAVE pattern, with the exception of the high rate of zero copula in the *V-ing* context for some speakers. Mesthrie's South African Indian English data also showed a trend unlike AAVE in its high ranking of predicate NP as a context for copula omission. The Singaporean English data from Platt's study differed from the others in exhibiting a notably flat distribution of zero copula across contexts. The individual Spanish speakers' longitudinal English data also showed no pattern among non-verbal predicates, but again higher rates of zero copula in *V-ing* contexts. L1 transfer was shown to underlie a number of these divergent distributions, although not always in the case of high rates of *be* omission with verbal predicates.

Discussion

Table 4 summarizes the patterns of grammatical conditioning across datasets. No single ordering of copula absence in relation to nominal, adjectival, and locative predicates is evident in the survey of L2 English acquisition. In particular, there is no evidence of a generalized second language learning preference for NP predicates to have an overt copula, as appears to be the case in AAVE and creoles. On the contrary, language transfer (as in the case of Dravidian languages) can actually

Table 4. Summary of patterns of copula absence

Dataset	Ordering of predicate contexts (lowest to highest rate of zero copula)								
A AAVE	NP	<	Loc	<	Adj	<	V-ing	<	gonna
B Creole	NP	<	Loc	<	Adj	<	V-ing	<	gonna
C Indian English (Indo-Aryan L1s)	gonna	<	Adj	<	NP	<	Loc	<	V-ing
D South African Indian English (Mesthrie 1992, Indo-Aryan L1s)	Loc	<	Adj	<	NP	<	V-ing	<	Loc
E South African Indian English (Mesthrie 1992, Dravidian L1s)	Loc	<	Adj	<	V-ing	<	NP	<	Loc
F Singapore English (Platt 1979, Malay-medium)	V-ing	≈	NP	<	Adj	<	Loc	<	Loc

G	Singapore English (Platt 1979, Chinese-medium)	___ <i>V-ing</i>	≈	___ NP	≈	___ Loc	<	___ Adj
H	Singapore English (Platt 1979, English-medium)	___ Loc	≈	___ NP	<	___ <i>V-ing</i>	<	___ Adj
I	Singapore English (Ho 1986, Chinese L1)	___ Loc	≈	___ NP	<	___ Adj	≈	___ <i>V-ing</i>
J	Spanish Learner of English (Butterworth & Hatch 1978)	___ NP	≈	___ Adj	≈	___ Loc	<	___ <i>V-ing</i>
K	Spanish Learners of English (Shapira 1978)					___ PredP	<	___ <i>V-ing</i>

encourage zero copula precisely in such contexts. The only fairly systematic pattern that appears to be shared across datasets is the higher rate of omission of auxiliary as opposed to copular *be*.

In this section, we first furnish evidence from typology that corroborates the apparent lack of a universal grammatical conditioning of copula use with nominal, locative, and adjectival predicates. We then seek an explanation for the more widely shared pattern of omission of auxiliary *be* with verbal predicates.

Further evidence against universality: Typological diversity in copula systems

A typological corollary to a universal L2 learning order would take the form of implicational universals: for example, any language permitting zero copula with nominal predicates would be predicted to permit zero copula with adjectival predicates, as zero copula in the former construction is predicted to be more marked in some way.¹⁷ If no universal implicational order exists for the copula, the absence of such an order in L2 acquisition should be reflected in a lack of evidence of typological markedness.

Stassen (1994), in a cross-linguistic overview of copula systems, identifies three systematic parameters of copula use that surface in a number of unrelated languages. One parameter he describes is semantic: the ‘permanency parameter’ associates zero realization of the copula with ‘permanent’ or ‘essential’ properties of the subject. The other two parameters in his discussion are structural: the ‘present parameter’ and the ‘third person parameter’. The ‘present parameter’ governs languages such as Russian, Swahili, and Arabic (and AAVE; see Pullum 1997) in which zero copula occurs only in the present tense. The ‘third person parameter’ governs languages such as Hungarian, which reserve zero copula for third person subjects.¹⁸

A parameter sensitive to predicate type could be hypothesized as another universal of this type, such that zero copula is systematically dispreferred with nouns, preferred with verbs, and intermediate elsewhere. However, Stassen presents

17. For example, Labov’s informal suggestion, noted in Rickford (1977:210), that nominal predicates might universally be more ‘copula demanding’ is one such typological claim. His speculation is disconfirmed by our survey.

18. AAVE also shows sensitivity to subject type, though not strictly the ‘third person’ parameter: Wolfram (1969) suggests greater copula omission with plural subjects and second person singular subjects. Interestingly, the Indian English data presented in Table 2 do show a slight preference for these two subject types, though not statistically significant in the case of plurality. Rickford (2006:Table 1) comments on the subject type constraint, citing evidence that it may be weaker than the predicate type constraint, which appears to be true for the Indian English data as well.

counter-evidence to such a universal. He observes that several Uralic and Papuan languages have a split system in which an overt copula is used with third person present locational predicates and zero copula with third person present nominal predicates. In other words, these languages reserve zero copula strictly for nominal predicates. Stassen goes on to observe that in his database of 385 languages, zero copula with predicate nominals is in fact more common than with locational predicates. While the two structural parameters account for tense and number restrictions cross-linguistically, splits according to predicate type do not appear to follow a single typological direction, and sometimes run counter to the AAVE ordering.

In sum, although intriguing universal tendencies do arise in cross-linguistic copula use, they do not appear to include a universal preference for the AAVE ordering based on following grammatical context. Thus, both the L2 data and the typological literature militate against a generalized pattern.

A universal bipartite distinction in the use of copular and auxiliary be?

While no universal ordering for copula omission has emerged for L2 English, the pattern of higher rates of omission of *be* in auxiliary as compared to copular uses is considerably more consistent. With the exception of the Singaporean datasets F and G in Table 4, the *V-ing* predicate is always one of the two most favorable grammatical environments for zero *be*, even in the absence of an L1 source for this pattern. Recall that Table 1 found this pattern to be present in longitudinal SLA studies as well. Furthermore, L1 English acquisition studies also consistently find greater omission of auxiliary *be* than copula *be*. We consider two potential explanations for this difference between verbal and non-verbal predicates here: (i) stage and individual level semantics, and (ii) redundancy avoidance.

In her work on L1 acquisition of *be*, Becker (2000, 2002) has proposed that semantic properties of predicates and the incremental development of children's functional projections can explain higher rates of omission of *be* with verbal predicates, locative predicates, and certain adjectival predicates. She provides precise rates of omission for four children (age range 2;0–3;4) for all grammatical contexts in L1 acquisition, and finds the systematic ordering in (8), also illustrated in Figure 7.¹⁹

19. Becker (2002) only presents non-verbal predicate data; rates for verbal predicates in (8) were kindly provided by Misha Becker via personal communication, based on her dissertation (Becker 2000). Brown (1973:380) also reported higher rates of omission of *be* before a progressive verb as opposed to other types of predicates in L1 acquisition, but did not provide a complete set of rates for these environments or differences within non-verbal predicates. L1 acquisition has

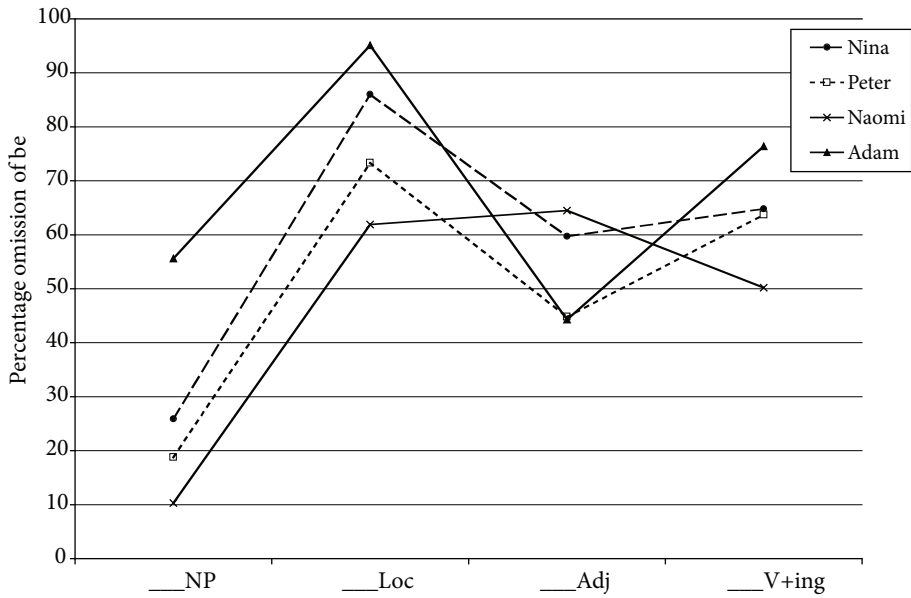


Figure 7. Copula Absence in Child English (drawn from Becker 2000, 2002).

N = 2428; see (8) for details.

Average R_s (correspondence to AAVE order): .3 (weak correlation);

Cronbach's α (internal consistency): 0.787.

- (8) L1 Child English: ___ NP < ___ Adj < ___ V-ing < ___ Loc
 27.6% (949) 51.8% (386) 64.5% (831) 79.1% (262)

The low rates of zero copula with NP predicates and the high rates of omission with verbal predicates resemble the AAVE pattern, although the highest rates of omission are clearly with locative predicates. The lack of an exact replication of the AAVE order is confirmed in the pairwise Spearman correlation of each child with the AAVE ordering listed in (1). The average correlation coefficient is .3, a relatively weak positive correlation, and the range of individual correspondences

also been examined among AAVE learners. In a study of copula acquisition by ten 3- to 5-year-old African American children, Wyatt (1996:105) found the ordering for non-verbal predicates shown in (i), similar to the AAVE order:

- (i) L1 Child AAVE: ___ NP < ___ Adj < ___ Loc
 (Total N = 1148) 18% 27% 35%

Wyatt shows that these children's pattern of copula omission closely mirrors the adult African American speakers' usage in her study. Such studies of AAVE acquisition (see also Kovac 1980) support the robustness of the AAVE copula omission system, even in early acquisition, but are not directly relevant to the present discussion, which focuses on the omission of the copula in situations where the target variety at least originally required an overt copula.

included three weak positive correlations ($R_s = .4$) and one confirmation of the null hypothesis ($R_s = .0$). By contrast, Becker's data is very consistent internally, with a Cronbach alpha value of .787, suggesting close conformity to the pattern in (8) rather than to the AAVE pattern.

In her analysis, Becker further divides her adjectival predicate data into two sets: stage-level and individual-level predicates. Individual-level predicates are those which denote permanent or inherent properties of the subject (e.g. *Mary is tall*; *Mary is a girl*), while stage-level predicates denote temporary or accidental properties (e.g. *Mary is tired*; *Mary is in the garden*; *Mary is sleeping*) (Carlson 1977; Kratzer 1995). Becker finds that the individual-level adjectival predicates have a lower rate of zero copula (31.7%), patterning more like NP predicates, which are also generally individual-level predicates. Stage-level adjectival predicates were found to have a higher rate of zero copula (53.8%), patterning slightly more like locative contexts, which tend to share the property of being stage-level. Becker argues for an analysis in which stage-level predicates may occur in non-finite clauses (permitted in child English), with temporal anchoring taking place with the head of an Aspect Phrase rather than Tense, whereas individual-level predicates are non-aspectual and restricted to finite clauses, requiring the functional projection Infl and thus generally corresponding to overt copula use. She also notes in passing that the relatively high rate of omission of *be* with progressive predicates could be subsumed under this account as the category of progressive is aspectual, like stage-level predication.²⁰

Becker's argument could account for verbal predicates developing higher rates of omission across L2, AAVE, and creole data based on their aspectual, or specifically stage-level, status. However, this analysis predicts equally high levels of copula omission with locative predicates and low levels of omission with nominal predicates; in other words it is too restrictive an explanation to cover the diversity of distributions we have found in L2 data, AAVE, and creoles.

Furthermore, although the stage/individual-level contrast in copula use has some cross-linguistic support, a comparison to Stassen's discussion suggests it may not be universal, and thus need not be expected to arise independently in all contact situations. Becker (2002:46) cites evidence of comparable effects in Russian, Spanish, and Portuguese. Hebrew also corresponds to Becker's data, as predicate nominals obligatorily require an overt copula but other constructions optionally

20. Becker (2002:54) notes, however, that the stage/individual distribution is not perfect in her data, and also that the longitudinal asymmetry in finiteness does not necessarily extend to main verbs in child English as might be expected.

permit zero copula (Rothstein 2001, Greenberg 1998).²¹ Here, as in Becker's data (and AAVE and creoles), zero copula is prohibited in NP-NP constructions due to their individual-level or 'permanent' reading. Crucially, however, Stassen's 'permanency parameter' mentioned earlier makes precisely the opposite claim about copula overtness and semantic permanency. Stassen's survey includes a robust group of languages which require zero copula specifically in NP-NP constructions, where a permanent attribute is predicated of a subject.²²

Becker's analysis does invite future examination of AAVE and creole copula data in terms of aspectual semantics, to see whether individual and stage-level adjectival predicates align with nominal and locative predicates, respectively, in terms of frequency of copula omission (see Cukor-Avila 1999 for an interesting exploration of this idea). If further conformity between Becker's data and AAVE and creole data (to the exclusion of L2 data) is found, and if West African L1 copula systems are not found to be sensitive to stage and individual-level readings, the results may feed into discussions of emergent UG principles in creoles (as in L1 learning). The absence of this pattern among L2 speakers in our datasets would then be explained by the availability of their specific L1 systems (although this does leave their high rates of omission with verbal predicates unexplained). For now we conclude that, while the stage/individual-level contrast is intriguing, it is unable to explain the shared omission of auxiliary *be* across almost all our sample data without over-predicting a shared (and unattested) pattern of copula *be* use. Note that even a UG explanation linking AAVE, creoles, and L1s would favor a single analysis of AAVE and creoles, distinct from L2 learning processes.

A second available explanation for the copula/auxiliary difference has been referred to as the 'Dummy Hypothesis' in the literature, and represents a view attributed by Stassen (1994) to a lineage starting as early as Meillet (1906), and including Lyons (1968) and Dik (1983). All of these approaches have treated the copula as a semantically empty element, devoid of independent lexical meaning, which carries grammatical information that cannot be marked on the predicate.

21. Furthermore, optional copula constructions in Hebrew are subject to individual-level readings when the copula is overt and stage-level readings when it is absent. For instance, when the copula in the Hebrew sentence *the sky is blue* is overt, the reading is that blue is always the color of the sky; when the copula is absent, the reading is that the sky is currently blue as opposed to being overcast (Rothstein 2001:233). Rothstein compares the stage- and individual-level contrast in optional copula omission in Hebrew to optional omission of *to be* in English in examples such as: *I consider John (to be) nice* and *I consider John *(to be) Mr. Smith*. However, she also suggests a critique of the 'generic' or 'permanency' analysis of overt copulas, offering counter-examples in Hebrew in which an obligatory copula is clearly associated with a temporary status meaning.

22. Ferguson (1971:142) also remarks on this systematic copula type, classifying them as Type B languages.

Safarewicz (1974:14) suggests that while a predicate consisting of a finite lexical verb provides both lexical and logical content, a non-verbal lexical predicate requires an additional form, the copula, to supply the logical content of existence. An extension of this view to variable copula omission could argue that learners do not identify finiteness morphology accurately and treat morphology such as the *-ing* verbal marker or the *gonna* verb form as bearers of logical (finiteness or tense) content, thus rendering use of the copula redundant.

This concurs with other studies which show evidence that second language learners initially favor unique form-function correspondences to optimize production and processing burdens (Ellis 1997; Pienemann 1998). Schumann (1974:148) has argued that the use of *be* + *V*stem and \emptyset + *V-ing* among learners represents an attempted ‘rejection of redundancy’ and Labov (1972:113) similarly noted that ‘it seems likely that the deletion of that *be* [auxiliary] (in its finite form) is connected with its redundant relation to the following *-ing* form.’²³

The attractiveness of this account is that, unlike the stage/individual-level account, it permits an explanation of the copula/auxiliary difference found across L1, L2, AAVE, and creole data, without making any claim about ordering within non-verbal predicate contexts, the ordering of which may instead derive from L1 influence. Based on the datasets we have examined, we provisionally favor the second view, which allows us to combine redundancy avoidance (to explain shared patterns of auxiliary *be* omission) with L1 transfer (to explain idiosyncratic patterns of copula *be* omission).²⁴

A return to the substrate in AAVE and creoles

Although specific L1 influence does not completely determine copula use by predicate type among the L2 speakers in the present paper, there is much clearer

23. Stassen (1994:112) offers a number of critiques of the Dummy Hypothesis; for instance, he notes that languages with ‘inert’ TMA systems might be predicted not to have a copula, as no form is needed to bear TMA information, yet these languages often have or even develop a copula. Nevertheless, we believe the redundancy view of copular and verbal marking could be an active factor in learner perceptions.

24. Bill Haddican (pers. comm.) makes the very interesting point that the copula/auxiliary *be* distinction we draw here may have a parallel in *have* usage among AAVE speakers, if these speakers accept *have* deletion in perfect sentences (*She (has) lived here a long time*) but not in possessive constructions (*I *(have) 4 dollars*). In such cases, the presence of verbal morphology to assist recovery of the perfect interpretation in the former construction — as with the presence of *-ing* in progressive constructions with *be* — may have the identical effect of licensing auxiliary deletion. Widely observed similarities in the behavior of possessive *have* and copula *be* (vs. auxiliary *have* and *be*) across languages would support such a parallel.

evidence of this factor than of a single shared system emerging in all situations. We provisionally suggest that specific L1 influence from shared West African origins remains one of the strongest available explanations for the shared AAVE and creole pattern, with the possible influence of a cross-linguistic ‘perceived redundancy avoidance’ tendency with verbal predicates.

John Holm (1976) was among the first to note the possibilities of substrate influence in AAVE from English-based creoles like Jamaican and Guyanese, which differentiate nominal, locative and adjectival predicates by the use of *(d)a*, *de* and \emptyset respectively, and/or from West African languages like Yoruba, which differentiate nominal, locative, adjectival, and verbal predicates by the use of entirely different forms like *u*, *ri*, *wà*, and *fe*, as well as \emptyset .²⁵ Under this view, creoles (often basilectal varieties in particular) maintain systems closest to the highly differentiated West African copula systems, while decreolization has led to mesolectal systems merging these distinctions to varying degrees. It is possible to see AAVE as one such merged system, in which copula absence differentiation by frequency in different predicate types is a corollary of its ancestral differentiation in African languages and creole English by form. Holm’s hypothesis continues to have a certain attractiveness, as English offers NO basis for a distinction in copula form or absence by predicate type: The distinction between English *is* and *are* is based only on the person and number of the subject, and is unaffected by whether the predicate is a noun, locative, adjective, or verb.

A full review of the precise nature of contact with creoles and African languages, which may have given rise to these substrate effects in AAVE, is beyond the scope of this paper; these sociohistorical conditions have been explored in other work (e.g. Dillard 1972; Mufwene 2000; Rickford 1997; Winford 1997). Certainly ‘indirect’ (creolized Caribbean) and ‘direct’ (West African) influences may have both played a part as social relations and demographics varied regionally and chronologically. For example, the former input type may have been a stronger source in the seventeenth century, with the latter type dominating in the

25. Winford (1998:110) cites two earlier instances of the idea of a substrate link between AAVE and creole copula systems: Bailey (1965) and in particular Bickerton’s (1971) comparison of mesolectal stages of decreolization in Guyanese creole to AAVE. The Yoruba example was introduced by Holm as representative of West African languages which may have played a role in the creation of Caribbean English-based creoles. It would be helpful to examine the copula systems of other potentially relevant West African languages (e.g. Twi, Vai, Igbo) to see whether their copula structures are similar. John Singler (pers. comm.) notes that widespread enslavement of the Yoruba did not begin until after 1750. But given that Bight of Biafra (primarily Nigeria) slaves constituted the largest single identifiable region from which British slave exports were made from 1711–1810, and given that the Yorubas were included in this group, their relevance is still clear.

eighteenth century, with changes in slave labor importation practices and demographics (Mufwene 2000:240–242).

Some challenges to the substrate influence explanation of the AAVE copula system remain. Why are adjectives widely regarded as a prototypical zero category in Caribbean English creoles, but not in AAVE (Mufwene 1992), and why are locative and nominal predicates, both marked by substantive (non-zero) forms in Yoruba and the creoles, regularly differentiated in terms of zero frequency in AAVE? And what of Hawaiian Creole, where the likelihood of African influence is low or nil? These and other problems with substrate and decreolization explanations are discussed in Rickford (1998:180–81, 186). Walker's (2000) proposal that the copula absence distributions by predicate type in AAVE and Early AAVE might be attributable to prosodic structure, however, has even greater weaknesses (see Sweetland, Rickford, and Hsu 2000, and Rickford 2006). The present study has shown that some form of substrate transfer remains a robust possibility for the variable copula systems examined here.

Conclusions

'Simplification' of the target language is of course a process widely, if not universally, associated with pidginization, creolization, and other kinds of language acquisition, both second language and first. It has been conventional to regard copula absence, where the target language has a full copula, as a variety of simplification; however, the phenomenon that is at issue is not copula absence per se, but rather the particular, relatively complex pattern of absence by following grammatical environment.

Using two statistical measures, we have shown in this paper that while there is a relatively consistent ordering of zero copula rates by following grammatical environment in AAVE and English-based creoles, this pattern has no consistent parallel ordering in the second language varieties of English we have examined (Indian, South African Indian, Singaporean, Spanish). This evidence weakens the claim that the creole and AAVE patterns might have developed independently based on universal strategies or mechanisms of 'Creative Construction'. General claims of this type have been popular in generativist models of SLA (Dulay and Burt 1974; Selinker 1992), even though these have been restricted to the learning of English (as pointedly discussed in Thomas 2002). There is scant support in the SLA literature and in the case studies examined here for a universal L2 learning trajectory across the five predicate contexts in (1).

The only shared pattern we find is the higher rate of *be* absence with *V-ing* in L1, L2, AAVE, and creoles. This may be attributed in part to the perceived

redundancy, on the part of the learner, in using an auxiliary with a verbal predicate. Auxiliary uses of *be* with *V-ing* and *gonna* involve verbal content or inflection (including suppletion of *is* and *are*) at a minimum of two points in the clause — the auxiliary and the progressive verb — whereas copular sentences with non-verbal predicates require morphological inflection at only one point. The fact that the morphological suppletion of *be* is itself a challenge for speakers is shown by the common tendency for speakers to use undifferentiated *is* among all numbers of persons in the present, and *was* in the past (Bickerton 1975:76). This perception of redundancy can explain high rates of auxiliary *be* omission insofar as speakers perceive progressive and *gon(na)* future predicates as bearing sufficient marking of verbal features, while non-verbal predicates require an overt bearer of verbal information in the form of a copula.

But these auxiliary/copula distinctions and theories of redundancy avoidance only go so far, leaving completely unexplained the shared internal distribution among NP, Loc, and Adj (copula) environments, or between *V-ing* and *gonna* (auxiliary) environments in AAVE and creoles. We have shown here that a notion of generalized imperfect learning (Hypothesis I) cannot account for this pattern, as there appears to be neither SLA nor typological support for a universal order across these contexts. The substrate language as a template (Hypothesis II) has proven to be more promising for the L2 distributions we examined, as in the case of zero copula with NP predicates among Dravidian language speakers, adjectival predicates among Chinese language speakers, and across all predicates among Malay speakers (lines E, F, and G in Table 4). A shared substrate explanation may also account for the strikingly parallel AAVE and creole orderings, given the apparent lack of a universal basis for their particular distribution and the strength of substrate effects in L2 copula systems.

Despite the popularity of substrate or transfer explanations in creole studies, and their resurgence in linguistics more generally, it must be admitted that they have their problems as well. The Indo-Aryan language speakers of South African Indian English in line D of Table 4 show a relatively high rate of zero copula before NPs as well (second only to *V-ing*), but there is no grammatical warrant for this in their native languages; in this case, contact with Dravidian speakers could be a factor. The Indian English speakers in line C, lack both an L1 and a contact model for their copula omission, and some of their variation remains unaccounted for; further investigation of stage- and individual-level semantics may reveal that this does in fact intervene in the absence of an L1 template and could explain the higher use of zero copula with locative predicates in line C in Table 4.

Nonetheless, our survey has generally supported a limited substrate explanation for emergent copula systems — rather than a shared learner system across varieties — with only a localized ‘universal’ learning tendency arising within the

context of *V-ing* predicates. A completely satisfactory solution to the question of where the highly consistent pattern of copula absence by predicate type in creole English and in AAVE originate from still eludes the field. However, this paper has shown that attributing it to independent inheritance of a 'universal' or general pattern of imperfect second language acquisition is not supported by the available empirical evidence.

Submitted: 9/28/04

Revised and resubmitted: 12/01/07

Accepted: 3/5/08

References

- Agresti, Alan, and Barbara Finlay (1997) *Statistical Methods for the Social Sciences* (3rd edition). New Jersey: Prentice Hall.
- Andersen, Roger W. (ed.). (1983) *Pidginization and Creolization as Language Acquisition*. Rowley, MA: Newbury House.
- Bailey, Beryl (1965) Toward a New Perspective in Negro English Dialectology. *American Speech* 40, 171–177.
- Bailey, Guy, and Natalie Maynor (1987) Decreolization? *Language in Society* 16:449–73.
- Baugh, John (1979) Linguistic style shifting in Black English. Unpublished PhD dissertation, University of Pennsylvania.
- Baugh, John (1980) A re-examination of the Black English copula. *Locating language in time and space*, ed. by William Labov. New York: Academic, 83–106.
- Baugh, John (1983) *Black street speech: Its history, structure, and survival*. Austin: University of Texas Press.
- Becker, Misha (2000) The Development of the copula in child English: The lightness of *be*. PhD dissertation, University of California, Los Angeles.
- Becker, Misha (2002) The Development of the copula in child English: The lightness of *be*. *Annual Review of Language Acquisition* 2: 37–58.
- Bickerton, Derek (1971) Inherent Variability and Variable Rules. *Foundations of Language* 7, 457–492.
- Bickerton, Derek (1975) *Dynamics of a creole system*. Cambridge: Cambridge University Press.
- Brown, Roger (1973) *A First Language*. Cambridge, MA: Harvard University Press.
- Butterworth, Guy, and Evelyn M. Hatch (1978) A Spanish-speaking adolescent's acquisition of English syntax. *Second Language Acquisition: A Book of Readings*, ed. by Evelyn M. Hatch. Rowley, MA: Newbury House, pp. 231–245.
- Carlson, Gregory (1977) Reference to Kinds in English. PhD dissertation, University of Massachusetts, Amherst.
- Cronbach, L. J (1951) Coefficient alpha and the internal structure of tests. *Psychometrika* 16, 297–334.
- Cukor-Avila, Patricia (1999) Stativity and Copula Absence in AAVE: Grammatical Constraints at the Subcategorical Level. *Journal of English Linguistics* 27: 341–55.

- Day, Richard (1973) Patterns of variation in copula and tense in the Hawaiian post-creole continuum. *Working Papers in Linguistics* 5.2. Department of Linguistics, University of Hawaii. (Published version of Day's 1972 University of Hawaii PhD dissertation.)
- Dik, Simon C. (1983) Auxiliary and Copula Be in a Functional Grammar of English. *Linguistic Categories: Auxiliaries and Related Puzzles*, ed. by F. Heny and B. Richards. Dordrecht: Reidel, 121–43.
- Dillard, J. L. (1972) *Black English: Its History and Usage in the United States*. New York: Random House.
- Dittmar, Norbert (1980) Ordering adult learners according to language abilities. *Second Language Acquisition*, ed. by Sascha Felix. Tübingen, 205–232.
- Dulay, H. C., and M. K. Burt (1974) Errors and Strategies in Child Second Language Acquisition. *TESOL Quarterly* 8, 2: 129–136.
- Ellis, Rod (1997) *Second Language Acquisition*. Oxford: Oxford University Press.
- Felix, Sascha W. (1978) *Linguistische Untersuchungen zum natürlichen Zweitspracherwerb*. München: Wilhelm Fink.
- Ferguson, Charles (1971) Absence of copula and the notion of simplicity: a study of normal speech, baby talk, foreigner talk and pidgins. *Pidginization and Creolization of Languages*, ed. by D. Hymes, 141–150. Cambridge: Cambridge University Press.
- Greenberg, Yael (1998) An overt syntactic marker for genericity in Hebrew. *Events and grammar*, ed. by Susan D. Rothstein. Dordrecht: Kluwer Academic Publishers, 125–144.
- Gutierrez Morales, Salome (2005) Morphosyntactic Expressions of Possession and Existence in Sinhala. *Proceedings from the Workshop on Sinhala Linguistics (Santa Barbara Papers in Linguistics 17)*, ed. by Robert Englebretson and Carol Genetti, 20–28.
- Hawkins, Roger (2001) *Second Language Syntax: A Generative Perspective*. Oxford: Blackwell Publishers.
- Herat, Manel (2005) Be variation in Sri Lankan English. *Language Variation and Change* 17, 181–208.
- Ho, Mian Lian (1986) The Verb Phrase in Singapore English. PhD dissertation, Monash University, Melbourne.
- Ho, Mian Lian, and John Platt (1993) *Dynamics of a Contact Continuum*. Oxford: Oxford University Press.
- Holm, John (1976) Copula variability on the Afro-American continuum. In *Conference preprints, first annual meeting of the Society for Caribbean Linguistics*, compiled by George Cave. Turkeyen: University of Guyana.
- Holm, John (1984) Variability of the copula in Black English and its creole kin. *American Speech* 59:291–309.
- Hymes, Dell (1971) Preface. *Pidginization and Creolization of Languages*, ed. by D. Hymes. New York: Cambridge University Press.
- Klein, Wolfgang, and Clive Perdue (1997) The basic variety. *Second Language Research* 13: 301–347.
- Kovac (Lucas), Ceil (1980) Children's acquisition of variable features. PhD dissertation, Georgetown University.
- Krashen, S. (1982) *Principles and practices of second language acquisition*. New York: Pergamon
- Kratzer, Angelika (1995) Stage Level and Individual Level Predicates. *The Generic Book*, ed. by G. Carlson & F.J. Pelletier. Chicago: University of Chicago Press.
- Krishnamurti, Bh. (1998) Telegu. *The Dravidian languages*, ed. by Sanford B. Steever. New York: Routledge.

- Labov, William (1969) Contraction, deletion, and inherent variability of the English copula. *Language* 45: 715–762.
- Labov, William (1972) *Language in the Inner City: Studies in the Black English Vernacular*. Philadelphia, PA: University of Pennsylvania Press.
- Labov, William (1995) The case of the missing copula: the interpretation of zeros in African American English. *Language. An Invitation to Cognitive Science*, Vol. 1 (second edition), ed. by L. Gleitman and M. Liberman, 25–54.
- Lefebvre, Claire (1998) *Creole Genesis and the Acquisition of Grammar*. New York: Cambridge University Press.
- Le Page, Robert, and Andre Tabouret-Keller (1985) *Acts of Identity: Creole-based Approaches to Language and Ethnicity*. Cambridge: Cambridge University Press.
- Lyons, John (1968) *Introduction to theoretical linguistics*. Cambridge: Cambridge University Press.
- Masica, Colin P. (1991) *The Indo-Aryan Languages*. Cambridge: Cambridge University Press.
- Mather, Patrick-André (2000) Creole Genesis: Evidence from West African L2 French. *Languages in Contact*, ed. by D.G. Gilbers, J. Nerbonne and J. Schaeken, 247–261. Amsterdam: Rodopi.
- McWhorter, John (2000) Strange bedfellows: Recovering the origins of Black English. *Diachronica* XVII:2, 389–432.
- McWhorter, John (2001) The world's simplest grammars are creole grammars. *Linguistic Typology* 5: 125–166.
- Meillet, Antoine (1906) La phrase nominale en indo-européen. In *Mémoires de la Société de Linguistique de Paris* XIV, p. 1–26.
- Meisel, Jürgen (1983) Strategies of second language acquisition: more than one kind of simplification. *Pidginization and Creolization as Language Acquisition*, ed. by R.W. Andersen, 12–57. Rowley, MA: Newbury House.
- Mesthrie, Rajend (1992) *English in Language Shift*. Cambridge: Cambridge University Press.
- Migge, Bettina, and Norval Smith (2007) Introduction: Substrate influence in creole formation. *Journal of Pidgin and Creole Languages* 22:1, 1–15.
- Miller, M.B. (1995) Coefficient alpha: A basic introduction from the perspectives of classical test theory and structural equation modeling. In *Structural Equation Modeling* 2(3), 255–273.
- Mufwene, Salikoko (1992) Ideology and facts on African American Vernacular English. *Pragmatics* 2(2)141–66.
- Mufwene, Salikoko (2000) Some Sociohistorical Inferences about the Development of African American English. *The English History of African American English*, ed. by Shana Poplack. Oxford: Blackwell.
- Mühlhäusler, Peter (1977) *Pidginization and Simplification of Languages*. (Pacific Linguistics 26). Canberra: Australian National University.
- Nicholas, Howard (1984) 'To Be' or Not 'To Be': Is That Really the Question? Developmental Sequences and the Role of the Copula in the Acquisition of German as a Second Language. *Second Languages: A Cross-Linguistic Perspective*, ed. by R. Andersen. Rowley, MA: Newbury House.
- Pienemann, Manfred (1981) *Der Zweitspracherwerb ausländischer Arbeiterkinder*. Bonn: Bouvier.
- Pienemann, Manfred (1998) *Language Processing and Second Language Development: Processability Theory*. Amsterdam: John Benjamins.

- Pillai, N. Nadaraja (1992) *A syntactic study of Tamil verbs*. Mysore: Central Institute of Indian Languages.
- Pit Corder, S. and Roulet, F. (eds.) (1977) *The notions of simplification, interlanguage and pidgins and their relation to second language pedagogy*. Droz, Geneva.
- Platt, John (1979) Variation and implicational relationships: Copula realization in Singapore English. In *General Linguistics* 19.1:1–14.
- Poplack, Shana. (ed.) (2000) *The English History of African American English*. Oxford: Blackwell.
- Poplack, Shana, and David Sankoff (1987) The Philadelphia story in the Spanish Caribbean. *American Speech* 62: 291–314.
- Pullum, Geoffrey K. (1997) Language that dare not speak its name, *Nature*, vol. 386, March 1997.
- Rasinger, Sebastian (2005) *Adult Second Language Acquisition in Immigrant Communities: A Study on Bengali-English in East London*. PhD dissertation, Sussex University.
- Rickford, John R. (1977) The question of prior creolization in black English. *Pidgin and creole linguistics*, ed. by Albert Valdman, 190–221. Bloomington: Indiana University Press.
- Rickford, John R. (1992) The creole residue in Barbados. *Old English and New: Studies in Honor of Frederic G. Cassidy*, ed. by Joan H. Hall, Nick Doane, and Dick Ringer, 183–201. New York and London: Garland.
- Rickford, John R. (1996) Copula variability in Jamaican Creole and African American Vernacular English: A reanalysis of DeCamp's texts. *Towards a Social Science of Language Vol. 1: Variation and Change in Language and Society*, ed. by Gregory R. Guy, Crawford Feagin, Deborah Schiffrin, and John Baugh, 357–372. Amsterdam/Philadelphia: John Benjamins.
- Rickford, John R. (1997) Prior creolization of AAVE? Sociohistorical and textual evidence from the 17th and 18th centuries. *Journal of Sociolinguistics* 1, 315–336.
- Rickford, John R. (1998) The Creole Origins of African-American Vernacular English: Evidence from copula absence. *African-American English: Structure, History, Use*, ed. by Salikoko Mufwene, John R. Rickford, Guy Bailey, & John Baugh. London: Routledge.
- Rickford, John R. (1999) Variation in the Jamaican Creole copula and its relation to the genesis of AAVE: New data and analysis. *Creole genesis, attitudes and discourse*, ed. by John R. Rickford and Suzanne Romaine, 143–56. Amsterdam/Philadelphia: John Benjamins.
- Rickford, John R. (2006) Down for the count? The Creole Origins Hypothesis of AAVE at the hands of the Ottawa Circle, and their supporters. *Journal of Pidgin and Creole Languages* 21: 97–155.
- Rickford, John R., Arnetha Ball, Renee Blake, Raina Jackson, and Nomi Martin (1991) Rap-pin on the copula coffin: Theoretical and methodological issues in the analysis of copula variation in African American Vernacular English. *Language Variation and Change* 3:1, 103–132.
- Rickford, John R., and Renee Blake (1990) Copula contraction and absence in Barbadian English, Samaná English and Vernacular Black English. *Proceedings of the Sixteenth Annual Meeting of the Berkeley Linguistics Society*, 16–19 February 1990, ed. by Kira Hall, Jean-Pierre Koenig, Michael Meacham, Sondra Reinman, and Laurel A. Sutton, 257–268. Berkeley, CA: Berkeley Linguistics Society.
- Rothstein, Susan (2001) *Predicates and their Subjects*. Dordrecht: Kluwer.
- Safarewicz, Jan (1974) The Syntactic Function of the Copula. *Linguistic Studies*, ed. by J. Safarewicz. The Hague: Mouton.

- Schiffman, Harold (1999) *A Reference Grammar of Spoken Tamil*. Cambridge: Cambridge University Press.
- Schumann, J. H. (1974) Implications of Pidginization and Creolization for the Study of Adult Second Language Acquisition. *New Frontiers in Second Language Learning*, ed. by J. H. Schumann and N. Stenson. Rowley, MA: Newbury House.
- Selinker, Larry (1992) *Rediscovering Interlanguage*. London: Longman.
- Shapira, Rina (1978) The Non-learning of English: Case study of an adult. *Second Language Acquisition: A Book of Readings*, ed. by Evelyn M. Hatch, 246–256. Rowley, MA: Newbury House.
- Siegel, Jeff (2003) Substrate influence in creoles and the role of transfer in second language acquisition. *Studies in Second Language Acquisition* 25 (2):185–209.
- Siegel, Sidney, and N. John Castellan, Jr. (1988) *Nonparametric Statistics for the Behavioral Sciences* (second edition). New York: McGraw-Hill.
- Stassen, Leon (1994) Typology versus mythology: the case of the zero-copula. *Nordic Journal of Linguistics* 17, 105–126.
- Stauble, Ann-Marie (1984) A comparison of a Spanish-English and a Japanese-English second language continuum: negation and verb morphology. *Second languages: a cross-linguistic perspective*, ed. by R. Andersen. Rowley, MA: Newbury House.
- Sweetland, Julie, John Rickford, and Joy Hsu (2000) Prosodic Conditioning of the Copula: A Second Opinion. Presented at *New Ways of Analyzing Variation (NWA)* 29. October 2000. Michigan State University, East Lansing, Michigan.
- Thomas, Margaret (2002) The specious battle between ‘contrastive analysis’ and ‘creative construction’. Paper in the *NAAHoLS Session of the Linguistic Society of America (LSA) Conference*, San Francisco.
- Trudgill, Peter (2004) Linguistic and social typology: The Austronesian migrations and phoneme inventories. *Linguistic Typology* 8: 305–320.
- Veenstra, Tonjes (2003) What verbal morphology can tell us about creole genesis: The case of French-related creoles. *The Phonology and Morphology of Creole Languages*, ed. by Ingo Plag and Sabine Lappe. Tübingen: Niemeyer.
- Veenstra, Tonjes (2007) Creoles as beyond the Basic Variety. Paper presented at the Society for Pidgin and Creole Linguistics meeting, Linguistic Society of America Annual Conference, January 2007, Anaheim, California.
- Véronique, Daniel (1994) Naturalistic adult acquisition of French as L2 and French-based creole genesis compared: insights into creolization and language change? *Creolization and language change*, ed. by Dany Adone and Ingo Plag. Tübingen: Max Niemeyer Verlag.
- Walker, James (2000) Rephrasing the Copula: Contraction and Zero in Early African American English. *The English History of African American English*, ed. by Shana Poplack. Malden and Oxford: Blackwell.
- Whinnom, Keith (1971) Linguistic hybridization and the ‘special case’ of pidgins and creoles. *Pidginization and Creolization of Languages*, ed. by Dell Hymes. New York: Cambridge University Press.
- Winford, Donald (1992) Another look at the copula in Black English and Caribbean creoles. *American Speech* 67(1): 21–60.
- Winford, Donald (1997) On the origins of African American Vernacular English: A creolist perspective. Part 1: The sociohistorical background. *Diachronica* 14 (2): 305–44.
- Winford, Donald (1998) On the Origins of African American Vernacular English: A Creolist Perspective. Part II: Linguistic Features. In *Diachronica* 15 (1): 99–155.

- Winford, Donald (2003) *An Introduction to Contact Linguistics*. Oxford: Blackwell.
- Winford, Donald (2004) Creole Formation and Second Language Acquisition. Ms. available at: http://www.ling.ohio-state.edu/~dwinford/recent_papers.htm
- Wolfram, Walt (1969) *A sociolinguistic description of Detroit Negro speech*. Washington DC: Center for Applied Linguistics.
- Wolfram, Walt (2000) Issues in reconstructing Earlier African American English. *World Englishes* 19 (1): 39–58.
- Wyatt, Toya (1996) Acquisition of the African American English Copula. *Communication Development and Disorders in African American Children: Research, Assessment, and Intervention*, ed. by A. Kamhi, K. Pollock, and J. Harris. Baltimore, Maryland: Paul H. Brooks Publishing Co.