Published in final edited form as:
J Speech Lang Hear Res. 2002 June ; 45(3): 505–518.

Methods for Characterizing Participants' Nonmainstream Dialect Use in Child Language Research

Janna B. Oetting and **Janet L. McDonald**
Louisiana State University Baton Rouge

Abstract

Three different approaches to the characterization of research participants' nonmainstream dialect use can be found in the literature. They include listener judgment ratings, type-based counts of nonmainstream pattern use, and token-based counts. In this paper, we examined these three approaches, as well as shortcuts to these methods, using language samples from 93 children previously described in J. Oetting and J. McDonald (2001). Nonmainstream dialects represented in the samples included rural Louisiana versions of Southern White English (SWE) and Southern African American English (SAAE).

Depending on the method and shortcut used, correct dialect classifications (SWE or SAAE) were made for 88% to 97% of the participants; however, regression algorithms had to be applied to the type- and token-based results to achieve these outcomes. For characterizing the rate at which the participants produced the nonmainstream patterns, the token-based methods were found to be superior to the others, but estimates from all approaches were moderately to highly correlated with each other. When type- and/or token-based methods were used to characterize participants' dialect type and rate, the number of patterns included in the analyses could be substantially reduced without significantly affecting the validity of the outcomes. These findings have important implications for future child language studies that are done within the context of dialect diversity.

Keywords

African American English; dialect; child language

According to Leonard (1998), conducting research in cross-linguistic contexts carries all of the methodological problems of other child language work and a few more. When Leonard wrote this statement, he was referring to child language studies that are conducted within the context of two or more languages, such as Italian and English or Spanish and French. For this type of cross-linguistic work, Leonard notes that some of the problems facing the researcher include finding native speakers of each language who are also experts in child language acquisition, identifying for each language an appropriate set of standardized tests to document a child's normal and/or impaired language status, and developing a method to ensure that children who speak different languages are developmentally similar.

Child language studies that are conducted within the context of more than one dialect of the same language are also cross-linguistic in nature. For these types of studies, an additional problem facing the researcher involves the characterization of each child's dialect. A dialect is defined as any language variety that is shared by a group of people. Dialects that are

Contact author: Janna Oetting, PhD, Louisiana State University, Department of Communication Sciences & Disorders, 163 Music and Dramatic Arts Building, Baton Rouge, LA 70803-2606. cdjanna@lsu.edu.

socially favored are often referred to as standard, and those that are socially stigmatized are often referred to as nonstandard or nonmainstream, even though all dialects of a language have been shown by linguists to be equally complex and systematic (Wolfram & Schilling-Estes, 1998). Most studies in child language have been conducted with participants who are learning a standard dialect of a language. These types of studies are so commonplace that a researcher rarely has to formally confirm and/or measure the participants' dialect use. When nonmainstream dialect speakers are included in research, however, formal documentation of the target dialect(s) is typically required. Nonmainstream dialect use can actually be broken down into two distinct research questions: (a) what nonmainstream dialect is being spoken, and (b) what is the rate of nonmainstream pattern use? Little research attention in the field of child language acquisition has been given to each of these questions. In this paper, we outline why measures of dialect type and rate are problematic for researchers who want to complete languages studies in the context of multiple nonmainstream dialects. Then, we describe three dialect coding methods that can be found in the literature. Following this review, we evaluate the utility of the three methods using an existing data set of child language samples.

Methodological Problems Facing the Researcher

Confirming Study Participants' Type of Nonmainstream Dialect

Describing why one study participant's nonmainstream dialect differs from another's can be extremely difficult for a researcher, especially if morphosyntax is the focus of the investigation. The reason for this is that the same types of surface patterns appear in a number of different nonmainstream varieties. To illustrate the overlapping nature of nonmainstream dialects, consider work by Washington and Craig (1994) and Oetting, Cantrell, and Horohov (1999). In the former study, children's use of a northern urban variety of African American English (AAE) was examined. In the latter, the focus was children's use of a rural variety of Southern White English (SWE). Of the 17 patterns that were coded as AAE by Washington and Craig and the 12 patterns that were coded as SWE by Oetting et al., 11 of them were the same. Moreover, the one SWE pattern that was not on the northern AAE list (i.e., *it/they* for existential *there*) is hardly unique to SWE varieties. To the contrary, Wolfram (1991) describes existential *they* as a southern pattern and existential *it* as a more general pattern that may be spreading.

The lack of a large set of unique pattern types does not threaten the distinctiveness of different nonmainstream varieties such as AAE and SWE because it is not the unique pattern sets that define a dialect. Other important differences between dialects include the ways in which particular patterns co-occur, the relative frequency at which some patterns are produced, and the grammatical entailments of and/or range of meanings expressed by a few select forms (Mufwene, Rickford, Baily, & Baugh, 1998; Wolfram & Estes-Schilling, 1998). Limited sets of unique patterns for different dialects, however, are problematic for researchers who want to study child language issues in the context of dialect diversity. Unlike dialectologists, child language researchers often make comparisons between linguistically defined groups of participants. When these comparisons include normal and impaired children, equivalency across groups also is needed. For these types of studies, the researcher needs to confirm the dialect type of every participant so that normal and impaired groups who speak the same nonmainstream variety can be examined.

Unfortunately, as demonstrated earlier, only a few unique patterns exist for some nonmainstream dialects, and none may exist for others. Use of these forms in conversational speech also can be infrequent. Indeed, of the 45 AAE speakers studied by Washington and Craig (1994), the number who actually produced each of the potentially unique AAE forms was as follows: *fitna/sposta/bouta* = 19, zero possessive = 12, zero present progressive = 9,

zero infinitive *to* = 5, indefinite article = 1, remote past *been* = 0. For Oetting et al. (1999), only 2 of the 31 children produced the one SWE pattern that did not appear on Washington and Craig's northern AAE list. If classifications of dialect type are based on the use of unique patterns, far too many participants would be excluded from research. If shared patterns are used, however, the type of nonmainstream dialect spoken by a participant is not necessarily confirmed. Thus, for determining a speaker's dialect type, the problem facing the researcher is a lack of established methods for distinguishing between speakers of different nonmainstream varieties.

Confirming Study Participants' Dialect Rate

Rates of nonmainstream pattern use vary considerably across speakers; and such factors as age, gender, race, educational achievement, socioeconomic status, region of country, type of community, and type of social network have been found to contribute to this variability (Wolfram & Schilling-Estes, 1998). Across-speaker differences in dialect rate occur even when one tries to control many of the variables listed above. Studies by Washington and Craig (1994) and Oetting et al. (1999) again can be used to illustrate this point. For Washington and Craig's study, all of the AAE speakers were African American and in preschool. All also lived in the same urban community and attended the same Head Start programs in Detroit, MI. Nevertheless, across participants, percentages of utterances with one or more nonmainstream patterns ranged from 0 to 39. For Oetting et al., all of the children were White and ranged in age from 4 to 6 years old, all lived in the same rural community, and all attended the same public schools or child development centers. Nevertheless, the percentages of utterances with one or more nonmainstream patterns in this study ranged from 1 to 20.

As will be demonstrated in the next section, multiple methods have been used to quantify the rate of nonmainstream pattern use by participants in previous studies. Thus, the problem of measuring a speaker's dialect rate is not due to a lack of methods. Instead, the major barrier facing the researcher when addressing this question is the lack of uniformity in the use of dialect rate measures across studies and the labor-intensive nature of rate-based methods.

Three Approaches to Dialect Coding

Three general approaches to dialect coding can be found in the literature. These methods include listener judgments, type-based counts of patterns, and token-based counts. Although all three methods can be used to generate dialect type and rate information, most researchers have used the first two methods to classify participants' dialect type and the third method to index their nonmainstream pattern rate.

Listener Judgment Methods—One approach to the characterization of dialect type involves listener judgments. Dialectologists often use listener judgments to identify research subjects, especially when little is known about the dialect and/or community under study (for a recent example, see Wolfram & Schilling-Estes, 1997). Sociolinguists also use listener judgments to identify target speech communities, but random sampling methods with stratification for such variables as age, sex, race, and so forth also are utilized (Milroy, 1987). Given that nonmainstream pattern use is often the dependent variable of interest within these investigations, additional measures of the target dialect(s) are typically not made for subject selection purposes.

Listener judgment tasks also have been used to examine adults' perceptions of race from speech (e.g., Grinstead, 1987; Haley, 1990; Lass, Mertz, & Kimmel, 1978). Many of these studies provide listeners with 20 seconds or less of an audiotaped sample. In addition,

listeners have had little formal training in the differences that exist between various English varieties. Although a detailed review of these studies is beyond the scope of this paper, speaker characteristics involving race, age, region, and community and listener characteristics involving race, region, and community have been shown to affect outcomes (Haley, 1990). Nevertheless, rates of correct racial identification involving samples from adolescents have ranged from 85% to 92% (Bailey & Bernstein, 1990). This finding is relevant to the current work because it suggests that listener judgments of communicative behavior, as a method, can lead to reliable outcomes.

Within the field of communication sciences and disorders, Seymour and Ralabate's (1985) study is one of the earliest we found that used listener judgments to classify the dialect type of participants. In their study, AA children who spoke AAE and White (W) children who spoke standard English were compared. The participants' dialect type was determined by listener judgments that were made by both authors; children were excluded if both authors did not agree on the type of dialect spoken. Wyatt (1991, 1996) and Cole (1980) also used listener judgments. In Wyatt's work, three clinicians were asked to judge AA children's use of AAE by listening to 3 minutes of an audiotaped sample. A seven-level Likert scale was provided to help the clinicians make their judgments. On this scale, stress and intonation, syntax, semantics, and phonological features of AAE were to be considered. From a pool of 17 children, the 10 who were perceived to produce the highest rate of AAE were chosen for study. Cole's (1980) rating scale involved three levels (high, moderate, low), and her judgments were made using the speech of the research participants' primary caregivers. Children were included in the study if their mothers received moderate to high ratings. In all of these studies, listener judgments were used to confirm the type of nonmainstream dialect spoken by the participants. Dialect rate differences between speakers were not highlighted or evaluated, even though rate-based information was collected by Wyatt and Cole.

Type-Based Methods—Counts of different nonmainstream pattern types also have been used by researchers to confirm a study participant's dialect variety. For example, Seymour, Bland-Stewart, and Green (1998) characterized AA children as AAE speakers if they produced at least three phonological and three syntactic features that matched descriptions of AAE. Jackson's (1998) criteria were similar, except participants were required to produce two, rather than three, patterns of AAE phonology and syntax. Smith, Lee, and McDade (2001) also used a type-based approach to examine linguistic differences between AAE speakers and standard English-speakers. For their study, AA children were selected for the AAE group if they produced at least five nonmainstream AAE patterns, and W children were selected for the standard English group if they produced no more than one nonmainstream pattern. For all three of these studies, a pool of target patterns was not specified, but dialect identification was completed by at least two clinicians who had extensive exposure to, and training in, AAE.

Terrell (1975) and Champion (1995) also used a type-based approach in their studies, but they specified a list of nonmainstream patterns. For Terrell's work, a language sample was considered representative of AAE if it contained at least five out of six possible AAE patterns. Champion's list included 28 AAE patterns, and children were selected if they produced two phonological and/or three syntactic patterns from this list.

Like listener judgment tasks, type-based counts of dialect are relatively easy to do. They also provide information about dialect use that is more objective than listener judgments. However, each of the type-based studies we reviewed counted AAE patterns that also occur in other nonmainstream dialects. Thus, for the researcher who works with multiple nonmainstream dialects, these particular methods may not be useful for distinguishing speakers of different nonmainstream varieties from each other. Note also that type-based

methods have never been used to characterize dialect rate differences among speakers, even though pattern types could be summed for this purpose.

Token-Based Methods—A third method employed by researchers involves token counts of nonmainstream pattern use. Token approaches are designed to provide researchers with information about the rate of a speaker's dialect, but they also provide researchers with some information about the dialect type of a speaker. Three slightly different token methods can be found in the literature. One method involves counting the number of utterances that contain one or more nonmainstream patterns and dividing this number by the total number of utterances analyzed for each child (Craig & Washington, 2000; Dollaghan, Brokaw, & Wunsche, 1999; Jackson & Roberts, 2000; Oetting et al., 1999; Washington & Craig, 1994, 1998).

A slightly different token method was used by Gidney and Deeney (2000). In this study, rate was calculated by counting the number of nonmainstream pattern tokens that a child produced and then dividing this number by the total number of words analyzed. Washington and Craig (1998) and Craig and Washington (2000) also used this method in addition to using token method 1. A third token-based approach was used by Oetting and McDonald (2001) and Jackson and Roberts (2001). In these studies, individual tokens of patterns were counted, but the sum of these was divided by the total number of utterances in the samples rather than by the total number of words.

There are practical differences among the three token-based methods. With the first, researchers need to tag only utterances that contain a nonmainstream pattern. With the others, more coding time is needed because each individual pattern within each utterance must be identified. Table 1 presents 12 hypothetical utterances that are coded with each method to illustrate the ways in which the results from the three approaches also differ. As demonstrated by the first three utterances, the results from the different methods vary in scale, but are parallel when utterance length and nonmainstream pattern frequency are held constant. Each of the next six utterances contains one nonmainstream pattern, but the number of words in each varies. For utterances such as these, methods 1 and 3 result in parallel results, but method 2 yields rates of dialect use that vary. Moreover, when nonmainstream pattern use is held constant, method 2 results in rates of dialect use that decrease as utterance length increases. Thus, as children grow older and their utterances become longer, it is possible that this coding method will underestimate the rate of a child's nonmainstream pattern use.

Another potential problem with method 2 is that it is vulnerable to sample size differences across children. Consider the results of calculating a dialect rate using five, rather than six, of the utterances in this set. Regardless of the utterance that is excluded, estimates from methods 1 and 3 remain unchanged. With method 2, rate varies depending upon which utterance is removed. If the first utterance is excluded, rate of use is .16; but if the last utterance is excluded, the rate is .20. Despite the potential weaknesses of method 2, it is important to note that this approach may be appropriate in some circumstances. For example, when Craig and Washington used this method for calculating dialect rate in their 1998 and 2000 studies, all of their children were in preschool, samples were limited to 50 utterances, and utterances consisted of no more than one independent clause. Thus, for these studies a number of factors that affect sample size and utterance length were tightly controlled.

Finally, the last three utterances in the table are all the same length, but they vary in nonmainstream pattern frequency. The first utterance contains one pattern, the second contains two, and the third contains three. For utterances such as these, results from methods

2 and 3 are parallel, and (as expected) rates of use increase as the number of patterns included within an utterance increases. In contrast, rates of use as calculated by method 1 remain unchanged for these utterances. This result illustrates the lack of sensitivity of method 1 for capturing within-utterance differences in nonmainstream pattern use.

In summary, all three dialect coding approaches have been used by more than one research group. Although all three methods provide a researcher with information about dialect type and rate, listener judgments and type-based methods have been used to confirm participants' dialect type, and token-based methods have been used to describe their dialect rate. Of the token-based approaches, the one treating pattern frequency as the numerator and utterance total as the denominator appears the most useful for samples that vary in size and/or language ability, but this claim is speculative because our analysis is based on a small number of hypothetical utterances.

All but two of the studies reviewed above focused on speakers of AAE, and only one examined more than one nonmainstream dialect. This study was ours (Oetting & McDonald, 2001), and it involved the language samples that are examined here. In our earlier work, race was used to initially group participants, and type of dialect was explored as a dependent variable. What we found were dialectal differences between the W and AA children that were consistent with previous descriptions of adult SWE and SAAE dialect varieties. Specifically, distinct profiles of SWE and SAAE pattern use were identified in the data, even though most of the coded patterns cut across racial boundaries.

Additional analyses with these data have revealed other ways in which the children's use of nonmainstream patterns varies as a function of dialect. For example, we have found that the SWE speakers are more likely to zero mark copula *be* when contexts require *are* as compared to other surface forms, but SAAE speakers are more likely to zero mark contexts requiring *is* and *are*, especially when these forms can be contracted (Wynn, Eyles, & Oetting, 2000). Also, the meanings of *had*+V*ed* expressions differ across the two dialects. For the SWE speakers, these forms carry past perfect meaning like standard English, but for the SAAE speakers, these forms express both past perfect and preterite meaning (Oetting, Ross, & Stapleton, 2001). Both of these findings are consistent with previous dialect work with adolescents/adults (e.g., Rickford & Rafal, 1996; Wolfram, 1974).

Unfortunately, transcription and coding for the Oetting and McDonald (2001) study took several years. An additional year was needed to examine the children's use of copula *be* and *had*+V*ed*. This amount of effort cannot be avoided when the dependent measure of interest is the nonmainstream dialect(s). For researchers who want to examine aspects of communication that are not dialectal in nature, however, requiring a comprehensive inventory of the participants' use of every possible nonmainstream surface pattern and/or an in-depth analysis of one or two particular patterns seems unreasonable, especially if other, less effortful, coding methods can be used. To study this issue more directly, we examined the utility of the different coding methods for characterizing study participants' dialect type and rate. Our specific questions were (a) How well do the three methods reviewed above classify individuals as speakers of either SAAE or SWE? and (b) How well do they capture the variability that exists in the rate at which individuals produce nonmainstream patterns? For both of these questions we also examined whether shortcuts in dialect coding can be made without affecting the validity of the outcomes.

Method

Data

The data consisted of 93 language samples that were collected, transcribed, and coded as part of four previously published studies (Oetting, 1999; Oetting et al., 1999; Oetting & Horohov, 1997; Oetting & McDonald, 2001). Forty of the samples were from AA children, and 53 were from W children; 56 were from males. A third of the samples were elicited from 6-year-olds who were classified as specifically language impaired (SLI). The others were elicited from equal numbers of normally developing age-matched (6N) and language-matched (4N) children. The samples included a total of 20,171 complete and intelligible utterances; the average number of complete and intelligible utterances per sample was 216 (*SD* = 64). The samples also contained coding for 35 different nonmainstream patterns of SWE and/or SAAE (see Appendix A). Details related to the characteristics of the participants and a description of the procedures used to elicit, transcribe, and code the samples can be found in Oetting and McDonald (2001).

Indices of Dialect Use

Listener Judgment—One-minute excerpts of each language sample were copied, in random order, on to four audiotapes (23 excerpts per tape for three; the last tape included 24). A tape of five excerpts from other child samples also was created for the listeners to use as practice. Excerpts were selected from the second half of the samples to maximize the number of child utterances included. The research assistants who copied the tapes were blind to the sociodemographic characteristics of the speakers on the tapes. However, if a personal name and/or other content that might suggest the racial identity of a child was heard during the copying, the assistants were asked to select a different part of the sample. The mean, median, and mode number of utterances per excerpt was 12 (range = 5 to 19).

Three White doctoral students in linguistics (two in progress and one recently graduated) at Louisiana State University independently listened to the excerpts and completed a dialect rating sheet for each one (see Appendix B). Each of the listeners had taken advanced course work in sociolinguistics, and one had completed a dissertation on Louisiana dialects. One was a native resident of Louisiana, another was from Alabama, and the third was from Georgia. The listeners were asked to complete the listener judgments in four different 23- to 24-minute sessions (one tape per session), and they were asked to take a break midway through each session. Throughout the sessions, the listeners were blind to the age, race, sex, and language ability of the children who produced the excerpts.

For each excerpt, the listeners first were asked to make a holistic judgment as to the type and rate of the child's dialect. Two seven-point scales, one for SWE and one for SAAE, were used to facilitate the ratings. On both scales, a score of 1 indicated no use of a nonmainstream dialect and a 7 indicated heavy use. The listeners were then asked to rate the confidence of their decisions using a three-point scale. They also were presented a list of four linguistic features (paralinguistics, phonology, syntax/morphology, and vocabulary) and asked to check the one(s) they used to make their judgments. Finally, they were asked to indicate whether aspects of the tape and/or sample quality hindered their judgment, and they were encouraged to write down comments about the dialect features they perceived.

Type-Based Counts—Two different analyses were completed to examine the usefulness of a type-based approach to dialect coding. First, the percentage of children who produced each of the 35 nonmainstream pattern types was calculated to examine whether certain patterns were used by more children in one dialect group than in the other. Second, the total number of patterns produced by each child was calculated to examine whether these type-

based totals could be used to differentiate the two dialects. These totals also were used to examine the usefulness of a type-based method for dialect rate information. Because 35 different pattern types were originally coded, this was the maximum score a speaker could receive. All patterns involved morphosyntax; phonological patterns were not included in the analysis because the samples were not originally coded for this information.

Token-Based Counts—Three different token-based counts of dialect were completed. They were percent of utterances with one or more nonmainstream patterns, percent of nonmainstream patterns as a function of words spoken, and percent of nonmainstream patterns as a function of utterances spoken. As discussed earlier, each of these methods has been used in previous studies.

Results

The results are reported as a function of the children's race because it provides an independent backdrop from which to compare and contrast the three methods. Grouping children this way to examine SWE and SAAE by no means implies that in this community and elsewhere all W children speak the same way nor that all AA children speak the same way. If this were the case, then there would be no reason to examine different dialect coding methods.

Characterizing Type of Dialect Spoken

Listener Judgment—For 79 (85%) of the 93 excerpts, there was 100% agreement across the three listeners in their selection of either the SWE or SAAE scale for determining the dialect type of each speaker; each of these classifications also was consistent with the corresponding child's race. When consistency was defined as agreement between at least two of the three listeners and the child's race, rate of agreement was 92% (86 of 93 samples). Six samples generated disagreements between more than one rater and the child's race, and one excerpt resulted in 100% agreement across raters, but the dialect classification differed from the child's race.

The excerpts for which more than one rater's dialect classifications differed from the child's race were evenly distributed across the four audiotapes and the three listeners. Disagreements also did not appear to be related to tape quality and/or intelligibility. At least one rater noted these concerns for 33% of the excerpts that generated the disagreements, but at least one rater noted these concerns on 31% of those that did not. Finally, disagreements did not appear to be related to the listeners' confidence estimates. The average confidence rating for the listeners who were in disagreement was 2.08 (SD = .57) on a scale of 1 to 3; the average confidence rating for listeners in agreement was 2.16 (SD =.69).

To further examine these six cases, the listeners were asked to complete the rating sheets again using 10-minute excerpts. This step allowed us to determine if the disagreements were related to the brevity of the excerpts or to the nature of the dialect spoken by the child. The longer excerpts were created using the same procedures as the shorter ones, except names were edited out of the segments because it was too hard to find 10-minute excerpts without them. The longer excerpts resulted in three samples for which dialect classification remained inconsistent across raters and, as mentioned above, one sample for which raters unanimously perceived the child to speak a dialect that differed from his race. In other words, dialect classifications of all but four (97%) of the excerpts were judged in the same way by at least two listeners, and each of these judgments matched the corresponding child's race.

The listeners also were asked to make holistic judgments using all aspects of the excerpt and to indicate which linguistic feature(s) (i.e., paralinguistics, phonology, vocabulary, and

morphology/syntax) they used to make their judgments. The percentages at which the raters used each linguistic feature are presented in Table 2. These percentages do not equal 100 because more than one feature could be used for each excerpt. Across samples, phonology was the most frequently used feature, followed by morphology/syntax, paralinguistics, and vocabulary. Morphosyntax and paralinguistics, but not the other two features, were used significantly more often to identify speakers as SAAE than as SWE: morphosyntax $t(91) = 4.7$, $p < .001$, $\eta^2 = .20$; paralinguistics $t(72) = 2.2$, $p < .05$, $\eta^2 = .06$. Finally, for 31 cases (13 W and 18 AA), cajun/creole flavoring was noted by the listeners. Interviews with the listeners indicated that these impressions were based on phonology. This finding is consistent with others who have studied cajun/creole accent markers in adults (Dubois & Horvath, 1998, 1999).

Type-Based Counts—Two of the 35 coded nonmainstream patterns, *had*+V*ed* and *i'ma* for *i'm going to*, were produced by the AA children only; 50% of the AA children produced the former pattern and 2.5% produced the latter. Two other patterns, completive *done* and use of existential *it/they*, were produced by the W children only, but the percentage of children in this group who produced each of these patterns was low, 3.8 and 5.6 respectively. The remaining 31 patterns were produced by children in both the W and AA groups. The 10 patterns that were produced by the greatest number of children in each group are presented in Table 3. As can be seen, the two lists vary in pattern order and in the percentage of children in this group who produced each form, but 7 of the 10 patterns appeared on both lists. These findings further highlight the extent to which different nonmainstream dialects present overlapping pattern sets, at least for the two nonmainstream dialects studied here.

Before type-based methods are thrown out as viable approaches to the characterization of study participants' dialect type, it is important to consider earlier findings reported in Oetting and McDonald (2001). Within that study using these same samples and token counts of the participants' nonmainstream pattern use, we were able to successfully classify the dialect type of 97% of the children in the SLI and 6N group by using discriminant function analyses. Discriminant analysis is ideally suited for dialect classification work because, with this method, weights are applied to individual patterns so that differences between two groups—in this case differences between the two dialects—are maximized. This method also allows one to exploit differences between dialects that relate to the co-occurrence of patterns and pattern frequency. Given that type-based counts of patterns take less time than token counts, it is worth examining whether a type-based method also can be used with regression algorithms to classify speakers of different nonmainstream varieties.

To run discriminant analyses with the type-based data, a child received a score of one for patterns that were produced at least once. If no tokens were found, the child received a zero for that pattern. Like our earlier study, the entry and deletion criterion was set to .10, and race was used to classify participants as either SWE or SAAE speakers. When all 35 patterns were included in the discriminant function, 95% of all the children were correctly classified as speakers of either SWE or SAAE. Step-wise methods reduced the model to seven patterns (*had*+V*ed*, zero regular past, be_2, zero regular third, S-V agreement with *be*, omission of auxiliary *have*, and alternative pronouns) and resulted in a dialect classification accuracy of 88%.

Token-Based Counts—Table 4 presents means, standard deviations, and ranges of the three token-based methods. As can be seen, there was consistency in the results of all three token-based approaches. Thus, although these methods produced different results when hypothetical data were examined in our literature review, many of the differences disappeared when real samples of 200+ utterances were considered. Unfortunately, with all

three methods, there was overlap in the scores that were generated by the AA children and the W children. Although much of the overlap disappears when the data are separated by child ability (SLI vs. 6N vs. 4N), the lack of clear dialect boundaries across the AA and W groups indicates that token-based methods, by themselves, are not particularly useful for subject selection/matching purposes. For these research purposes, one ideally needs selection criteria to remain stable across various language ability profiles.

Nevertheless, when discriminant analyses were applied to the token counts, methods 2 and 3 generated outcomes that could be used to characterize the nonmainstream dialects of participants. (Method 1 could not be examined because individual patterns were not counted.) As before, race was used to assign the children to a dialect group, and .10 was used as the criterion for entry. When a full model discriminant analysis was completed with data generated from either token method 2 or 3, 97% of the children were correctly classified as speakers of either SWE or SAAE. The step-wise models reduced the patterns to 11 for method 2 and 6 for method 3, and both reduced models resulted in dialect classification accuracies of 95%. Six of the patterns in the reduced algorithms were the same for methods 2 and 3. They were zero *be*, zero regular third, S-V agreement with *be*, zero irregular past, zero present progressive, and *had*+V*ed*. For method 2, the five additional patterns included in the reduced model were: *fixing to, what* for *that*, zero regular past, omission of auxiliary *have*, and be_2. These findings are consistent with what was reported in Oetting and McDonald (2001) when only the SLI and 6N children were considered in the discriminant analysis and token method 3 was used. Adding the younger normal controls and using two different token methods further confirm the use of regression algorithms for dialect classification work.

Across-Method Comparisons—For 90% (84/93) of the participants, dialect classifications were the same across the three methods when 10-minute excerpts and full model discriminant functions were considered. A greater percentage of disagreements and/or misclassifications occurred for children in the SLI (19%) and 4N (10%) groups than in the 6N group (0%), and a greater percentage of disagreements was found for boys (13%) than for girls (5%). The child's race did not appear to influence the percentage of disagreements that were identified (W = 9%, AA = 10%). When results from the one-minute experts and the step-wise analyses were compared, consistency across methods was 80% (74/93). Again, a greater percentage of disagreements across methods occurred for children in the SLI (29%) and 4N (26%) groups than for the 6N group (6%), and a greater percentage of misclassifications also was found for the boys (25%) than for the girls (14%). For the W and AA children, the percentage of disagreements was 24 and 15, respectively. Importantly, though, the dialect type of no single child was misclassified by all three methods. Instead, the inconsistent/misclassified cases varied as a function of the method used.

Characterizing Dialect Rate Differences Between Study Participants

The second question guiding the work related to the usefulness of the three coding methods for examining dialect rate differences between children. Recall that the token methods are ideally suited for this purpose, but these methods are labor intensive. If listener judgments, a type-based method, and/or a shortened version of a token-based method can be used instead, a great deal of time and effort may be saved on the part of the researcher.

The utility of the listener judgment approach was evaluated by calculating the percentage of children who received average dialect ratings at each Likert scale level (see Table 5). Recall that a score of 1 on the scales reflected no use of nonmainstream patterns and a 7 reflected heavy use. Although the scales had a restricted range (1–7), the average ratings of the children did distribute across the levels. Also, many (78%) of the AA children received

average ratings of 5 or greater, whereas many (82%) of the W children received average ratings that were less than 5. Finally, average dialect ratings were higher for the AA children (mean = 5.71, *SD* = 1.04) than for the W children (mean = 4.11, *SD* = .91); $t(91) = 7.8$, $p < .001$, $\eta^2 = .40$.

The utility of the type-based approach was examined by calculating the total number of different patterns each child produced. The type-based scores ranged from 7 to 23 and 5 to 19 for the AA and W children, respectively. Like the listener judgments, average type-based totals were higher for the AA children (mean = 16.58, *SD* = 3.93) than for the W children (mean = 11.34, *SD* = 3.56); $t(91) = 6.7$, $p < .001$, $\eta^2 = .33$.

As demonstrated earlier in Table 4, all three token-based methods also resulted in higher dialect ratings for the AA children than for the W children: Method 1: $t(64) = 10.9$, $p < .001$, $\eta^2 = .64$; method 2: $t(66) = 9.5$, $p < .001$, $\eta^2 = .64$; method 3: $t(58) = 10.1$, $p < .001$, $\eta^2 = .58$. Across participants, token method 3 generated a greater range of scores (3–67) than methods 1 (3–52) and 2 (1–13). Token method 3 also generated a greater range of scores than the listener judgment ratings (2–7) and the type-based totals (5–23). This finding suggests that token method 3 is superior to all of the other methods for rate-based information because a greater range of scores allows individual differences in a data set to be maximized.

Correlational analysis is another way to evaluate the utility of the dialect coding methods. Rates produced by the three token-based methods were highly correlated to each other (*r*s > .92). Rates from token methods 1 and 3 also were moderately to highly correlated to the participants' average listener judgment ratings and their type-based scores (*r*s ranged from .60 to .70). Slightly lower correlations were observed when scores from token method 2 were compared with the listener judgments and type-based counts and when estimates from these latter two approaches were compared to each other (*r*s ranged from .56 to .59). One factor contributing to the lower correlations between token method 2, the type-based counts, and the listener judgments may have been the restricted range of scores generated by these methods. It also is important to remember that the listener judgments were based on the children's use of all linguistic features of dialects as compared to the type and token counts that included morphosyntax only.

Finally, given our interest in identifying shortcuts to dialect coding, we examined whether rates of nonmainstream pattern use could be predicted by counting a few select patterns rather than the full set of 35. Token method 3 was used for this analysis. In total, 10 separate correlations were completed. The first compared the full pattern estimates (i.e., token counts of all 35 patterns) to those generated by counting the tokens of the most frequent pattern, zero *be*. The next correlation compared the full pattern estimates to those generated by counting the tokens of zero *be* and the next most frequent pattern, zero regular third. The rest of the correlations involved adding token counts of the next most frequent pattern until the 10 most frequently produced patterns were included in the set. The order of pattern entry after zero *be* and zero regular third was omission of auxiliary *do*, S-V agreement with *don't*, zero irregular past, multiple negation, S-V agreement with *be*, alternative pronoun, zero regular past, appositive.

As shown in Table 6, when the dialect groups were combined, rate estimates based on token counts of only one pattern, zero *be*, were moderately to highly correlated ($r = .75$) to rates based on token counts of all patterns. Token counts of four patterns (zero *be*, zero regular third, omission of auxiliary *do*, and S-V agreement with *don't*) increased the correlation to .90. Adding more patterns to the set increased the magnitude of the correlation, but gains were incrementally minimal. When the dialect groups were separated, correlations were

lower for both groups, and they were especially low for the SAAE-speakers. Also, more patterns were needed for the SAAE group to reach a correlation that approximated .90 than for the SWE group (SAAE = 10 vs. SWE = 5).

Discussion

Two methodological problems face the researcher who wants to conduct child language research in the context of multiple nonmainstream dialects. The first problem involves a lack of methods for confirming and/or describing study participants' use of different nonmainstream varieties. The current results indicate that one or more of the methods studied here should be used to solve this problem. In the current work, we were able to classify the dialect type of 88% to 97% of the participants, depending upon the method and/or shortcut used. Across methods using the 10-minute excerpts and the full model algorithms, the dialect classifications of 90% of the participants were the same. Rate of agreement across methods was 80% when results from the one-minute excerpts and the reduced algorithms were compared. From a practical standpoint, all of these rates of dialect classification are manageable for research. It is important that these rates are higher than what one would get if unique pattern sets were required for dialect classification purposes. Recall that only 2 (6%) of the participants studied by Oetting et al. (1999) and 0–19 (0%–42%) of the participants studied by Washington and Craig (1994) would have been confirmed as SWE and AAE speakers if one or more unique patterns were required for dialect classification purposes.

Of the three approaches studied here, the listener judgment method is the easiest and quickest to complete because language samples do not need to be transcribed or coded. Therefore, we recommend that in future work researchers consider this method for classifying speakers of different nonmainstream dialects. If morphosyntax is the focus of one's research and language sample analyses are planned, then type and/or token-based counts of patterns that make use of regression algorithms also should be tried. Without discriminant function analyses, however, it is important to understand that the type- and token-based approaches that have been used in previous studies will be unable to distinguish speakers of different nonmainstream varieties from each other.

Type- and token-based counts of pattern use take more time than listener judgment tasks. One way to expedite the coding process is to reduce the number of patterns coded. Current findings indicate that regression algorithms involving type- and/or token-based counts can be run with no more than 11 salient patterns. Our preference would be to use a token-based method in future work because, in the current study, this approach required fewer patterns and led to a higher rate of dialect classification than the type-based method. Of course, with either method, one needs to know what the salient patterns in a dialect are before one can begin making shortcuts. The patterns identified here have been established for two dialects spoken in one rural area of Louisiana. These patterns may or may not be relevant to dialects spoken elsewhere.

The second problem facing the researcher involves methods for estimating dialect rate differences between participants. Recall from the literature review that three different token-based methods have been used for rate-based information, but listener judgments and type-based counts have not. Interestingly, rate estimates from all three token-based methods were highly correlated with each other. Token methods 1 and 3 also were moderately to highly correlated with the listener judgments and type-based totals. Because listener judgments are easy and quick, we recommend that researchers consider this method for rate-based information, especially if the goal is to provide readers with descriptive information about the participants. If the goal of the research is to examine the effect of dialect use on other

child variables, then a more objective and specific measure of nonmainstream pattern use should be considered. For studies of morphosyntax, our preference would be to use token method 3 because it generated the greatest range of scores across the participants. If token method 3 is used, researchers should be able to estimate a participant's relative dialect rate by coding the 10 highest frequency patterns. With large data sets (e.g., $n \sim 100$) that also include multiple dialects and dialect densities, token counts of as few as four patterns may appropriately capture rate differences among speakers.

An obvious caveat about the current study is the need to test the generalization of the findings with other nonmainstream dialects. Given that our data included language varieties that co-exist in the same rural area, we speculate that our findings regarding the usefulness of the three dialect coding methods are conservative. All three methods studied here may prove even more useful in studies that examine nonmainstream varieties from different areas (e.g., urban vs. rural, North vs. South). One also needs to study the usefulness of these methods with participants of different ages. Work by sociolinguists indicates that nonmainstream pattern use is most distinguished in adolescence (Chambers, 1995). Our finding that the dialects of more 6N children were classified in the same way across methods than those in the 4N groups is consistent with this work. We speculate that the current set of methods will work less well with children who are younger than those studied here, but better with those who are older.

Even if the current set of findings are replicated, it is critical to remember that dialect use is a sociolinguistic phenomenon. Nonmainstream pattern use is highly dependent upon a host of external factors as well as those that are internal to the individual speaker. Different environments, genres, topics, conversational partners, and speech acts influence when and how particular forms of language are produced. Any description of a participant's dialect type and rate for research purposes reflects nothing more than a snapshot. For causal-comparative studies, it is important to examine whether dialect measures collected during the subject-selection process carry over to the experimental protocol. And, of course, any measures of nonmainstream pattern use that occur within a language experiment will probably not mirror the everyday language behaviors of the participants. Nevertheless, providing readers with information about study participants' dialect type and rate should facilitate across-study comparisons of findings. This information also will help us, as social scientists, to better represent the people and communities we study.

Future topics for research include learning more about which patterns are the most important to code. In the current work, slightly different subsets of patterns surfaced when shortcuts were identified for classifying the participants' dialect type as compared to their nonmainstream pattern rate. For dialect type, patterns selected by the discriminant analyses were those that maximized differences between the dialects. For dialect rate, patterns were selected based on their frequency of use. Two patterns, zero *be* and zero regular third, surfaced in both analyses. This finding underscores the importance of these two patterns in future studies.

We also want to know more about those participants who present nonmainstream dialects that are difficult to confirm. The nonmainstream variety of the one W child who was unanimously perceived as speaking SAAE is particularly interesting to us. For this child, we completed a separate listener's judgment task as part of Oetting and McDonald (2001). In that study, using a different 10-minute excerpt that occurred earlier in the language sample, the dialect of this child was unanimously identified as SWE. This child was also classified as a SWE speaker in the type-based analyses. Although the first author's perception of this child's entire audiotaped sample is that his dialect falls within the family of SWE dialects that are spoken in Louisiana, determining whether this child's nonmainstream variety

reflects a mix of SWE and SAAE, a different dialect that is somewhere in close proximity to SWE and SAAE, or code switching cannot be determined at this time. To examine this issue, one would need to actively recruit more of these types of participants into research, broaden the analyses to include phonology and paralinguistics, vary the genre and speaking partner, and collect information about social networks and attitudes. Use of one or more of the three coding methods studied here should facilitate this type of work.

Acknowledgments

This project was made possible by a grant from NIDCD awarded to the first author and an Interdepartmental Summer Research Stipend from Louisiana State University awarded to both authors. Appreciation is extended to the parents, children, and staff of Ascension Parish, LA, who made collection of the language transcripts possible; to Megan Melacon, Sarah Ross, and Beth Stapleton who served as the expert listeners; and to Lisa Green who served as an external project consultant. In addition, thanks are extended to Lesley Ellison, Lesley Eyles, Lesli Habans, and Anita Hall who assisted with the dialect coding, and to Whitney Posey, Samantha Simpson, and Christy Wynn who edited the audiotaped stimuli and collected the data for the listener judgment task.

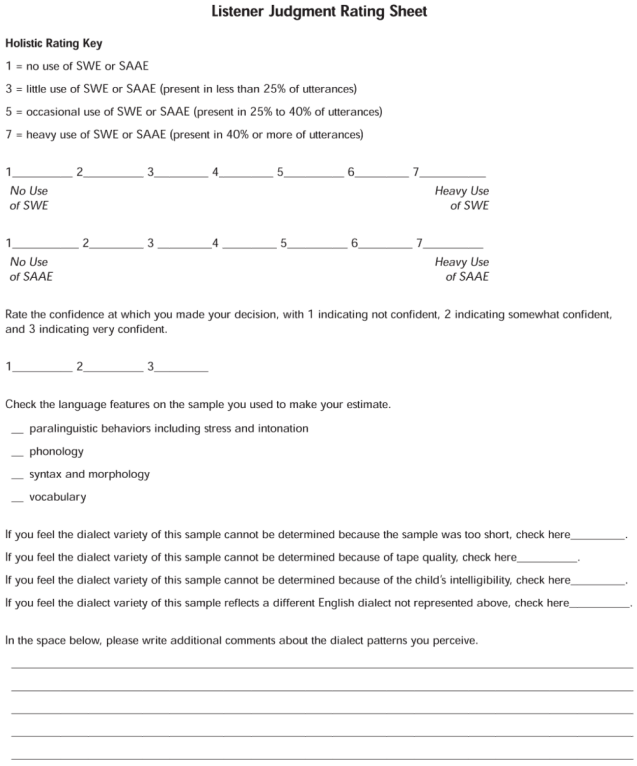
Appendix A

Appendix A

List of 35 Nonmainstream Patterns.
(See Oetting and McDonald, 2000, for descriptions and examples of these nonmainstream patterns.)

zero *be*	overregularization	*been* and *BIN*
be_2 (also referred to as habitual *be)*	participle as past	*done*+verb
i'ma for *i'm going to*	*ain't*	*fixing+verb*
SV agreement with *be*	multiple negation	undifferentiated pronoun
omission of auxiliary *do*	indefinite article	reflexive
omission of auxiliary *have*	zero present progressive	demonstrative
zero regular third	zero plural	dative
zero irregular third	zero possessive	*y'all* varieties
SV agreement with *don't*	zero infinitive *to*	appositive
zero regular past	for to/to	existential *it* and *they*
zero irregular past	zero *of*	Wh- noninversion
had+past	*what/that* or zero *that*	

Appendix B



Listener Judgment Rating Sheet

Holistic Rating Key

1 = no use of SWE or SAAE

3 = little use of SWE or SAAE (present in less than 25% of utterances)

5 = occasional use of SWE or SAAE (present in 25% to 40% of utterances)

7 = heavy use of SWE or SAAE (present in 40% or more of utterances)

1_____ 2_____ 3_____ 4_____ 5_____ 6_____ 7_____

No Use of SWE *Heavy Use of SWE*

1_____ 2_____ 3 _____4 _____ 5_____ 6_____ 7_____

No Use of SAAE *Heavy Use of SAAE*

Rate the confidence at which you made your decision, with 1 indicating not confident, 2 indicating somewhat confident, and 3 indicating very confident.

1_____ 2_____ 3_____

Check the language features on the sample you used to make your estimate.

__ paralinguistic behaviors including stress and intonation

__ phonology

__ syntax and morphology

__ vocabulary

If you feel the dialect variety of this sample cannot be determined because the sample was too short, check here_____.

If you feel the dialect variety of this sample cannot be determined because of tape quality, check here_____.

If you feel the dialect variety of this sample cannot be determined because of the child's intelligibility, check here_____.

If you feel the dialect variety of this sample reflects a different English dialect not represented above, check here_____.

In the space below, please write additional comments about the dialect patterns you perceive.

Figure B.

References

Bailey G, Bernstein C. The idea of Black English. The SECOL Review. 1990; 14:1–24.

Chambers, J. Sociolinguistic theory. Blackwell Publishers; Malden, MA: 1995.

Champion, T. Unpublished dissertation. University of Massachusetts; Amherst: 1995. A description of narrative productions and development in child speakers of African American English.

Cole, L. Unpublished dissertation. Northwestern University; Evanston, IL: 1980. A developmental analysis of social dialect features in the spontaneous language of preschool black children.

Craig H, Washington J. An assessment battery for identifying language impairments in African American children. Journal of Speech, Language, and Hearing Research. 2000; 43:366–379.

Dollaghan, C.; Brokaw, H.; Wunsche, D. Conversation-based measures of language impairment in school-aged children; Poster presented at the 20th Symposium on Research in Child Language Disorders; Madison, WI. 1999;

Dubois S, Horvath B. Let's tink about dat: Interdental fricatives in Cajun English. Language Variation and Change. 1998; 10:245–261.

Dubois S, Horvath B. From accent to marker in Cajun English: A study of dialect formation in process. English World-Wide. 1999; 19:161–188.

Gidney, C.; Deeney, T. Does dialect use relate to phonological awareness in African American children?; Paper presented at the Fourth Biennial Memphis Research Symposium; Memphis, TN. 2000;

Grinstead T. Listeners' responses to regional and ethnic accents in broadcasting. SECOL Review. 1987; 11:115–134.

Haley K. Some complexities in speech identification. The SECOL Review. 1990; 14:101–113.

Jackson, J. Unpublished dissertation. University of Massachusetts; Amherst: 1998. Linguistic aspect in African-American English speaking children: An investigation of aspectual ‘be.

Jackson, S.; Roberts, J. Complex syntax production of African American preschoolers; Paper presented at the Fourth Biennial Memphis Research Symposium; Memphis, TN. 2000;

Jackson S, Roberts J. Complex syntax production of African American preschoolers. Journal of Speech, Language, and Hearing Research. 2001; 44:1083–1096.

Lass N, Mertz P, Kimmel K. The effect of temporal speech alterations on speaker, race, sex identifications. Language and Speech. 1978; 21:279–290. [PubMed: 732401]

Leonard, L. Children with specific language impairment. MIT Press; Cambridge, MA: 1998.

Milroy, L. Language and social networks. Basil Blackwell; New York: 1987.

Mufwene, S.; Rickford, J.; Bailey, G.; Baugh, J., editors. African-American English. Routledge; New York: 1998.

Oetting J. Children with SLI use argument structure cues to learn verbs. Journal of Speech, Language, and Hearing Research. 1999; 42:1261–1275.

Oetting J, Cantrell J, Horohov J. A study of specific language impairment (SLI) in the context of nonstandard dialect. Clinical Linguistics and Phonetics. 1999; 13:25–44.

Oetting J, Horohov J. Past tense marking by children with and without specific language impairment. Journal of Speech, Language, and Hearing Research. 1997; 40:62–74.

Oetting J, McDonald J. Nonmainstream dialect use and specific language impairment. Journal of Speech, Language, and Hearing Research. 2001; 44:207–223.

Oetting, J.; Ross, S.; Stapleton, B. Had+Ved: A narrative discourse structure in AAE. Poster presented at the annual convention of the American Speech-Language-Hearing Association; New Orleans, LA: 2001.

Rickford J, Rafal C. Preterite had+V-ed in the narratives of African American Preadolescents. American Speech. 1996; 71:227–254.

Seymour H, Bland-Stewart L, Green L. Difference versus deficit in child African American English. Language, Speech, and Hearing Services in Schools. 1998; 29:96–108.

Seymour H, Ralabate P. The acquisition of a phonological feature of Black English. Journal of Communication Disorders. 1985; 18:139–148. [PubMed: 3988978]

Smith T, Lee E, McDade H. An investigation of T-units in African American English-speaking and standard American English-speaking fourth grade children. Communication Disorders Quarterly. 2001; 22:148–157.

Terrell F. Dialectal differences between middle-class Black and White children who do and do not associate with lower-class Black children. Language and Speech. 1975; 20:377–383. [PubMed: 616514]

Washington J, Craig H. Dialectal forms during discourse of poor, urban, African American Preschoolers. Journal of Speech and Hearing Research. 1994; 37:816–823. [PubMed: 7967568]

Washington J, Craig H. Socioeconomic status and gender influences on children’s dialectal variations. Journal of Speech and Hearing Research. 1998; 41:618–626.

Wolfram W. The relationship of White southern speech to Vernacular Black English. Language. 1974; 50:498–527.

Wolfram, W. Dialects and American English. Prentice Hall; Englewood Cliffs, NJ: 1991.

Wolfram, W.; Schilling-Estes, N. Hoi toide on the Outer Banks. University of North Carolina Press; Raleigh, NC: 1997.

Wolfram, W.; Schilling-Estes, N. American English. Blackwell Publishers, Inc; Malden, MA: 1998.

Wyatt, T. Unpublished dissertation. University of Massachusetts; Amherst: 1991. Linguistic constraints on copula production in Black English child speech.

Wyatt, T. Acquisition of the African American English copula. In: Kamhi, A.; Pollock, K.; Harris, J., editors. Communication development and disorders in African American children. Paul Brooks; Baltimore, MD: 1996. p. 95-116.

Wynn, C.; Eyles, L.; Oetting, J. Nonmainstream dialect use and copula be: Analysis of linguistic constraints. Poster presented at the annual convention of the American Speech-Language-Hearing Association; Washington, DC: 2000.

Table 1

Three token-based methods for coding nonmainstream dialect use.

Example Utterances	Token-Based Approach 1[a]	Token-Based Approach 2[b]	Token-Based Approach 3[c]
Oscar in the can.	100	.25	1
Oscar is in the can.	0	0	0
Him and John run.	100	.25	1
What you did?	100	.33	1
How you get up?	100	.25	1
I bought me a book.	100	.20	1
My daddy once went by hisself.	100	.17	1
Course I brung him up real fast.	100	.14	1
One day I had went on the elevator way up.	100	.01	1
He don’t got a car.	100	.20	1
He don't got no car.	100	.40	2
Him don't got no car.	100	.60	3

[a] Percent of utterances with one or more pattern

[b] Percent of patterns produced as a function of the number of words spoken

[c] Percent of patterns produced as a function of the number of utterances spoken

Table 2

Use of linguistic features by raters within listener judgment task.

	Phonology (%)	Morphology Syntax (%)	Paralinguistics (%)	Vocabulary (%)
Overall	86	61	41	24
W Samples (n = 53)	83	48	36	26
AA Samples (n = 40)	90	78	48	22

Table 3

Percent of children in each group who produced each pattern.[a]

W (*n* = 53)	AA (*n* = 40)
zero be (89%)*	zero be (100%)*
multiple negation (72%)*	zero regular third (100%)*
zero regular third (70%)*	zero regular past (90%)
omission of do (66%)*	S-V agreement with be (85%)*
S-V agreement with don't (60%)*	multiple negation (82%)*
S-V agreement with be (58%)*	S-V agreement with don't (78%)*
appositives (57%)	zero irregular past (75%)*
over-regularization of regular past (55%)	omission of do (70%)*
zero irregular past (51%)*	zero irregular third person (70%)
alternative pronoun form (51%)	zero possessive (68%)

[a]The 10 most frequently used pattern types for each group are listed in descending order. Patterns that appear on both lists are indicated by *.

Table 4

Token-based indices of nonmainstream dialect.[a]

	W (*n* = 53)	AA (*n* = 40)	Groups combined
Percent of utterances with one	12	29	19
or more nonmainstream dialect	(6)	(9)	(11)
	3–35	10–52	3–52
Percent of dialect patterns out	3	7	5
of total words	(2)	(2)	(3)
	1–8	3–13	1–13
Percent of dialect patterns out	13	34	22
of total utterances spoken	(7)	(12)	(14)
	3–42	11–67	3–67

[a]Means are presented first, standard deviations are presented second, and ranges are presented third.

Table 5

Percent of children in each group as a function of their average Likert scale rating.

Average ratings	W	AA
1–1.99	0	0
2–2.99	6	0
3–3.99	40	10
4–4.99	36	12
5–5.99	13	22
6–6.99	5	48
7	0	8

Table 6

Correlation coefficients between rates of use based on set of 35 patterns and rates of use based on reduced sets of patterns.

	Number of patterns in set[a]									
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
Counts of All 35 Patterns										
Groups Combined	.75	.88	.88	.90	.90	.91	.93	.94	.94	.96
SWE Group Only	.79	.81	.88	.88	.87	.91	.93	.96	.96	.97
SAAE Group Only	.36	.66	.66	.66	.73	.74	.80	.84	.84	.88

[a]Patterns included in the reduced set are, in order: zero *be*, zero regular third, omission of auxiliary *do*, S-V agreement with *don't*, zero irregular past, multiple negation, S-V agreement with *be*, alternative pronoun, zero regular past, appositive.

J Speech Lang Hear Res. Author manuscript; available in PMC 2012 July 05.