Average C-Unit Lengths in the Discourse of African American Children From Low-Income, Urban Homes

Holly K. Craig
Julie A. Washington
Connie Thompson-Porter
University of Michigan
Ann Arbor

This investigation reports average length of communication units (C-units) in words and in morphemes for 95 4- to 6 1/2-year-old African American boys and girls from lower-income homes in metropolitan Detroit. Mean C-units increased across the age span of this sample, and kindergartners produced significantly longer C-units than preschoolers. The syntactic complexity of the children's language samples correlated positively with increases in C-unit length, and regression analyses revealed that syntactic complexity was the best predictor of mean C-unit length. Children with longer average C-unit lengths produced greater frequencies of all types of syntactic complexity. Their language samples were distinguished from children with shorter mean C-unit lengths by clauses linked with coordinate and subordinate conjunctions. The findings indicate that average C-unit length will be useful as a quantitative index of linguistic growth in research designs focusing on young school-age African American children living in poverty.

KEY WORDS: African Americans, children, low income, sentence lengths, language development

Most African Americans speak African American English (AAE) at least to some degree (Battle, 1993; Smitherman-Donaldson, 1977). African American children from low-income homes use more dialectal forms than their middle-class peers, and the discourse of boys evidences more dialect than that of girls (Washington & Craig, in press). In some of our earlier work, as many as 39% of the utterances of African American preschoolers from urban, low-income homes were found to include morphosyntactic dialectal forms (Washington & Craig, 1994). When as much as approximately one-third of a child's discourse may incorporate morphosyntactic differences from Standard American English (SAE), it is not surprising that African American children can be disadvantaged in schools where SAE is the primary mode of instruction (Carnegie Council on Adolescent Development, 1989). In particular, African American children appear over-represented on the special education caseloads of the nation's public schools (Statistical Profile of Special Education in the United States, 1994).

Very little information is available concerning the language development of the African American child. The earliest research confirmed that the dialect use of the child and youth was rule governed like that of

adult African Americans and, accordingly, dialect should be considered a sociolinguistic difference rather than a language deficit (Baratz, 1970; Dillard, 1972; Fasold & Wolfram, 1970; Stewart, 1970; Wolfram, 1971; Wolfram & Fasold, 1974). Subsequently, the modification of widely used language measures into versions that would not falsely identify a dialect user as language impaired has been advocated (Terrell & Terrell, 1993). The modification of formal language tests has been emphasized particularly but with disappointing results (Hemingway, Montague, & Bradley, 1981; Washington & Craig, 1992a; Wiener, Lewnau, & Erway, 1983). There seems to be no quick alternative to undertaking a sustained inquiry into the characteristics of the language used by African American children and determining the developmental course.

This line of research would benefit from a quantitative metric of language development. Average sentence length is an heuristic of this type and is well established as a subject-matching variable used to equate groups in research designs across a diverse set of investigative purposes. Average sentence length data offer the researcher a number of advantages over formalized measures like standardized tests. Most fundamentally, they are usually computed from spontaneous language samples derived during natural interactions. Naturalistic contexts such as freeplay with toys are comfortable settings for children. Freeplay places fewer external constraints on the child's language formulation than responses to structured tasks, and thereby should elicit more representative samples of the child's linguistic skills (Bloom & Lahey, 1978; Brown, 1973; Lund & Duchan, 1988; Prutting, Gallagher, & Mulac, 1975). In addition, with the recent advent of computerized transcription, coding, and analysis programs such as the Children's Data Exchange System (CHILDES; MacWhinney, 1994) and Systematic Analysis of Language Transcripts (SALT; Miller & Chapman, 1987), these calculations can be performed more automatically and accurately than some other matching measures.

Measures of average sentence length have taken a variety of forms over the years. Some of the earliest researchers calculated mean number of words per sentence, Mean Length of Response (MLR; McCarthy, 1930; Templin, 1957). As information accumulated about the morphosyntactic growth of emerging sentences, Mean Length of Utterance (MLU), which calculates mean number of morphemes per utterance, became the primary measure. Utterance-based analyses preserve emerging and, therefore, often incomplete grammatical units for analysis, and reflect the very young child's gains in morphology (Brown, 1973; Rondal, Ghiotto, Bredart, & Bachelet, 1987; Scarborough, Wyckoff, & Davidson, 1986; Wells, 1985). Utterance boundaries for MLU analyses are usually defined by pauses, terminal intonation contours,

and syntactic and semantic structure (Miller, 1981; Miller & Chapman, 1987). MLU remained a primary subject descriptor variable through the pragmatic/sociolinguistic research era of the 1980s and into the 1990s because its utterance base allowed inclusion of brief but important conversational forms such as stereotypic acknowledgments to comments and the contingent query (Craig & Gallagher, 1982; Gallagher, 1981).

MLU is useful at early language stages, but its viability for older children has been questioned. Although Miller (1991) indicated that MLU increases steadily up to age 13, it may no longer index grammatical development. Beyond 42 months of age or when MLUs are greater than 4.00, this measure can lack reliability and varies by context (Brown, 1973; D'Odorico & Franco, 1985; Klee & Fitzgerald, 1985; Kramer, James, & Saxman, 1979; Scarborough et al., 1986; Wells, 1985).

Alternatively, Terminable units (T-units) segment written clauses (Hunt, 1970) and include a main clause, its subordinate clauses, and other nonclausal embedded structures. T-units separate main clauses joined by simple coordinate conjunctions (*and, but, or*) into separate units, unless the second clause evidences a co-referential subject deletion. Division of strings of clauses linked only by simple coordinate conjunctions is advantageous for children of school age who are capable of producing very long utterances. T-unit length in words has been found to increase systematically with grade (Hunt, 1970; Klecan-Aker & Hedrick, 1985; O'Donnell, Griffin, & Norris, 1967).

Communication units (C-units), defined by Loban (1976) as independent clauses plus their modifiers, offer some important advantages over other potential segmentation units like the T-unit. Like T-units but unlike the utterance, C-units segment discourse into main clauses when conjoined with simple coordinate conjunctions unless a co-referential subject deletion characterizes the second clause. Developed for oral language analysis purposes, however, like the utterance, C-units preserve more of the interaction than the T-unit by including in the corpus stereotypical single words (for example: "yes," "no," "oh," etc.) and other nonclausal units if they are responses to an immediately prior adult request. In young children who may not be producing coordinated structures consistently, the C-unit and utterance likely function similarly. C-units have the advantage, however, of being applicable to older children and to the segmentation of spoken discourse, features recommending this segmentation unit for the study of African American children's oral language at the time of school entry.

Average C-unit length was one of the measures used by Laine (1978) when characterizing the linguistic skills of successful and unsuccessful readers who were African American, Chicano, and Anglo 7- and 10-year-old boys

from middle and from lower SES homes, residing in metropolitan Los Angeles. Oral responses to questions and spontaneous picture descriptions were the elicitation contexts, and reading success was based upon teacher report. Unfortunately, the outcomes are somewhat difficult to interpret because corpus size per subject was not provided, and the tables note that only the values for children from the LSES samples have been reported. Laine states that average C-unit lengths were not significantly different relative to ethnicity or SES. Successful readers, however, did have significantly longer C-units than unsuccessful readers. These findings are suggestive; average C-unit lengths may be useful for characterizing a range of language skills of African American children.

It will be important to evaluate the relationship between C-unit length and linguistic complexity. Average C-unit length would be most useful to clinicians and researchers if it indexed linguistic growth. In addition, it will be important to determine whether gender influences the values obtained when pursuing this line of research with African American children. In other research, Washington and Craig (in press) found that amount of dialect varied systematically by gender in the discourse of African American kindergartners. Accordingly, average C-unit length may be influenced by gender, and by amount of dialect use.

Mean length of C-unit in words (MLCU-w) and mean length of C-unit in morphemes (MLCU-m) were selected for these purposes. Word-based calculations can be advantageous for school-age children because many linguistic gains during the school years reflect the operation of large units like relative clauses and adverbial relationships. Increasing numbers of words in C-units should be sensitive to these types of linguistic growth. Morpheme-based calculations are another widely used measure of linguistic growth, and demonstrate systematic increases through age 13 (Miller, 1991).

The purpose of this investigation was to contribute to the process of describing the oral language samples of normally developing African American children. Preschoolers, kindergartners, and children from low-income homes were the focus of investigation to continue building a developmental reference profile for language skills in this population that was begun in some of our earlier work (Craig & Washington, 1994, 1995; Washington & Craig, 1992a, 1992b, 1994, in press). In order to begin to address these issues, the following questions were posed.

1. What is the average C-unit length in words and in morphemes for young African American children who are apparently normally developing?

2. How does mean C-unit length in words relate to mean C-unit length in morphemes for young African American children?

3. Are there systematic variations in average C-unit length relative to the following variables: gender, chronological age, syntactic complexity, and AAE?

Methods
Subjects

The subjects were 95 4- to 6 1/2-year-old African American boys (n = 45) and girls (n = 50) residing in metropolitan Detroit. African American children comprised more than 75% of the student body in the school districts participating in this investigation. All of the children were dialect users.

All of the subjects were from lower socioeconomic status (LSES) homes determined by the demographics of the children's communities, and their participation in a school-based program designed for children living in poverty. The younger children were enrolled in one of Michigan's state-funded preschool programs for children "at-risk." The current subjects were considered at-risk because family income levels fell below the poverty line. The kindergartners participated in a free or reduced-price federal lunch program for low-income families at their schools. The family incomes of the children who participate in the lunch programs must be between 135% and 185% of the federally established poverty guideline for a family of four during the year in which the data were collected.

Based upon teacher and parental judgments, all of the subjects were developing normally, free from medical, emotional, or sensory problems. None were enrolled in special education or speech and language services of any type. No formal tests were administered to confirm the reports of their normal developmental status because no culturally valid language tests are available for this population (see Washington, 1996, for additional discussion). Prior to data collection, each child passed a bilateral hearing screening at 25dB for 500, 1,000, and 4,000 Hz (ANSI, 1978).

Data Collection

Spontaneous language samples were collected during dyadic interaction with one of a team of six African American female examiners in a freeplay context. Each examiner was an experienced tester who spoke AAE to the children. The children and examiners were previously unacquainted.

Language samples were approximately 15 to 20 minutes in duration. Each child was presented with three toy sets and allowed to select one for play. This attempted to control for interest level across children and the choices were successful in maintaining the

children's interest for the duration of the tapings. The three toy sets comprised action toys: Barbie and Ken dolls with a Burger King playset, action figures and props, and the Fisher-Price School.

Each child wore an individual microphone and the samples were audio-recorded using a microphone mixer. The data collection took place in a quiet room in the child's school.

Scoring

The freeplay samples were transcribed using the segmentation criteria of Loban (1976) for C-units. Successive main clauses linked by a simple coordinate conjunction were segmented into separate C-units if the second clause included a subject. For example: "I got a gun and I'm gonna shoot you" was segmented into "I got a gun" and a second C-unit "and I'm gonna shoot you." If a co-referential subject in the second clause was elliptically omitted, the utterance was considered a single C-unit (for example: "she only needs to do her hair and put on a skirt"). C-unit corpora include nonclausal utterances if they are a response to a prior adult question. Similarly, we included single word stereotypical acknowledgments to a prior adult comment and productions of "what" functioning as a contingent query by the child subsequent to an adult utterance.

C-units were transcribed orthographically using CHAT conventions of the Children's Data Exchange System (CHILDES; MacWhinney, 1994). The first 50 wholly intelligible C-units were stored in separate CHAT files. These 50 C-unit files were scored for the presence of AAE, using Washington and Craig's (1994) definitions. Appendix A presents their definitions with examples. The 50 C-unit samples were also scored for the presence of complex syntax, using Craig and Washington's (1994) taxonomy. Appendix B presents definitions and examples of the types of complex syntax. More than one type of complex syntax code was possible per C-unit. In the earlier investigation, Craig and Washington excluded pragmatic connectives that occur to initiate a turn and link the message to that of the prior speaker. In the present investigation, pragmatic connectives again were excluded as part of the conjunction analysis. For example, an examiner said: "Do you have Ken at home?" and a subject responded: "Yeah but he got hair." This turn initial conjunction was not scored as an instance of complex syntax. In addition, when the coordinate conjunctions *and, but, or* were judged to initiate a new C-unit, they were not scored as part of the conjunction analysis. Consequently, only coordinate conjunctions that were internal to C-units were coded as part of the syntactic complexity analysis. In the example: "I got a gun" "and I'm gonna shoot you" the C-unit initial

conjunction "and" was not scored as an instance of complex syntax, whereas the "and" in the C-unit "she only needs to do her hair and put on a skirt" was scored as an instance of complex syntax.

For each subject's 50 C-unit file, mean length of C-units in words was generated automatically using the Mean Length of Turn (MLT) analysis program in CLAN. Consistent with Loban's procedures, MLT excludes mazes from word counts when tagged during transcription with the appropriate conventions. Unlike the Loban procedures, however, MLT does not count contractions as separate words. Mean length of C-units in morphemes were generated automatically by CLAN using the MLU command for the 50 C-unit files. The frequency command (FREQ) generated types and tokens of AAE codes and types and tokens of complex syntax codes for the 50 C-unit files.

Reliability

Transcription reliabilities were established for all subjects. Approximately 10% of each transcript was re-transcribed by an independent observer. A point-to-point comparison at the level of the morpheme was high, 88%, when the number of agreements was divided by the number of agreements plus disagreements.

Whereas the coding of AAE constituted a core analysis for all aspects of our research program, coding agreements were established for the 50 C-unit files of every subject. An independent observer re-scored each 50 C-unit file for AAE, and a point-to-point comparison was calculated by dividing the number of agreements by the number of agreements plus disagreements. Percentages of agreement for types and for tokens were high, 97% and 85%, respectively. Eleven transcripts were then randomly selected, representing just over 10% of the subjects. A point-to-point comparison for C-unit segmentation was 98%, and for the occurrence of complex syntax was 91% for types and 83% for tokens.

In addition, verbal behaviors of the examiners were investigated in two ways in order to explore the impact of potential variations on the verbal behaviors of the children. Whereas the number of subjects tested by a specific examiner was determined based upon scheduling issues, the specific numbers of language samples involving any particular examiner varied. For analysis purposes, five interactions were randomly selected for each of the four examiners who had collected at least five of the samples reported in this investigation, representing approximately 20% of the total interactions. The frequencies of both AAE tokens, $F(19) = 4.53$, $p < .05$, and types, $F(19) = 3.81$, $p < .05$, varied significantly (see Table 1). A Tukey Honestly Significant Difference (HSD) revealed that one of the examiners used significantly

Table 1. Means (M) and standard deviations (SD) for frequencies of AAE tokens, types, and rate of utterances for examiners, and for the children relative to the examiners.

Examiners	AAE tokens		AAE types		Utterances	
	M	SD	M	SD	M	SD
Examiner behaviors						
1	10.0	5.6	2.0	1.0	.27	.05
2	* 6.6	3.4	1.6	0.9	.18	.03
3	*,**17.8	7.5	3.8	1.9	.32	.22
4	**6.6	5.0	1.6	0.5	.21	.02
Child behaviors						
1	10.2	5.0	3.8	0.4	NA	NA
2	8.8	4.4	2.6	0.9	NA	NA
3	12.8	9.5	3.4	1.8	NA	NA
4	8.8	10.2	2.8	1.8	NA	NA

*p ≤ .05
**p ≤ .01
NA = not appropriate

more dialect than two other examiners but all other AAE comparisons were nonsignificant. Although dialect use varied by examiner, the rate of adult utterance production was not statistically different, $F(19) = 1.45, p > .05$. Rate of production was determined by dividing the total number of adult utterances by the amount of time it took the children to produce 50 C-units. Next, the impact of the variations among examiners on the behaviors of the children were estimated by comparing the frequencies of AAE types and tokens produced by the children relative to the examiners (see Table 1). Despite a difference in the dialect use of some examiners, no statistically significant differences were apparent for the frequencies of AAE types, $F(19) = 0.81, p > .05$, or tokens, $F(19) = 0.30, p > .05$, produced by the children. Considered together, these findings indicate that the verbal behaviors of the children were not affected in systematic ways by the circumscribed variations in dialect across examiners.

Results

Eighty-nine of the children (94% of the sample) produced 50 intelligible C-units or more in the sampling period. Average language sample lengths from those who did not produce 50 C-units ranged from 14 to 45 intelligible C-units. Three of these 6 children with short samples were preschoolers and 3 were kindergartners. Their mean number of words per intelligible C-unit are included in the subsequent analyses so that the distributions obtained would be maximally representative rather than reflecting only highly verbal children (as recommended by Miller & Chapman, 1981).

Average C-unit lengths calculated in words and in morphemes evidenced an extremely strong, positive, statistically significant relationship to each other ($r = .94, p < .001$). A number of relationships to each measure of average C-unit length were explored and are reported in Table 2. There was no statistically significant difference in mean C-unit lengths based upon gender [MLCU-w: $t(93) = 0.33, p > .05$; MLCU-m: $t(93) = 0.21, p > .05$] and this variable was collapsed in subsequent analyses.

Age and syntactic complexity related to MLCU-w and to MLCU-m. Pearson product moment correlations showed a very low, positive statistically significant correlation between chronological age in months and the two measures of MLCU (MLCU-w: $r = .20, p = .049$; MLCU-m: $r = .22, p = .03$). See Table 3. Age groups in years are presented in Table 2 and show a small but steady increase in MLCUs. For words, a one-way ANOVA revealed a statistically significant difference between the groups, $F(92) = 4.89, p ≤ .01$. A Tukey-HSD post hoc analysis revealed that the average C-unit length of the 6-year-olds was significantly greater than the 4-year-olds ($p < .05$) but the mean for the 5-year-olds was not significantly different from the younger or the older groups. The same relationships were found for morphemes, $F(92) = 5.35, p < .01$, with a statistically significant difference observed only between the MLCU-m of the 4-year-olds and the 6-year-olds.

Table 2 also displays MLCUs by grade. An independent t-test revealed that the MLCUs of the kindergartners were significantly greater than those of the preschoolers for both word, $t(93) = 2.44; p ≤ .01$, and morpheme, $t(93) = 2.94; p ≤ .01$, length.

AAE includes some forms that optionally add words (for example: multiple negation, double modals, appositive pronoun) or delete words (e.g., zero cop/aux, zero

Table 2. Means (M) and standard deviations (SD) of the C-units (n = 95) for words (MLCU-w) and for morphemes (MLCU-m).

	n	MLCU-w		MLCU-m	
		M	SD	M	SD
Males	45	3.33	.65	3.72	.63
Females	50	3.38	.70	3.76	.82
Age (in years)					
4	29	*3.14	.70	**3.48	.75
5	53	3.36	.63	3.76	.71
6	13	*3.81	57	**4.24	.57
Grade					
Preschoolers	52	*3.21	.70	**3.55	.75
Kindergartners	43	*3.54	.60	**3.98	.65

*p ≤ .05
**p ≤ .01

Table 3. Intercorrelations among mean C-unit lengths, age, frequencies of African American English (AAE) averaged across words, complex syntax (types) for words (MLCU-w) and for morphemes (MLCU-m).

	MLCU-w (words)	Age (months)	AAE/ words	Complex syntax (types)
MLCU-w (words)	—	*.20	*.22	**.72
Age (months)		—	.14	.12
AAE/words (tokens)			—	.09
Complex syntax (tokens)				—

	MLCU-m (mor- phemes)	Age (months)	AAE/ (words)	Complex syntax (types)
MLCU-m (morphemes)	—	*.22	.15	**.65
Age (months)		—	.14	.12
AAE/words			—	.09
Complex syntax (types)				—

$^*p \leq .05$
$^{**}p \leq .01$

"to") and could affect a word-based calculation. Other AAE forms affect bound morphemes (for example: zero possessive, subject-verb agreement) and these may influence a morpheme-based calculation of C-unit length. Accordingly, a relationship between AAE and both the word- and the morpheme-based calculations of MLCU were explored. In order to do this, the frequencies of AAE occurring in each subject's 50 C-unit sample were divided by the number of words in the sample. Averaging AAE tokens across words was an attempt to avoid creating an artifactual relationship between increased opportunities for AAE to occur in longer sentences, by definition comprised of a greater number of words and morphemes. A Pearson product moment correlation revealed a very low, positive, statistically significant relationship between MLCU-w and the frequencies of AAE forms/words $(r = .22, p = .03)$ and a nonsignificant relationship between MLCU-m and the frequencies of AAE forms averaged across words $(r = .15, p > .05)$. See Table 3.

Finally, the relationship between the complex syntax used by the children and their average C-unit length was examined as well. Most of the children (90 of 95) used complex syntax, but the amounts evident in their samples varied widely, with frequencies ranging from 1 to 16 occurrences. Again, in order to avoid an artifactual relationship between opportunities for occurrences of complex syntax with the greater number of words and morphemes inherent in longer C-units, the number of different complex syntax types produced in the standard 50 C-unit samples, rather than token frequencies were

examined. A Pearson product moment correlation revealed a strong, positive, statistically significant relationship for MLCU-w and the number of different types of complex syntax used by the children $(r = .72, p \leq .001)$. Similarly, MLCU-m correlated strongly with the number of different types of complex syntax occurring in the samples $(r = .65, p \leq .01)$. See Table 3.

Simultaneous regression analyses were pursued using the variables that correlated significantly with the word- and morpheme-based measures to probe their impact on mean C-unit length. Age, AAE frequencies averaged across words, and number of different types of complex syntax were examined for MLCU-w. When age was part of the regression model, it did not contribute significantly to the amount of variance explained $(p = .16)$. Frequencies of complex syntax and AAE had significant effects on average C-unit length whether or not age was part of the regression model, $R^2 = .55, p < .001$; $R^2 = .54, p < .001$, respectively. Table 4 summarizes the simultaneous regression model when these two variables are regressed on mean C-unit length. Complex syntax had greater relative effect on explaining MLCU-w ($\beta = .70$) than dialect use ($\beta = .16$).

For MLCU-m, age and number of different types of complex syntax were examined to evaluate the impact of these two significantly correlated variables. Again, when age was part of the regression model, it did not contribute significantly to the amount of variance explained $(p = .06)$. Frequencies of complex syntax had significant effects on MLCU-m whether or not age was part of the regression model, $R^2 = .44, p < .001; R^2 = .42, p < .001$, respectively. Complex syntax explained a considerable amount of MLCU-m variation ($\beta = .65$).

Overall, these analyses indicate that mean C-unit length varied by grade, was weakly related to a child's

Table 4. Summary of simultaneous regression analyses for variables predicting MLCU-w and frequencies of complex syntax .

Variable	B	SE B	β
MLCU-w			
Complex syntax (types)	.26	.03	.70**
AAE/words	3.16	1.44	.16*
$R^2 = .54$			
Complex Syntax			
MLCU-w	1.94	.19	.72**
$R^2 = .52$			
MLCU-m			
Complex syntax (types)	.26	.03	.65**
$R^2 = .42$			

$^*p \leq .05$
$^{**}p \leq .001$

chronological age, and is better understood as an index of syntactic complexity. As a check on this interpretation, simultaneous regression analyses were computed differently, with types of complex syntax functioning as the dependent variable in each regression model. When the number of different types of complex syntax was the dependent variable, both age and AAE frequencies did not contribute significantly to the amount of variance explained in the model for MLCU-w ($p = .95$, $p = .30$, respectively). However, mean C-unit length explained the amount of variance in the frequencies of complex syntax at a statistically significant level ($R^2 = .52$, $p < .001$). Similarly for morphemes, mean C-unit length explained the amount of variance in the frequencies of complex syntax at a statistically significant but lower level ($R^2 = .42$, $p < .001$).

In order to confirm the quantitative finding that increasing C-unit lengths reflected more complex syntax, the subject sample was subdivided into the top and bottom quartiles for the MLCU-w measure. Figure 1 displays the percentage of children in the bottom quartile, those using the shortest C-unit lengths (MLCU-w ≤ 2.98) to those at or above the 75th percentile (MLCU-w ≥ 3.88). This more qualitative analysis confirms that longer C-unit lengths represent greater use of complex syntax. The portion of the subject sample who produced the longest C-units also produced the same types of complex syntax as the portion of the subject sample with the shortest C-units, but to a greater extent. Further, the top quartile was distinguished from the bottom quartile by greater coordinate and subordinate conjunction use.

Discussion

In the present study, average C-unit length evidenced a significant increase by grade, correlated positively with age, and predicted relative amounts of complex syntax. Both word- and morpheme-based averages yielded remarkably similar results, and their values correlated very highly with each other as well. These similarities are not surprising because many language gains beyond the early years affect free rather than bound morphemes, for example, subordinate clauses, embedded phrases, advanced infinitives, post-verb modifications, and so forth (Hunt, 1965; Loban, 1976; O'Donnell et al., 1967; Scott, 1988). Both measures would reflect the inclusion of more words in the calculations.

MLCU-w correlated significantly with amounts of dialect in the samples but the relationship was very weak ($r = .22$), and MLCU-m was not significantly related to the children's dialect use. In earlier work, Washington and Craig (1994) found that AAE forms mark fewer than half of the utterances produced by African American preschoolers. Average C-unit length appears to be unaffected by variations in dialectal marking at these ages and, therefore, should be useful as a way to globally characterize the sentence production skills of African American children despite our lack of information about the developmental course of AAE.

The C-units were collected during freeplay interactions and the African American examiners in these contexts spoke AAE to the children. The low structure of freeplay, the interactions with an adult of the same race,

Figure 1. Percentages of subjects in the top (*n* = 24) and bottom (*n* = 24) quartiles using each type of complex syntax. ISS = Infinitive Same Subject, NIW = Simple Noninfinitive Wh-Clause, NPC = Noun Phrase Complement, LET = Let(s)/Lemme and Infinitive, REL = Relative Clause, IDS = Infinitive with a Different Subject, UNI = Unmarked Infinitive, WHI = Wh-Infinitive Clause, GER = Gerunds and Participles, TAG = Tag Questions, COOR = Coordinate Conjunctions, SUB = Subordinate Conjunctions.
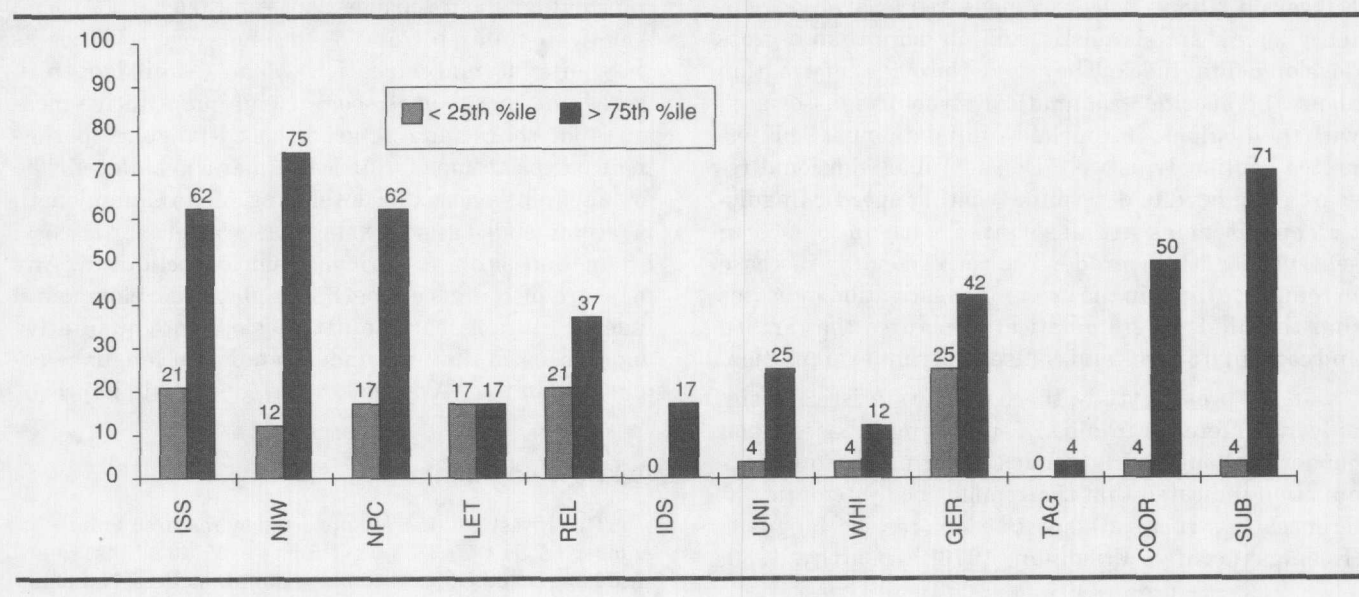
and the examiners' use of dialect all should have placed the linguistic choices more under the control of the children, compared to what has been possible in the administration of prior published reports with formal tests (Hemingway, Montague, & Bradley, 1981; Washington & Craig, 1992a; Wiener, Lewnau, & Erway, 1983). The analysis of spontaneous language samples can be labor intensive. In contrast, average C-unit length can be calculated as mean number of words or morphemes automatically and quickly with available computerized software packages. Preserving the spontaneity of freeplay discourse while improving the efficiency of quantitative analyses recommend average C-unit length as a global measure of linguistic growth for young African American children from low-income, urban homes.

This investigation is no more than a first step in the process of fully developing an average C-unit length measure for quantifying the oral language of young African American children. Prior research with SAE-speaking children has revealed systematic differences in average sentence lengths when context varies, for example, between the home and the clinic with the mother or clinician as interlocutor (Kramer, James, & Saxman, 1979; Olswang & Carpenter, 1978; Scott & Taylor, 1978). Context and interactant effects remain unknown when the children are AAE speakers, and should be pursued. In addition, different discourse genres might have yielded different values. Further, the subjects in this study were all young children and all from low-income homes. Average C-unit length should increase systematically for older African American children and there may be differences based upon SES. Considerable additional research is necessary to understand more fully the potential utility of this measure for studying the discourse of African American children.

As cautioned by others (Cazden, 1968; Klee & Fitzgerald, 1985), however, measures of average sentence length are simplistic, and no more than a gross developmental index. They do not provide information about the specific grammatical structures associated with their values. Our mean C-unit data must be interpreted cautiously as well. Considerable additional research will need to determine whether specific linguistic competencies are associated with increases in sentence length by grade. The corresponding increase in complex syntax in the present investigation indicates that the kinds of grammatical structures that embed and coordinate verbs and clauses warrant examination.

In the present study, there was no statistically significant difference in mean C-unit length based upon gender. Although early studies of normal language acquisition indicated that there might be systematic differences in structural aspects of expressive language (Koenigsknecht & Friedman, 1976; McCarthy, 1930;

Winitz, 1959), more recent research has indicated that there are more similarities than differences between the language performances of boys and girls and that the differences that are apparent relate more to pragmatic use than structural knowledge (Craig & Evans, 1991; Ely, Berko-Gleason, & McCabe, 1996; Macaulay, 1978; Sheldon & Rohleder, 1996). These prior studies focused upon the language development of boys and girls who were SAE-speakers. In one of the few investigations of gender influences on the language of African American boys and girls, Washington and Craig (in press) reported significantly greater frequencies of dialectal forms in the discourse of kindergarten boys than girls. This pattern confirms earlier observations for gender-based differences between African American men and women (Abrahams, 1973; Wolfram, 1969) and extends the findings to include systematic differences in children as young as 5 and 6 years. Considered relative to the findings of the present investigation, gender-based differences in the discourse of African American children may be circumscribed to the frequency of use of dialectal forms, and not be evident for other aspects of language production.

Mean C-unit length correlated significantly with age, but the positive correlation was low suggesting that the relationship was a weak one. A number of investigators have reported weak correspondences between age and MLU (Brown, 1973; D'Odorico & Franco, 1985; Klee & Fitzgerald, 1985; Scarborough et al., 1986). As children enter school, grade may be a better correlate to increasing C-unit length than age and, indeed, provides the data analysis framework for most research on increasing sentence length (see review by Scott, 1988). This is not surprising because practical considerations would prioritize grade placements in school-based research programs.

In summary, this investigation provides descriptive information about the average C-unit length of African American children from low-income, urban homes as they enter formal education. Mean C-unit length in words and morphemes seem useful quantitative measures for characterizing grade-based language performance expectations for African American children who are apparently normally developing. Mean C-unit length is recommended as a first step in characterizing the oral production skills of African American children. Any measure of sentence length is a global developmental index. Future research must develop more qualitative approaches so that our understanding of language acquisition in the AAE-speaker can continue to improve.

Acknowledgments

Institutes of Health. Special thanks to the principals, teachers, parents, and students of the Oak Park, MI public school system for supporting this project.

References

Abrahams, R. (1973). The advantages of Black English. In J. DeStefano (Ed.), *Language, society, and education: A profile of Black English* (pp. 97–106). Worthington, OH: Charles A. Jones Publishing Company.

American National Standards Institute. (1978). *Methods for pure tone audiometry* (ANSI S.3.21-1978). New York: ANSI.

Baratz, J. C. (1970). Educational considerations for teaching standard English to Negro children. In R. W. Fasold & R. W. Shuy (Eds.), *Teaching Standard English in the inner-city* (pp. 20–39). Washington, DC: Center for Applied Linguistics.

Battle, D. E. (1993). Introduction. In D. E. Battle (Ed.), *Communication disorders in multicultural populations* (pp. xv–xxiv). Boston: Andover Medical Publishers.

Bloom, L., & Lahey, M. (1978). *Language development and language disorders.* New York: John Wiley and Sons.

Brown, R. (1973). *A first language.* Cambridge: Harvard University Press.

Carnegie Council on Adolescent Development. (1989). *Turning point: Preparing American youth for the 21st century.* New York: Carnegie Corporation.

Cazden, C. (1968). The acquisition of noun and verb inflection. *Child Development, 39,* 433–438.

The Council for Exceptional Children. *Statistical profile of special education in the United States.* (1994). Reston, VA: Author.

Craig, H., & Evans, J. (1991). Gender and children's turn-taking. *Journal of Speech and Hearing Research, 34,* 866–878.

Craig, H., & Gallagher, T. (1982). Gaze and proximity as turn regulators within three-party and two-party child conversations. *Journal of Speech and Hearing Research, 25,* 32–42.

Craig, H. K., & Washington, J. A. (1994). The complex syntax skills of poor, urban, African-American preschoolers at school entry. *Language, Speech, and Hearing Services in Schools, 25,* 181–190.

Craig, H. K., & Washington, J. A. (1995). African-American English and linguistic complexity in preschool discourse: A second look. *Language, Speech, and Hearing Services in Schools, 26,* 87–93.

Dillard, J. (1972). *Black English.* New York: Random House.

D'Odorico, L., & Franco, F. (1985). The determinants of baby talk: Relationship to context. *Journal of Child Language, 12,* 567–586.

Ely, R., Berko-Gleason, J., & McCabe, A. (1996). Why didn't you talk to your mommy honey?: Gender differences in talk about past talk. *Research on Language in Social Interaction, 29,* 7–25.

Fasold, R., & Wolfram, W. (1970). Some linguistic features of Negro dialect. In R. Fasold & R. Shuy (Eds.), *Teaching Standard English in the inner-city* (pp. 41–86). Washington, DC: Center for Applied Linguistics.

Gallagher, T. (1981). Contingent query sequences within adult-child discourse. *Journal of Child Language, 8,* 51–62.

Hemingway, B. L., Montague, J. C., & Bradley, R. H. (1981). Preliminary data on revision of a sentence repetition test for language screening with black first grade children. *Language, Speech, and Hearing Services in Schools, 12,* 153–159.

Hunt, K. W. (1965). *Grammatical structures written at three grade levels.* Urbana, IL: National Council of Teachers of English.

Hunt, K. W. (1970). Syntactic maturity in schoolchildren and adults. *Monographs of the Society for Research in Child Development, 35* (1, Serial No. 134).

Klecan-Aker, J. S., & Hedrick, D. L. (1985). A study of the syntactic language skills of normal school-age children. *Language, Speech, and Hearing Services in Schools, 16,* 187–198.

Klee T., & Fitzgerald, M. D. (1985). The relation between grammatical development and mean length of utterance in morphemes. *Journal of Child Language, 12,* 251–269.

Koenigsknecht, R., & Friedman, P. (1976). Syntax development in boys and girls. *Child Development, 47,* 1109–1115.

Kramer, C. A., James, S. L,. & Saxman, J. H. (1979). A comparison of language samples elicited at home and in the clinic. *Journal of Speech and Hearing Disorders, 44,* 321–330.

Laine, J. E. (1978). A language analysis of successful and nonsuccessful readers: Comparing linguistic ability in Black, Chicano, and Anglo boys. *Journal of Black Studies, 8,* 439–451.

Loban, W. (1976). *Language development: Kindergarten through age twelve.* Urbana, IL: National Council of Teachers of English.

Lund, N. J., & Duchan, J. F. (1988). *Assessing children's language in naturalistic contexts.* Englewood Cliffs, NJ: Prentice-Hall, Inc.

Macaulay, R. (1978). The myth of female superiority in language. *Journal of Child Language, 5,* 353–363.

MacWhinney, B. (1994). *The CHILDES project: Tools for analyzing talk* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum.

McCarthy, D. (1930). *The language development of the preschool child.* Minneapolis: University of Minnesota.

Miller, J. (1981). *Assessing language production in children.* Baltimore: University Park Press.

Miller, J. (1991). Quantifying productive language disorders. In J. Miller (Ed.), *Research on child language disorders: A decade of progress* (pp. 211–220). Austin, TX: Pro-Ed.

Miller, J. F., & Chapman, R. S. (1981). The relation between age and mean length of utterance in morphemes. *Journal of Speech and Hearing Research, 24,* 154–161.

Miller, J. F., & Chapman, R. S. (1987). *Systematic Analysis of Language Transcripts (SALT): A computer program designed to analyze free speech samples.* Madison, WI: University of Wisconsin-Madison.

O'Donnell, R. C., Griffin, W. J., & Norris, R. D. (1967). Syntax of kindergarten and elementary school children: A transformational analysis. *Research Report No. 8.* Champaign, IL: National Council of Teachers of English.

Olswang, L. B., & Carpenter, R. L. (1978). Elicitor effects on the language obtained from young language-impaired children. *Journal of Speech and Hearing Disorders, 43,* 76–88.

Prutting, C., Gallagher, T., & Mulac, A. (1975). The expressive portion of the N.S.S.T. compared to a spontaneous language sample. *Journal of Speech and Hearing Disorders, 40,* 40–49.

Rondal, J. A., Ghiotto, M., Bredart, S., & Bachelet, J. (1987). Age-relation, reliability and grammatical validity of measures of utterance length. *Journal of Child Language, 14,* 433–446.

Scarborough, H., Wyckoff, J., & Davidson, R. (1986). A reconsideration of the relation between age and mean utterance length. *Journal of Speech and Hearing Research, 29,* 394–399.

Scott, C. M. (1988). Spoken and written syntax. In M. Nippold (Ed.), *Later language development: Ages nine through nineteen* (pp. 49–96). Boston, MA: College Hill Press.

Scott, C. M., & Taylor, A. E. (1978). A comparison of home and clinic gathered language samples. *Journal of Speech and Hearing Disorders, 43,* 482–495.

Sheldon, A., & Rohleder, L. (1996). Sharing the same word, telling different stories: Gender differences in co-constructed pretend narratives. In D. I. Slobin, J. Gerhardt, A. Kyratzis, & J. Guo (Eds.) *Social Interaction, social context, and language* (pp.613-632). Mahwah, NJ: Lawrence Erlbaum Associates.

Smitherman-Donaldson, G. (1977). *Talkin and testifyin: The language of Black America.* Boston: Houghton Mifflin.

Stewart, W. (1970). Toward a history of American Negro dialects. In F. Williams (Ed.), *Language and poverty* (pp. 51–70). Chicago: Markham Publishing.

Templin, M. (1957). Certain language skills in children. *Institute of Child Welfare Monograph Series, No. 26.* Minneapolis, MN: University of Minnesota Press.

Terrell, S. L., & Terrell, F. (1993). African-American cultures. In D. E. Battle (Ed.), *Communication disorders in multicultural populations* (pp. 3–37). Stoneham, MA: Butterworth-Heinemann.

Washington, J. A. (1996). Issues in assessing the language abilities of African American children. In A. Kamhi, K.

Pollock, & J. Harris (Eds.), *Communication development and disorders in African American children* (pp. 35–54), Baltimore, MD: Paul H. Brooks Publishing Co.

Washington, J. A., & Craig, H. K. (1992a). Performances of low-income, African-American preschool and kindergarten children on the Peabody Picture Vocabulary Test–Revised. *Language, Speech, and Hearing Services in Schools, 23,* 329–333.

Washington, J. A., & Craig, H. K. (1992b). Articulation test performances of low-income, African-American preschoolers with communication impairments. *Language, Speech, and Hearing Services in Schools, 23,* 203–207.

Washington, J. A., & Craig, H. K. (1994). Dialectal forms during discourse of poor, urban, African American preschoolers. *Journal of Speech and Hearing Research, 37,* 816–823.

Washington, J. A., & Craig, H. K. (in press). Socioeconomic status and gender influences on children's dialectal variations. *Journal of Speech, Language, and Hearing Research.*

Wells, G. (1985). *Language development in the preschool years.* New York: Cambridge University Press.

Wiener, F. D., Lewnau, L. E., & Erway, E. (1983). Measuring language competency in speakers of Black American English. *Journal of Speech and Hearing Disorders, 48,* 76–84.

Winitz, H. (1959). Language skills of male and female kindergarten children. *Journal of Speech and Hearing Research, 2,* 377–386.

Wolfram, W. (1969). *A sociolinguistic description of Detroit Negro speech.* Washington, DC: Center for Applied Linguistics.

Wolfram, W. (1971). Black and white speech patterns revisited. In W. Wolfram & N. Clarke (Eds.), *Black-white speech relationships* (pp. 139–161). Washington, DC: Center for Applied Linguistics.

Wolfram, W., & Fasold, R. (1974). *The study of social dialects in American English.* Englewood Cliffs, NJ: Prentice Hall.

Contact author: Holly K. Craig, PhD, University of Michigan, 1111 E. Catherine Street, Ann Arbor, MI 48109-2054. Email: hkc@umich.edu