# On the Boundaries of Linguistic Competence: Matched-Guise Experiments as Evidence of Knowledge of Grammar<sup>\*</sup>

Emily M. Bender Stanford University

March 17, 2004

# 1 Introduction

According to the standard definition, *linguistic competence*, the object of study of generative grammar, is knowledge of language. As it is used in generative linguistics, however, *knowledge of language* is also a technical term, referring to some system of rules that determines the set of sentences in a language. As such, any linguistic knowledge that does not bear on the grammaticality or linguistic meaning of sentences is necessarily not a part of the systems that grammars are meant to model.

The widespread adoption of this conception of the object of study appears to be partly due to historical accident. The notion of a language as a set of sentences comes from Formal Language Theory (developed by Chomsky and others). In 1955, Chomsky applied this conception of language to the study of natural languages and laid out a research program of building grammars that generate the appropriate strings. In this early work, the cognitivist position is not evident. Unconcerned with issues of linguistic knowledge, Chomsky was free to define the object of study in whatever way was convenient. Later, cognitivist theories of linguistic competence (including prominently Chomsky's own) inherited and incorporated this definition of the object modeled by a grammar: they set out to describe the internalized systems that determine which sentences speakers will accept as grammatical.

<sup>\*</sup>This paper is based on Chapter 4 of Bender 2001, and further details of methodology can be found there. A preliminary version of this work appeared as Bender 2000. I would like to thank Tom Wasow, Penny Eckert, Ivan Sag, John Rickford, Arnold Zwicky, Stacy Fambro, Andrea Kortenhoven, Claude Steele and Kathryn Campbell-Kibler for helpful discussion and encouragement; two anonymous reviewers for *Lingua* and the audience at the Penn Linguistics Colloquium for useful feedback; Jennifer Iljas and Mark Thomas's class; Kristofer Jennings for help with statistical analysis; and the Stanford Humanities and Sciences Graduate Alumni Association for financial support in the form a Dissertation Fellowship. All errors remain my own.

However, theories concerned with actual speakers' knowledge of language do not have the freedom to arbitrarily define the object of study. To my knowledge, there is no evidence that speakers possess a separate module of knowledge limited to that which is usually included in competence grammar.

The aim of this paper is to illustrate one particular methodology for discovering whether speakers have knowledge of another kind of pattern in language: the soft or non-categorical grammatical constraints which are evident in the distribution of sociolinguistic variables. I leave to future work the question of the relationship of this kind of knowledge to that which bears on grammaticality (but see Bender 2001, Chapter 6 for some discussion).

The remainder of this paper is structured as follows: Section 2 provides some background on the sociolinguistic variable that will serve as a case study. Section 3 describes the methodology as it was applied to this case study. The results of the experiment were suggestive, but not conclusive, due to the relatively small sample size and other factors. These results are presented in Section 4. Section 5 discusses ways in which the methodology could be improved as well as possible extensions.

# 2 Variable Copula Absence in AAVE

The particular sociolinguistic variable studied here is variable copula absence in African American Vernacular English (AAVE). In AAVE, in addition to sentences with the full (1a) and contracted (1b) forms of the copula, sentences such as (1c), with no (overt) copula are also grammatical.

- (1) a. She is my piano teacher.
  - b. She's my piano teacher.
  - c. She my piano teacher.

While the variants in (1) are extremely similar (if not quite identical) in their semantics, the choice between copula presence and copula absence is constrained by both social and grammatical factors. Because this variable has been studied so extensively (see e.g., Labov 1969, Wolfram 1969, Baugh 1979, Bailey and Maynor 1987, Rickford et al. 1991), we know that some of the non-categorical constraints it is subject to are consistent across many different speech communities.

The single most widely studied non-categorical grammatical constraint on copula absence, and the one that is the focus of the experiment, is the effect of the following grammatical environment. Table 1, excerpted from Rickford's Table 6.16 (1998:190), provides a summary of nine data sets from studies of copula absence in various AAVE speaking communities. The left half of the table gives information about the data set: whether it concerns *is*, *are*, or both; the population studied; and the author. The right hand side of the table gives information about the effect of following grammatical environment on copula absence. With the exception of Wolfram's study (#4), the values are Varbrul weights, with copula absence as the distinguished variant.<sup>1</sup> Wolfram's results were reported as raw frequency data, given in percentages. The five environments considered are noun phrase (\_\_NP), locative (\_\_Loc), adjective (\_\_Adj), -in(g) form verb (\_\_V+ing), and gon(na) (\_\_gon). These environments are arranged according to the ranking in (2), where environments to the left are the least favorable to copula absence and the environments to the right are the most favorable. The consistency of the pattern across the different studies is striking, despite some disagreement as to the ranking of following locatives (\_\_Loc) and adjectives (\_\_Adj). (3) to (7) illustrate each of the environments with examples from Wolfram 1969.

		Studies		E	Environm	ients		
#	Form	Place	Source	NP	_Loc	Adj	V+ing	gon
1	is	NYC	Labov 1969	.2	.36	.48	.66	.88
		Thunderbirds						
2	is	NYC Jets	Labov 1969	.32	.52	.36	.74	.93
3	is	NYC Cobras	Baugh 1979	.14	.31	.72	.59	.78
4	is+are	Detroit WC	Wolfram 1969	37%	44%	47%	50%	79%
5	is	LA	Baugh 1979	.32	.29	.56	.66	.69
6	are	LA	Baugh 1979	.25	.69	.35	.62	.64
7	is+are	Texas kids	Bailey &	.12	.19	.25	.41	.89
			Maynor 1987					
8	is+are	Texas adults	Bailey &	.09	.15	.14	.73	.68
			Maynor 1987					
9	is+are	East Palo	Rickford	.29	.42	.47	.66	.77
		Alto	et al. 1991					

Table 1: Copula absence in AAVE in different communities, by following grammatical environment. (Values given are Varbrul factor weights, with the exception of the data from Wolfram (1969), which is given in percentages.)

- (2) NP < Loc < Adj < V+ing < gon
- (3)  $\_$ NP: She a nurse.
- (4) <u>Loc</u>: They out there in space.
- (5) <u>Adj</u>: She real nice.
- (6)  $\_V+ing$ : Do anything if you fighting.
- (7) \_\_\_\_\_gon: I really don't think John gonna make it.

<sup>&</sup>lt;sup>1</sup>On Varbrul, see Cedergren (1973), Sankoff (1975), Rousseau and Sankoff (1978) and Guy (1988).

Within each of these studies, the pattern found was argued to be robust (statistically significant)—i.e., something more than an accident of which data happened to be recorded. That (roughly) the same pattern is found across many communities is a strong confirmation. This pattern of non-categorical grammatical constraints on copula absence appears to be a property of AAVE, one that must either follow from something else in the language, follow from functional constraints on performance, or be learned, for there is no other way for it to be reproduced in community after community.<sup>2</sup>

Notice that the question of whether speakers have knowledge of this pattern is (somewhat) independent of the question of what causes the pattern. That is, the existence of a functional explanation for the pattern does not preclude speakers learning it as a pattern. On the other hand, it is possible that this particular pattern of non-categorical constraints is a purely arbitrary fact about AAVE. In this case, speakers must have some representation of the pattern, otherwise it would not be reproduced so consistently. Here I will leave aside the question of the origin of the pattern and focus on whether speakers can be shown to have knowledge of it.

# 3 Methodology

If speakers have knowledge of non-categorical constraints on variation, it is almost certainly tacit and inaccessible to introspection. Fortunately, Labov's (1963) finding that sociolinguistic variation is socially meaningful provides a jumping off point for constructing an experiment.<sup>3</sup> If sociolinguistic variation is socially meaningful, then the social value of variants might interact with the non-categorical constraints. In particular, Wolfram makes the following observation concerning the data in Table 2 (from Wolfram 1969:172):<sup>4</sup>

The relatively high frequency with which zero realization is found preceding intentional future *gonna* among middle-class informants suggests that zero realization preceding *gonna* is less stigmatized than zero realization in other environments. (1969:172–173)

Generalizing to allow for social values other than stigmatization, I propose the following two-part hypothesis:

<sup>&</sup>lt;sup>2</sup>Of course, just because an effect can be found by considering the following grammatical environment doesn't mean that that's really what's going on. It could be that the correlation is between semantic properties and rates of copula absence, and that the strength of this effect together with a reasonably close pairing of syntax and semantics are enough to let the effect shine through even though we've been counting the 'wrong' thing. However, semantic constraints would still be grammatical constraints. Further, it seems unlikely that a reanalysis of the production data in terms of semantic categories would turn up only categorical constraints, given minimal sets such as in (1). Syntactic or semantic, the pattern shown in Table 1 must either be a part of the grammar of AAVE or follow from something else in the grammar or in performance.

<sup>&</sup>lt;sup>3</sup>See also Eckert 2000 and the papers in Eckert and Rickford 2001.

<sup>&</sup>lt;sup>4</sup>See also Sylva and Zwicky's (1975) observation that the stylistically marked syntactic rules they discuss can be more or less stylistically marked depending on the lexical items they interact with. For example, they observe that existential *there* is relatively 'formal', but only with verbs other than *be*.

	Following grammatical environment					
Speaker group	Adj	NP	Loc	V+ing	gon	
Middle class	1.6%	4.2%	13.3%	11.3%	33.3%	
Working class	36.5%	47.3%	44.4%	50.0%	78.9%	

Table 2: Wolfram's copula absence by following grammatical environment and social class

- I Copula absence/presence in AAVE is associated with some social value.
- II Copula absence/presence in AAVE is more strongly associated with that social value the more marked the environment is for each variant.

For example, if copula absence sounds confident, then copula absence before a noun should sound especially confident and copula absence before a verb somewhat less so.

Note that Part II of this hypothesis entails that speakers have knowledge of non-categorical constraints. However, what is at issue here is more than just the knowledge of the constraint. Previous formal approaches to constraints on variation (e.g., Labov 1969, Anttila 1997, and Boersma and Hayes 1999) if they consider social constraints at all, treat them as separate from grammatical constraints. The hypothesis being tested here is that the two kinds of constraints interact. That is, social constraints are conceptualized as the social meaning of the variable, and grammatical constraints as the intensifying or attenuating effect of the grammatical environment on the social meaning or social value of the variable.<sup>5</sup>

The experimental design was based on the matched-guise methodology of Lambert et al. 1975.<sup>6</sup> In order to test for knowledge of non-categorical constraints on copula absence, I extended the matched-guise methodology to test the social evaluation of an individual linguistic feature, as described in the following subsection.

### 3.1 Stimuli

#### 3.1.1 Test sentences

As the purpose of this study was to test the social evaluation of a particular linguistic feature (copula absence) and the effect of the grammatical context on that evaluation, the speech samples had to be much shorter than the 20-second-long samples used in Lambert et al.'s study. This is because the longer the speech sample, the more likely the introduction of confounding variables. Similarly, the sentences were designed to avoid AAVE- and SAE-specific features as much as possible. Even if held constant across all of these stimuli, a feature other than copula presence/absence that is strongly stereotyped could have flooded the effect of the variable under consideration. Nonetheless, since I was looking at the evaluation of copula absence within the system of AAVE, it was important to choose sentences that are a

<sup>&</sup>lt;sup>5</sup>For some discussion of how this might be formalized, see Bender 2001, Chapter 6.

<sup>&</sup>lt;sup>6</sup>See also Lambert 1967, Giles 1971, Anisfeld and Lambert 1964, Lambert et al. 1966, S. Lambert 1973, and Lambert and Tucker 1975.

part of AAVE, and to make sure that their actual production conformed to AAVE phonology, etc. Further, in testing for the effect of the non-categorical constraint in question, it was important to control for all other known non-categorical constraints.

The non-categorical constraint to be tested is the effect of the following grammatical environment, where the environments are ordered from least to most favoring of copula absence as in (8).

$$(8) \_NP < \_Loc < \_Adj < \_V+ing < \_gon$$

Ideally, one would want to test all of the environments to see if they are ordered in perception as they are in production. However, in order to keep the experimental task reasonably short, only one pair was tested. *Gon* is somewhat problematic because it is itself an AAVE-specific feature and the high rates of copula absence before *gon* suggest that it may be in the process of being reanalyzed as a modal.<sup>7</sup> Leaving out *gon*, the V+ing and NP environments allow for the greatest contrast.

The test sentences, designed to meet all of these considerations, are as follows:

- (9) a. Yeah I know her. She's teachin me piano at Music World.
  - b. Yeah I know her. She's my piano teacher at Music World.
  - c. Yeah I know her. She teachin me piano at Music World.
  - d. Yeah I know her. She my piano teacher at Music World.

Sentences (9a) and (9b) represent the copula presence condition, while (9c) and (9d) represent the copula absence condition. (9a) and (9c) represent the V+ing condition, and (9b) and (9d) represent the NP condition.

The sentences are about as close as they can be in phonological and semantic content given the requirement of setting up the different conditions, and are constant with respect to all other known constraints: the subject is always a pronoun, and always phonologically identical. There is no known effect of following stop vs. nasal.

The initial and final phrases (Yeah I know her and at Music World) are included to give the sentences some length.

In addition to the test sentences, the following sentences were also recorded as filler stimuli:

- (10) a. Yeah I know her. She useta teach me piano at Music World.
  - b. Yeah I know her. She useta be my piano teacher at Music World.
  - c. Yeah I know her. She taught me piano at Music World.

<sup>&</sup>lt;sup>7</sup>Although not one with all of the properties of a true auxiliary: for example, it doesn't invert (Salikoko Mufwene, p.c., Jan., 2000).

d. Yeah I know her. She was my piano teacher at Music World.

The filler sentences match the test sentences fairly closely in their semantics, but not as closely as the test sentences match each other. Likewise, there is more variation in phonological content.

The talkers<sup>8</sup> that I recruited to record the stimuli confirmed that all of these sentences would sound natural in the middle of an AAVE conversation.

Eight talkers were recruited from the Stanford community to record the sentences. All were (at least) bidialectal, commanding both AAVE and SAE.<sup>9</sup> All were female. Each talker was assigned to either test or filler stimuli. The recordings range from 2.482 seconds to 2.698 seconds in length for test stimuli (mean = 2.592 seconds) and from 2.36 to 3.208 seconds for filler stimuli (mean = 2.711 seconds). These times include 0.1 seconds of silence at the beginning and end of each recording. This study did not follow Lambert et al.'s strategy of lengthening the stimuli by repeating them on the grounds that this would give listeners more of a chance to consciously identify the variation in the realization of the copula.<sup>10</sup>

#### 3.1.2 Test scales

Lambert et al. were researching the language attitudes of members of the community they studied, and indirectly the ethnic identification of people growing up in a bicultural environment. Accordingly, they used the scales to assess the stigma or prestige associated with each language variety represented in their recordings. Since I was focusing on one feature (copula absence) there was the possibility that it has some more specific social value, akin to the way ing/in is associated with formality/informality (Fischer 1958), and that one of the scales chosen would be close enough to this social value to reflect it. This argues for including as many scales as possible. On the other hand, too many scales would end up making the experiment too long and could also lead to the last several scales being presented too long after the listener heard the stimulus. In the end, I settled on the seven scales listed in Table 3.

The choice of scales was partially informed by the fact that AAVE is a stigmatized variety. It seemed fairly likely *a priori* that listeners might associate a particular feature of AAVE (such as copula absence) with lack of education or employability. At the same time, varieties that are globally stigmatized may nonetheless have positive value for their speakers (cf. Trudgill's (1974) notion of covert prestige), and indeed individual speakers may recognize global stigmatization and local value simultaneously. If copula absence is

<sup>&</sup>lt;sup>8</sup>To avoid ambiguity, I will use the term 'talker' for the speakers who made the recordings in a matchedguise experiment because the term 'speaker' is used to refer to a person with competence in a given language, regardless of what s/he is using the competence to do in the situation being considered. The participants who listened to the recorded stimuli will be referred to as 'listeners' in general, but also as 'speakers' when it is important to highlight their competence of a particular variety.

<sup>&</sup>lt;sup>9</sup>I did not systematically collect geographical data on the talkers. Although all were currently living in California, some at least were from other regions of the country.

<sup>&</sup>lt;sup>10</sup>For further details about the recordings, including strategies for eliciting natural-sounding tokens, see Bender 2001, Chapter 4.

ţ.
educated
ob

Table 3: Scales used for the experiment

used to index solidarity or in-groupness, for example, in-group listeners may rate talkers who use copula absence as 'likeable' or 'reliable' (i.e., someone they feel solidarity with). 'Confidence' could reasonably be expected to go either way: Speakers who defiantly use stigmatized variants could be taken as confident. Alternatively, listeners may assume that speakers of a stigmatized variety have internalized some of that stigma and therefore lack confidence, especially about language. Finally, 'comical' and 'polite' were included to round out the space of characteristics being considered.

Note that these scales correspond to traits that are not necessarily correlated. For example, someone could sound confident without sounding well-educated. Furthermore, the scales differ in whether they relate to relatively long-term characteristics (education, employability) or to characteristics that might vary from moment to moment (politeness, confidence). The framing of each scale as relating to long-term or short-term properties of the talkers was also indicated by the question used to present each scale: "How likeable/reliable/well educated does this person sound?" "How polite/comical is this person being?" "How confident do you think this person is feeling?" and "How good of a job do you think this person has?"<sup>11</sup>

### 3.2 Participants

Participants were recruited from a introductory psychology course at a community college in a community in California with a large African-American population. Participants earned extra credit in their course for their participation and were also paid \$5.

Since participants could earn extra credit, the experiment had to be open to anyone in the class. The people who chose to participate fell into the five ethnically and linguistically defined groups given in Table 4. These groups will form the basis of the analysis below, and will be referred to by the numbers assigned to them in Table 4.

Familiarity with AAVE was determined in an exit interview as described in §3.3 below. Whether or not a person spoke (any variety of) English natively was determined on the basis of the demographic questionnaire. Any listener who grew up in a non-English-speaking country was counted as non-native. Note that membership in Group III only requires familiarity with AAVE and not native-speaker status in English. In fact, one listener in Group

<sup>&</sup>lt;sup>11</sup>The form of this last question is not grammatical for all speakers. Whether it was grammatical for all of the speakers in this study and, if not, whether it would have had any effect on the results, is unclear.

Ι	AAVE speakers (all African American)	N = 11
II	African Americans who do not identify as AAVE speakers, but who are	N = 4
	familiar with AAVE	
III	participants who are familiar with AAVE but are not African American	N = 7
IV	native speakers of English with no familiarity with AAVE	N = 6
V	non-native speakers of English with no familiarity with AAVE	N = 7

#### Table 4: Groups of listeners

III grew up in the Philippines and another in Haiti.<sup>12</sup> Groups II and III are distinguished on the basis that the African Americans' experience with AAVE is substantially different from the others'. Group III listeners reported that they were familiar with AAVE because they went to predominantly African American high schools and/or participate in Hip Hop culture.

Table 5 summarizes the information on age and gender by group. Table 6 gives the ethnic self-identification of the participants as well as information on where they grew up.<sup>13</sup>

Group	age range	median age	# men	# women
Ι	19 - 55	26	3	8
II	24 - 33	27	1	3
III	18 - 27	22	2	5
IV	19 - 40	25	2	4
V	18-34	21	2	5

Table 5: Age and gender information, by group

### **3.3** Running of the experiment

The experiment took place in an instructor's office at the community college, with one to two participants at a time (there were two computers). Once the demographic questionnaire was completed, the participants donned headphones attached to the computers they would use. The stimuli were presented to the subjects by the program PsyScope (Cohen et al. 1993).

Each trial began with the message "loading speaker" being displayed for 0.5 seconds to give the participants a signal that they should get ready to listen. Then the screen went blank and one of the 32 test or filler stimuli was played over the headphones. The stimuli

<sup>&</sup>lt;sup>12</sup>In Bender 2001, this speaker was actually included in Group II. Moving him to Group III did not make an interesting difference in the overall results.

<sup>&</sup>lt;sup>13</sup>Some of the African American participants gave their ethnicity as 'Black' and some as 'African American'. These responses have been regularized to 'Black' in Table 6. The speaker from Nigeria self-identified as African American.

Group	ethnic self-identification	home state/country <sup><math>a</math></sup>	Ν
Ι	African American	California	5
	$(didn't ask)^b$	California	4
	African American	Florida	1
	African American	Louisiana, California	1
II	African American	California	2
	African American	Alabama	1
	Mixed	California	1
III	White	California	1
	Jewish	Texas	1
	Filipino	Philippines, California	1
	White	California	1
	Caucasian	Ohio, New Mexico, California	1
	Asian	Illinois, California	1
	Black	Haiti	1
IV	(didn't ask)	California	2
	Indian	US Virgin Islands, British Virgin Islands,	1
		California	
	Latina	California	1
	White	California	1
	Mexican American	California	1
V	Chinese	China, California	3
	Asian Pacific	Korea	1
	Asian	China	1
	Vietnamese	Vietnam, California	1
	African American	Nigeria	1

 $^a\mathrm{Where}$  the list ener lived between the ages of 2 and 18

 $^b{\rm I}$  had originally expected to recruit only African American participants. The question asking about ethnic identification was added after the first session.

Table 6:	Ethnicity	and	home	state,	/country,	by	group
	•				• • •	•/	<u> </u>

were presented in a near-random order. (The order was near-random rather than random in order to minimize the probability that stimuli from the same talker would be presented one after the other.) The order of presentation was determined separately for each listener.

Each recording was followed by the seven scales. The scales were presented in a random order, with the position of the positive and negative poles selected randomly (i.e., sometimes the negative adjective was on the left, and sometimes it was on the right). The presentation of the scales, together with the contextualizing question, was as in Figure 1.



Figure 1: Presentation of scales

The question and scale remained on the screen until the participant put the mouse on or near one of the seven bullets and clicked, or until the question timed out at 20 seconds. The timeout was included as a mechanism for allowing participants to pass on individual questions, as specified in the consent form. A click on or near the bullet closest to the negative end of the scale (e.g., 'not a good job') was coded automatically by PsyScope as 1, the next one over as 2, etc. After all seven scales were presented in this fashion, the next trial began with the "Loading speaker..." message.

The last part of the experiment to be presented on the computer was a language attitudes questionnaire, which will not be discussed further here. When the computer-based portion was complete, there was an exit interview in which I informed the participants of the purpose of the study and determined their level of familiarity with AAVE. Whenever possible, the exit interview was done with two participants at a time. The participants were shown a list of the eight sentences (test and filler) and the exit interview proceeded roughly according to the following script:

These are the sentences that the people were saying. How many different speakers did you think there were? ... Actually, there were eight speakers. Four said each of these sentences [the test sentences] and four said each of these sentences [the filler sentences]. These sentences [the test sentences] are the ones I'm interested in. The rest are filler sentences. As you can see, these two have the verb *is* in them and these two don't. In AAVE, both ways of saying it are

grammatical. I was interested to see if the choice between saying *is* and leaving it out would affect how people rated the speakers.

Also, in these two sentences the predicate is a verb, *teachin me piano*, while in these two it's a noun, *my piano teacher*. When you record people speaking AAVE, it turns out that they're much more likely to leave the *is* out when the predicate is a verb, like *teachin*, than when it is a noun, like *piano teacher*. So I am interested to see if that will have any effect on how people rate the speakers. My hypothesis was that if, for example, it sounds confident to leave *is* out, it should sound even more confident to leave it out in the unusual case.

Of course, this should only be true for listeners who are familiar with AAVE. Are you familiar with AAVE?

The phrasing of the last question was fortuitous. I chose to ask about familiarity rather than actual speaking because I didn't want anyone to feel like I was accusing them of speaking a non-standard variety and because in the first exit interview I had one African-American and one non-African-American participant. By asking the question this way, it was applicable to both of them. As it turns out, some African-American participants responded to this question by saying something like "Yeah, I'm familiar with it. We talk that way at home." Other African-American participants said, "I don't talk like that, but I hear it from ...." One man said he "winced" at the preacher's bad grammar at his church. All of the African-American participants were at least familiar with AAVE. Also, some of the non-African-American participants said they were familiar with AAVE, most commonly from Hip Hop culture or from going to high schools with large African-American populations. The differences between these three groups are very interesting and will be highlighted in the discussion below.

### 3.4 Missing data

Due to a variety of causes, a small number of data points (338 of 3920 ratings of a talker by a listener for a sentence on a scale, or 8.6%) are missing. Six listeners (listeners H, L, M, N, f and i) allowed one scale for one talker to time out rather than answering it. A seventh (S) allowed two scales to time out. Due to a technical difficulty with one of the computers, 10 listeners (Q, R, U, V, Y, Z, d, e, and g) were not presented with the stimuli from one test talker. Two listeners (O and T) stopped the experiment part-way through, one because the computer crashed and the other because she decided she'd had enough. In all cases, the data that was available was included in the analysis.

## 4 Results

This section presents the results of the experiment. For reference, the hypotheses are repeated here:

I Copula absence/presence in AAVE is associated with some social value.

II Copula absence/presence in AAVE is more strongly associated with that social value the more marked the environment is for each variant.

The results of the judgment task show that most listeners associated copula absence with some social value. That is, for some scales, their responses were sensitive to the presence or absence of the copula. For the two African-American groups (Groups I and II), there is a further effect of the grammatical environment, such that marked cases were rated more to the extremes of the scales. This effect is not present for Groups III-V.

### 4.1 Copula presence vs. copula absence

This section presents the results relevant to Part I of the hypothesis, that copula absence/presence in AAVE is associated with some social value. The sentences are repeated here in (11). In the labels for the sentences, P stands for copula presence, A for copula absence, V for following verb and N for following noun.

- (11) a. AN: Yeah I know her. She my piano teacher at Music World.
  - b. AV: Yeah I know her. She teachin me piano at Music World.
  - c. PN: Yeah I know her. She's my piano teacher at Music World.
  - d. PV: Yeah I know her. She's teachin me piano at Music World.

Figure 2 gives an overview of the results, showing the average rating of each sentence (from 1 to 7) on each scale across all listeners and talkers. With the exception of 'comical', all of the scales show a clear division between the absence cases (AN and AV) and the presence cases (PN and PV). Furthermore, all of the differences go in the same direction: on average, the listeners found that the talkers sounded *more* confident, *better* educated, like they had *better* jobs, *more* likeable, *more* polite, and *more* reliable in the copula presence condition than in the copula absence condition.

A two-tailed Wilcoxon signed rank test<sup>14</sup> shows the presence v. absence contrasts on all of the scales except 'comical' to be significant (p < 0.0001). The result for 'comical' is not significant, even for p < 0.05.

These results shows that six of the scales were relevant to the social value (or 'social meaning') of copula absence/presence for the group as a whole. However, given the diversity in linguistic experience of these speakers (see Table 4), it's not reasonable to expect that they all share the same evaluation of copula absence. As shown in Table 7, Groups I and V found fewer scales to be relevant to the meaning of copula absence than the other groups, and for Group V in particular, the effect is statistically weaker than for the other groups.

<sup>&</sup>lt;sup>14</sup>A non-parametric alternative to the student's t-test, which, unlike the t-test, does not assume that the populations being compared are distributed normally. As this test compares particular values within each population (here, the rating given to one talker saying sentence PN by a particular listener on a particular scale to the value for that same talker, listener and scale but for the sentence AN, and likewise for PV and AV), when any particular data point was missing (see §3.4), the corresponding point was thrown out as well.



Figure 2: Average ratings for all listeners

Group	comical	confident	educated	job	likeable	polite	reliable
Ι	n.s.	n.s.	$6.7 \mathrm{x} 10^{-7}$	$2.6 \mathrm{x} 10^{-5}$	0.025	n.s.	0.008
II	n.s.	0.015	$1.0 \mathrm{x} 10^{-5}$	$2.8 \mathrm{x} 10^{-5}$	0.002	0.003	0.001
III	n.s.	n.s.	$1.6 \mathrm{x} 10^{-6}$	$1.1 \text{x} 10^{-4}$	0.010	$4.9 \mathrm{x} 10^{-5}$	$5.0 \mathrm{x} 10^{-5}$
IV	n.s.	0.016	$4.0 \mathrm{x} 10^{-6}$	$6.0 \mathrm{x} 10^{-6}$	$4.5 \mathrm{x} 10^{-4}$	$8.5 \text{x} 10^{-5}$	0.001
V	n.s.	n.s.	0.016	0.036	0.005	n.s.	0.028

Table 7: P values for each scale, by group

In fact, given how little is known about the means by which speakers learn (and update) social meaning, given the artificiality of the task, and given that the scales may or may not have been directly relevant to the actual social meaning of copula absence/presence for these listeners, it wouldn't be surprising to find great individual variation, even within groups, as to which scales were deemed relevant. Furthermore, as discussed in §4.2 below, part II of the hypothesis is dependent on part I in such a way that testing it requires finding out which scales were relevant to the social meaning of copula absence/presence for each listener. Fortunately, the Wilcoxon signed rank test is sensitive enough to apply to the 8 pairs of sentences (N and V cases, across 4 talkers) judged by each listener on each scale. The results are shown in Table 8, where an 'X' in a cell indicates that the listener rated the P sentences significantly differently from the A sentences on that scale (p < 0.05). A '/' indicates that the results approached significance (0.1 ).

Table 8 reflects the results of a two-tailed Wilcoxon test, as it wasn't possible to predict in general which way the difference would go. Examination of the data shows that (consistent

with the results shown in Table 2), the P sentences are rated higher than the A sentences in almost all cases where there is a difference. The exceptions are marked with a superscripted 'a' in Table 8.

In contrast with the results of Lambert et al.'s (1975) study, there is scant evidence for a positive evaluation of copula absence among AAVE speakers (Group I). While it was to be expected that AAVE speakers should rate the talkers in the copula presence condition as sounding more educated and like they have better jobs, the same isn't necessarily true for the other scales. As mentioned above, the scale 'reliable' was included in the hopes that it would index in-groupness. It patterned with 'job' and 'educated' instead. One possible explanation for this is that the participants interpreted 'reliable' as having to do with corporate reliability—i.e., reliable in the eyes of one's employer. However, these data alone do not support any firm conclusions. One might also expect listeners to score the more vernacular sentences higher on the scales 'likeable' and 'confident', but this didn't happen.

It is possible that these results reflect 'linguistic insecurity' (Labov 1966), the stigmatization of AAVE features by AAVE speakers. However, it is also possible that they reflect the influence of the experimental environment. The experiment was carried out at the community college the participants were attending, by a white experimenter and with a computer. The listeners in Group I may have been evaluating the sentences they heard with respect to the norms of the wider community. If this is the case, it might also explain why the results are weaker for Group I listeners than for Group II listeners, who presumably would not have felt such a conflict between their own evaluation of the talkers and the evaluation they expected from the wider community.

The listeners in Group II all indicated that they were familiar with AAVE, but then distanced themselves from the variety. They did this by saying things like "I wince whenever my preacher splits his infinitives"<sup>15</sup> or "I'm trying to teach my daughter to speak Standard English, but she's picked up AAVE at school." It seems that these speakers may have decided at some point to use Standard English rather than AAVE, although both may have been available to them. Having made such a decision, they would probably have negative associations with AAVE features, and such features would probably be very salient for them.

The general picture that emerges from these data is as follows: Groups II-IV tended to select more scales as relevant than Groups I and (especially) V. The most commonly selected scales were 'education' and 'job'. Given the unidirectionality of the data (when there was a significant effect, it was almost always in the direction of copula presence being rated higher than copula absence, the one exception being on the scale 'comical'), it's hard to say whether any of the scales specifically accessed anyone's social meaning for this variable, or whether copula absence was just evaluated negatively, especially by Groups II-IV. The fact that Group I didn't select more scales as relevant, and especially the fact that they didn't rate the A sentences higher than the P sentences on any scales, may have been the result of a conflict between their own attitudes towards AAVE and the normative attitudes towards

<sup>&</sup>lt;sup>15</sup>Split infinitives are not an AAVE-specific feature, of course, but this comment expresses the participant's attitudes twoards grammatical 'correctness'. If he's bothered by split infinitives, he is unlikely to overlook missing copulas.

Listener	com.	conf.	ed.	job	lik.	pol.	rel.
Group I							
А		Х	Х	Х	Х	Х	Х
В			Х	Х			Х
С	Х		Х	Х			
D			Х	Х			
Е			Х				/
F							Х
G				Х			
Н							
Ι							
J							
K							
Group II	[				•		
L			Х	Х	Х	/	Х
М			Х	Х	Х		Х
Ν			Х	Х		/	
$O^b$					/		
Group II	Ι						
Р	$/^a$		Х	Х	Х	Х	Х
$\mathbf{Q}^{c}$	X <sup>a</sup>		Х	Х			Х
$\mathbf{R}^{c}$			Х	Х		Х	/
S	Х					Х	
$T^b$			/				
$\mathbf{U}^{c}$			/				
$\mathbf{V}^{c}$							
Group I	V			-			
W			Х	Х	Х	Х	Х
Х		Х	Х	Х		Х	
$\mathbf{Y}^{c}$		/	Х	Х			/
$\mathbf{Z}^{c}$			Х	Х	/		
a							
b							
Group V							
С			/	Х			Х
$d^c$	/	a					
ec					/		
f	$/^a$						
$g^c$							
h							
i							

<sup>*a*</sup>A sentences rated higher than P sentences. <sup>*b*</sup>Listener stopped part way through the task. <sup>*c*</sup>Data missing for one talker (25% of the stimuli).

Table 8: Significant6 scales, by listener

AAVE that they were sensitive to in the experimental session.

Nonetheless, the evidence presented in this section provides support for part I of the hypothesis: In the experimental task, listeners based their social evaluation of the talkers in part on the presence vs. absence of the copula. The heterogeneity in the pattern of scales selected (especially in Group I) raises intriguing further questions, but the fact that the majority of listeners selected some scales is sufficient to allow for testing of part II of the hypothesis.

### 4.2 Effect of the following grammatical environment

This section reports the results that pertain to part II of the hypothesis: that there should be an interaction between the grammatical environment and the social value of the variable. In particular, the hypothesis states that copula absence/presence in AAVE is more strongly associated with its social value the more marked the environment is for each variant (as shown by production studies). That is, part II of the hypothesis will be confirmed, if, for any scale treated as relevant by a listener, copula absence before a noun or copula presence before a verb (AN and PV, the relatively marked cases) are rated more towards the extremes of the scale than copula absence before a verb or copula presence before a noun (AV and PN, the less marked cases). Since copula presence was almost always rated higher than copula absence, when there was a difference (see  $\S4.1$  above), the hypothesis predicts the configuration in Figure 3. If a talker sounded relatively uneducated to a listener in the copula absence condition, they should have sounded especially uneducated when using copula absence before a noun. To put it yet differently, to the extent that a disfavorable social value is associated with copula absence, that disfavorable social value should be heightened in the AN sentence with respect to the AV sentence. Conversely, to the extent that a favorable social value is associated with copula presence, that favorable social value should be heightened in the PV sentence with respect to the PN sentence. Such a result would show that speakers have knowledge of the non-categorical constraint.



Figure 3: Predicted ordering of sentences on a relevant scale

This hypothesis is tested by once again applying the Wilcoxon signed rank test, this time in a one-tailed version because there is a prediction as to the direction of the difference. This test compares the ratings for N and V sentences while holding constant the talker, the listener, the scale, and P vs. A. Since part II of the hypothesis is dependent on part I (it only makes sense to look for an effect of the following grammatical environment on the social value of the variable where the social value has been detected in the first place), only those listener/scale pairs that showed a significant difference for P vs. A (those marked with an X in Table 8) are included in the analysis. Further, as the specific configuration predicted is

Group	AN < AV	PN < PV
Ι	n.s.	0.019
II	0.009	0.005
III	n.s.	n.s.
IV	n.s.	n.s.
V	n.s.	n.s.

Table 9: Significance values for N < V, by group

different for listener/scale pairs where polarity is reversed (copula absence rated higher than copula presence), only those cases with copula presence rated higher than copula absence are included.<sup>16</sup> The results (p values) for each group are shown in Table 9.

These results show that, for Groups I and II only, there was a significant effect of the grammatical environment on the social evaluation of forms of the copula. When listeners in Group I rated copula presence higher (on some scale) than copula absence, they also rated copula presence before a verb higher than copula presence before a noun significantly more often than chance. This means that the social value of copula presence for these listeners is intensified in the marked environment. There are two possible ways that this could come about. The first is that these listeners know that a following verb is a marked environment for copula presence, and judge that a speaker would only use copula presence in that environment if s/he were particularly emphatic about expressing the social value of copula presence. The second is that the social value of the copula presence is encoded in their grammars as being dependent on the part of speech of the predicative phrase. In this case, the fact that copula presence is rarer before verbs than before nouns would follow from speakers avoiding copula presence before verbs except when they really wanted to express that more intense social value. Either way, the listeners have some direct, if tacit, knowledge which underlies the pattern found in production studies. Similar remarks hold for Group II, in both the copula absence and copula presence conditions.

Why should there be a difference between Groups I and II? The result that Group I listeners only showed an interaction with the grammatical environment in copula presence would follow if copula absence is socially and stylistically unmarked for Group I, while copula presence is socially meaningful. A second possibility is that Group I and Group II participants related to the 'normative' environment of the experiment differently, and that this affected the results for part II of the hypothesis. There was wide variation within Group I in the number of scales individual participants selected as relevant. As suggested above (§4.1), an experimental setting and design better suited to accessing in-group evaluation of the variable may well turn up more robust results for Group I.

Although there are important differences between the results for Groups I and II, both groups' responses are sensitive to the same non-categorical constraint. One might ask why the listeners in Group II, as non-AAVE speakers, were aware of this pattern while the

<sup>&</sup>lt;sup>16</sup>The direction of the difference was determined by running one-tailed Wilcoxon signed rank tests.

listeners in Group III, who also claimed to be familiar with AAVE, were not. The answer is most likely that the African Americans (Group II) had much more extensive and involved experience with AAVE as listeners than did the listeners in Group III who were not African American. The contrast between Groups I and II, on the one hand, and Groups III-V on the other, is an important result: only speakers who are sufficiently familiar with AAVE showed knowledge of the non-categorical constraint, although speakers in other groups did find copula absence/presence relevant to their evaluations of the talkers.<sup>17,18</sup> This shows that the non-categorical constraint is specific to AAVE, and not something latent in the structure of English in general.

# 5 Conclusion

The population studied in this experiment was too small for the results to be conclusive. The results are, however, promising as preliminary evidence that speakers do have knowledge of the non-categorical constraint tested. Aside from increasing the participant pool, there are several ways in which this methodology could be improved and extended. This section briefly considers some of these possibilities.

One big drawback to this experiment was there was only one token set tested. Among other things, this makes significance testing by items (Clark 1973) impossible. If several token sets were used, the experimental task could be kept to a manageable time limit by presenting each listener with only one stimulus from each token set while still ensuring that each listener judges multiple examples of each sentence type (a within subjects design using balanced questionnaires with the particular token representing each sentence type as a random factor). This would necessitate some changes in the statistical methods used, however, as it would remove the possibility of directly comparing individual listeners' judgments of minimal pairs/sets of the sentences.

The selection of scales could also be refined with prior ethnographic work to determine roughly what the social value of the variable actually is. More carefully selected scales would likely turn up more robust results. In addition, if the number of scales considered could be reduced, participants could consider more sentences in the time allotted.

In a larger study, it would also be interesting to see whether the gender of the talker interacts in interesting ways with the social value of the variable and/or the non-categorical

<sup>&</sup>lt;sup>17</sup>The one-tailed test run in the other direction (V < N) comes up significant to p < 0.05 for Group I in the absence condition and Group V in the presence condition and significant to p < 0.01 for Group IV in the absence condition. That the only significant results in Groups III-V should go in the 'wrong' direction is not surprising, as these groups wouldn't be expected to have sufficient experience with AAVE to have internalized the non-categorical constraint. It is not clear at this time what may be producing that pattern, nor whether the same explanation could be applied to the result for Group I.

<sup>&</sup>lt;sup>18</sup>An anonymous reviewer brings up the possibility that, for a given speaker, the following grammatical environment may affect only some aspects of the social evaluation and not others. For example, if the speakers in Groups III-V were asked how 'Black' the talkers sounded in the different guises, they may differentiate the N and V cases, even though they didn't on the scales given. I leave this interesting question to future research.

grammatical constraints on that variable. Gender of talker was excluded as a variable in this study, primarily because including both male and female talkers in the stimulus set would have made it easier for listeners to recognize individual talkers when they heard them for a second, third or fourth time. This could have led to listeners trying to be consistent in the way they rated different talkers. However, in a design that allows for multiple token sets (as described above) there should also be room to include more speakers.

Finally, this methodology might be used to gather information about the grammaticality of certain syntactic structures or other grammatical features in varieties spoken only by people who also control at least one other dialect. The grammaticality judgments of such speakers (especially with regards to a stigmatized variety) are problematic: it's not at all clear whether bidialectal speakers can reliably separate the systems in a grammaticality judgment task. Using methodology similar to that described here, one could investigate the social evaluation of talkers who use features or sentence structures from one variety when (according to phonological cues, etc.) they are ostensibly speaking another. If the features in question truly don't belong to the 'matrix' variety, then the talkers might be judged as 'posing' or (ineffectively) trying to sound like they belong to some other group. From such results, one could make inferences about the grammatical systems of the two varieties. However, any such study must be done carefully, so as to rule out the possibility that the feature in question is part of the variety to be studied, just one that is socially marked.

In this paper, I hope to have shown how the matched guise methodology can be extended to investigate speaker's knowledge of non-categorical constraints (both social and grammatical) on variation. If the preliminary results presented here can be confirmed in larger studies, I believe that they will provide a fruitful area in which to explore the boundaries of linguistic competence.

# References

- Anisfeld, E., and Wallace E. Lambert. 1964. Evaluational reactions of bilingual and monolingual children to spoken language. *Journal of Abnormal and Social Psychology* 69:89–97.
- Anttila, Arto. 1997. Deriving variation from grammar. In F. Hinskens, R. van Hout, and W. L. Wetzels (Eds.), Variation, Change and Phonological Theory, 35–68. Amsterdam: John Benjamins.
- Bailey, Guy, and Natalie Maynor. 1987. Decreolization? Language in Society 16:449–473.
- Baugh, John. 1979. Linguistic Style Shifting in Black English. PhD thesis, University of Pennsylvania.
- Bender, Emily. 2000. Non-categorical constraints in perception. In M. M. Fox, A. Williams, and E. Kaiser (Eds.), Penn Working Papers in Linguistics 7(1), 15–26.
- Bender, Emily M. 2001. Syntactic Variation and Linguistic Competence: The Case of AAVE Copula Absence. PhD thesis, Stanford University.

- Boersma, Paul, and Bruce Hayes. 1999. Empirical tests of the gradual learning algorithm. Unpublished ms., University of Amsterdam and UCLA.
- Cedergren, Henrietta J. 1973. Interplay of Social and Linguistic Factors in Panama. PhD thesis, Cornell University.
- Chomsky, Noam. 1955. The logical structure of linguistic theory. Unpublished ms., Published 1975 by New York: Plenum.
- Clark, Herbert H. 1973. The language as fixed effect fallacy: A critique of language statistics in psychological research. *Journal of Verbal Learning and Verbal Behavior* 12:335–359.
- Cohen, Jonathan D., Brian MacWhinney, Matthew Flatt, and Jefferson Provost. 1993. PsyScope: A new graphic interactive environment for designing psychology experiments. Behavioral Research Methods, Instruments, and Computers 25:257–271.
- Eckert, Penelope. 2000. Linguistic Variation as Social Practice. Oxford: Blackwell.
- Eckert, Penelope, and John R. Rickford (Eds.). 2001. *Style and Sociolinguistic Variation*. Cambridge: Cambridge University Press.
- Fischer, John L. 1958. Social influences on the choice of a linguistic variant. Word 14:47–56.
- Giles, Howard. 1971. Our reactions to accents. New Society (October 14th):713–715.
- Guy, Gregory R. 1988. Advanced Varbrul analysis. In K. Ferrara, B. Brown, K. Walters, and J. Baugh (Eds.), *Linguistic Change and Contact*, 124–136. Austin, Texas: University of Texas Department of Linguistics.
- Labov, William. 1963. The social motivation of a sound change. Word 19:273–309.
- Labov, William. 1966. The Social Stratification of English in New York City. Washington, DC: Center for Applied Linguistics.
- Labov, William. 1969. Contraction, deletion, and inherent variability of the English copula. Language 45(4).
- Lambert, Sylvie. 1973. The role of speech in forming evaluations: A study of children and teachers. Master's thesis, Tufts University.
- Lambert, Wallace E. 1967. A social psychology of bilingualism. *Journal of Social Issues* 23:91–109.
- Lambert, Wallace E., H. Frankel, and G. R. Tucker. 1966. Judging personality through speech: A French-Canadian example. *Journal of Communication* 16:305–321.
- Lambert, Wallace E., Howard Giles, and Omer Picard. 1975. Language attitudes in a French-American community. *Linguistics* 158:127–152.

- Lambert, Wallace E., and G. Richard Tucker. 1975. White and Negro listeners' reactions to various American-English dialects. In J. Dillard (Ed.), *Perspectives on Black English*, Vol. 4 of *Contributions to the Sociology of Language*, 369–377. The Hague: Mouton.
- Rickford, John R. 1998. The creole origins of African-American vernacular English: Evidence from copula absence. In S. S. Mufwene, J. R. Rickford, G. Bailey, and J. Baugh (Eds.), *African-American English*, 154–200. New York: Routledge.
- Rickford, John R., Arnetha Ball, Renee Blake, Raina Jackson, and Nomi Martin. 1991. Rappin on the copula coffin: Theoretical and methodological issues in the analysis of copula variation in African-American Vernacular English. Language Variation and Change 3:103–132.
- Rousseau, Pascale, and David Sankoff. 1978. Advances in variable rule methodology. In D. Sankoff (Ed.), *Linguistic Variation: Models and Methods*, 57–69. New York: Academic Press, Inc.
- Sankoff, David. 1975. Varbrul version 2. Unpublished ms.
- Sylva, Clare M., and Arnold Zwicky. 1975. Discord. In R. W. Fasold and R. W. Shuy (Eds.), Analyzing Variation in Language: Papers from the Second Colloquium on New Ways of Analyzing Variation, 203–219. Washington, DC: Georgetown University Press.
- Trudgill, Peter. 1974. The Social Differentiation of English in Norwhich. Cambridge: Cambridge University Press.
- Wolfram, Walter A. 1969. A Sociolinguistic Description of Detroit Negro Speech. Center for Applied Linguistics.