Theoretical and Empirical Bases for Dialect-Neutral Language Assessment: Contributions from Theoretical and Applied Linguistics to Communication Disorders

Barbara Zurer Pearson, Ph.D.¹

ABSTRACT

Three avenues of theoretical research provide insights for discovering abstract properties of language that are subject to disorder and amenable to assessment: (1) the study of universal grammar and its acquisition; (2) descriptions of African American English (AAE) Syntax, Semantics, and Phonology within theoretical linguistics; and (3) the study of specific language impairment (SLI) cross-linguistically. Abstract linguistic concepts were translated into a set of assessment protocols that were used to establish normative data on language acquisition (developmental milestones) in typically developing AAE children ages 4 to 9 years. Testing AAE-speaking language impaired (LI) children and both typically developing (TD) and LI Mainstream American English (MAE)-learning children on these same measures provided the data to select assessments for which (1) TD MAE and AAE children performed the same, and (2) TD performance was reliably different from LI performance in both dialect groups.

KEYWORDS: Universal grammar, unique structures of AAE syntax, cross-linguistic study of LI, developmental milestones

Learning Outcomes: As a result of this activity, the participant will be able to identify (1) the theoretical bases for a dialect-sensitive language screener and a dialect-neutral language assessment, and (2) the goals considered to be essential in the selection of the items for the proposed screener and diagnostic language assessments.
**THE THEORETICAL FRAMEWORK**

The need for a linguistically and culturally fair language test in communication disorders has been recognized since the 1960s, but until now there have been no standardized speech and language instruments specifically designed and constructed for learners of dialects other than Mainstream American English (MAE). In the interval, research in several fields has contributed to the conceptual basis for such a test. These research traditions are (1) the study of universal grammar and its acquisition within theoretical linguistics; (2) descriptions of African American English (AAE) Syntax, Semantics, Pragmatics, and Phonology within theoretical linguistics; and (3) the study of specific language impairment (SLI) cross-linguistically. Seymour, Roeper, de Villiers, and de Villiers (this issue) used insights from all three fields to create a practical instrument that takes into consideration typical development across dialects, and also the abstract properties of grammars to diagnose language disorder.

As described by Seymour in the introduction to this issue, following the social revolutions of the 1960s, sociolinguists gave us two important legacies. The first is an appreciation of the rule-governed nature of AAE, demonstrating that it is not a defective version of Mainstream English, but a logical and systematic language variation in its own right. Sociolinguists of the 1960s and 1970s also devised a useful inventory of points of contrast between MAE and AAE, which has been the basis of significant research in the intervening time. For example, if a language had noun phrases with the noun at the end of the phrase (such as Japanese), it would affect the typical word order of sentences in the language, the types and position of relationship words, such as prepositions, and the structure of relative and other adjunct clauses in the language. There are many other examples of the integral relatedness of parts of grammar within contemporary linguistics, and these informed the kinds of language probes described in this issue.

In the abstract study of phonology, likewise, Optimality Theory provides a mechanism for us to view the greater underlying commonality between languages and dialects, while at the same time appreciating their surface differences. In Optimality Theory, researchers posit a restricted set of rules shared by all languages. Different languages “weight” the universal rules differently—that is, apply them in different orders or with different priorities—to give rise to many different outputs observed in different languages and dialects.

With these frameworks as a background, the authors of this issue were motivated to look for underlying language principles as the foundation for the test, and as ways to diagnose disorder that would not confuse its signs with surface differences due to dialect. One such area, for example, is in the fundamental properties of movement rules and how movement is blocked, exemplified by *wh*-questions (see Roeper). The theory behind syntactic barrier items, found in the discussion of syntax evaluation, derives from Ross’s work on “islands,” later elaborated by Chomsky in Barriers. These principles were applied to child language acquisition theory by Roeper and de Villiers and their colleagues. These principles were applied to child language acquisition theory by Roeper and de Villiers and their colleagues.

**Universal Grammar**

During the same period, abstract theoretical linguistics was elaborating a framework focused on a more abstract level of grammatical description. Linguists such as Chomsky and colleagues posited a system of universal grammar within which all languages share a common underlying set of principles, but differ by how settings of certain parameters, or “switches,” are set. This “principles and parameters” research provided a mechanism for understanding how small differences in the fundamental structures of languages would have effects in many different areas of the surface structure of languages.

**Research on the Abstract Structure of AAE**

Our work has also profited from greater exploration of AAE within the framework of abstract linguistics. *Wh*-barriers, for example, were investigated in an AAE-learning population, and typically developing AAE-learning children were shown to respect the universal principles involved, much the same as children learning MAE or French or Greek, for
example. In 1998, Coles extended that work to provide new insights into the mechanisms governing negative concord, another AAE property, whereas Green investigated the structural description of the AAE verb phrase. Green’s work provides a possible explanation for why AAE verbs would not receive person marking (as in third-person –s) in the same way as MAE verbs. As Green points out, the AAE verbal system has a rich system of “aspectual” elements that have no counterpart in the grammar of MAE. (“Aspect” refers to other means aside from tense to describe the time element in an action—whether it was a continuing action, or a repeated action, one that happened all at once, or one that lasted a long time; e.g., “the leaf fell” versus “the leaf was falling.”) For example, AAE has special meanings for verb auxiliaries, such as invariant be for habitual actions (“they [always] be running”), done for perfective, been done for remote past—all shades of meaning that require adverbs to express them in MAE. Green posits an additional branch in the AAE structural tree. As in Russian, which also has a richly marked system of aspect in its grammar, the aspect particles, which appear superficially the same as MAE tense markings, do not interact with the agreement system. If Green is right, this would give a deep structure reason for the lack of person agreement in the AAE verb phrase.

Other theoretically informed studies of AAE have been carried out by Terry, Jackson, Wyatt, Dayton, and others. Although not all of these findings have found direct translation into assessment probes, they emphasize the regularities in certain features of AAE, which were often considered instances of imperfect learning of MAE structures. In this framework, rather than being imperfect, the use of these same features can be viewed as learning an alternative grammar with properties as rich and complex as any other natural language.

Cross-Linguistic Study of SLI

Another important avenue of research is found in the recent interest in language impairment (LI) in languages other than English. The study of LI and SLI in different languages adds to our understanding of the deeper principles of language involved in such impairment. For example, much research on LI in English-learning children has focused on the acquisition of morphosyntax, and several theories attributed the difficulty to a problem with, for example, unstressed words and parts of words. More recently, inspired by modern linguistic analyses of the functional categories, Rice and Wexler have pointed to a particular difficulty in SLI in the area of tense. One of the recommendations of this line of work is to take a measure of a “finite verb morphology composite.” This measure represents the overall percentage with which children use morphemes such as past -ed, present third person singular -s, and copula and auxiliary be forms.

Leonard concludes, “For many children with SLI, the use of this collection of morphemes is unusually weak.” However, researchers studying children learning languages with more developed systems of morphosyntax inflections report little difficulty with tense marking. Instead, other difficulties come to light; for example, in agreement, as observed by Clahsen or the incorporation of pronoun forms (clitics) into the verb phrase, as in French. The problems with tense forms observed in English LI may be less a question of the morphology of tense, but rather a consequence of a more general problem in the elaboration of the verb phrase as a whole, which takes different forms depending on the computations required by the structural properties of the different languages.

In the case of AAE, even normally developing children measured by a standard such as the “finite verb composite” of MAE would be identified as having serious linguistic problems, because AAE is different in each one of these forms. As researchers, we sought to avoid confusing dialect variation and disorder and to capture the sophisticated abstract knowledge that all typically developing children learn without being taught, but which create problems for children with language impairment. We do not yet know why these problems might arise—whether it is, as has been suggested, a question of incomplete grammars, or processing difficulties, or missed parameters. Perhaps in SLI, the normal components of language...
that feed growth to one another get into misalignment through delay in some area. Each of these, or yet another possibility, may be at work for different subsets of children with disorders. The theoretical basis for SLI is still an issue with many unresolved questions. Our final set of suggested tasks attempts to delve more deeply into areas where the child could have fundamental problems.

THE EMPIRICAL RESEARCH PROGRAM

Step 1: Developing Items

One way to go about translating theoretical background into an empirical research program that would identify dialect and culture-free assessment items was to mine protocols from acquisition experiments within Syntax, Semantics, Pragmatics, and Phonology for candidate items. Several of de Villiers and Roeper’s experiments provided formats for our question probe to identify language impairment. Work in the acquisition of passive by Roeper, question-asking by de Villiers, quantifiers by Mattei and Roeper, articles by Schafer and de Villiers, and lexical contrasts by Waxman and Hatch, for example, all suggested still other formats in the different language domains. During the initial National Institutes of Health Screener grant and the first phase of a contract to Seymour for a comprehensive language test, items were adapted from experiments and tried out with AAE learners in Hartford, CT, and Springfield, MA, to establish their utility before being submitted to nationwide field testing.

The goal of the piloting and then the first round of field testing was to narrow down the candidate items to the most effective ones. We had two objectives in mind, Dialect Identification and Diagnosis of Language Impairment; thus, useful items were considered to be those that could be embodied in clear-cut paper-and-pencil-based format and would meet these criteria:

1. they showed steady development across age

For identifier items,
2. they showed a clear distinction between dialect groups

For diagnostic items,
3. they were neutral with respect to the dialect spoken, and
4. there was a clear distinction between typical and disordered performance.

Step 2: Field Testing/Developmental Milestones

The field testing research for this experimental test also served as the data collection for determining developmental milestones for AAE learners. It aimed to establish the milestones of development, a metaphor for the road markers that tell the distances between what are here “landmarks” of language behavior. At what age do we expect a typically developing child to understand how to give an exhaustive response to a double question? At what age do we expect most children to use the language for thought to tell about a character’s actions and motivations in a short narrative? In addition, by extension, at what age is it a sign of delay not to do so?

Despite progress understanding the unique structures of AAE adult language, no comprehensive picture of AAE child language and the stages of its development has yet emerged. Therefore, the developmental milestone research was designed primarily to find the course of mastery in the AAE population for the noncontrastive items in the set; that is, features shared with MAE, such as past tense copula were or rules for embedding clauses. Neither has the prevalence of contrastive items in the speech of AAE-learning children yet been established across a range of ages. Thus, a portion of the developmental milestone research was devoted to determining which AAE features are most persistent, and at what ages and in what contexts AAE features are a significant factor in AAE-child speech.

In theory, children speaking either AAE or MAE should find items based on noncontrastive features equally easy or hard; but there are several reasons why the dialects could differ even on noncontrastive features in development. Features might be similar yet not
identical, or they might not occur with equal frequency. They may interact differently with other features that are contrastive. For example, the tendency for plural nouns to appear without an “-s” marking (e.g., “two cup”) in AAE may affect the child’s likelihood of producing a bare noun as a singular (e.g., “cup” instead of “the cup”). Or, if the third-person /-s/ is absent in AAE, then number agreement may be a much less salient property overall. What is the impact of that on the use of quantifiers, such as *every*, which takes a singular noun, versus *all*, which generally modifies a plural noun? The AAE-speaking child may find it easier than the MAE child to treat the two quantifiers alike, and then take longer to recognize the special properties of *every*. It is not clear, or even likely, that equivalence of structures in the adult dialects would always translate into a similar schedule for learning those structures in the child language of both dialects.

**Tryout Research/Developmental Milestones Procedures**

For this research phase it was important to obtain a diverse sample of 1257 children from across the United States (see Table 1). To accomplish this, 477 speech-language pathologists [footnote to TPC] were recruited to assist in “trying out” more than 300 test items. The items were divided into 14 subdomains, and they assessed more than 30 different language constructs.

**Table 1  Research Sample for the Developmental Milestones Research**

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>AAE</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TD</td>
<td>108</td>
<td>131</td>
<td>143</td>
<td>15</td>
<td>33</td>
<td>27</td>
<td>37</td>
<td>19</td>
<td>27</td>
<td>540</td>
</tr>
<tr>
<td>LI</td>
<td>27</td>
<td>45</td>
<td>50</td>
<td>16</td>
<td>29</td>
<td>15</td>
<td>30</td>
<td>14</td>
<td>16</td>
<td>242</td>
</tr>
<tr>
<td><strong>MAE</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TD</td>
<td>60</td>
<td>61</td>
<td>73</td>
<td>16</td>
<td>20</td>
<td>20</td>
<td>21</td>
<td>18</td>
<td>31</td>
<td>320</td>
</tr>
<tr>
<td>LI</td>
<td>21</td>
<td>29</td>
<td>34</td>
<td>9</td>
<td>19</td>
<td>12</td>
<td>14</td>
<td>6</td>
<td>11</td>
<td>155</td>
</tr>
<tr>
<td>Total</td>
<td>216</td>
<td>266</td>
<td>300</td>
<td>56</td>
<td>101</td>
<td>74</td>
<td>102</td>
<td>57</td>
<td>85</td>
<td>1257</td>
</tr>
</tbody>
</table>

AAE, African American English; TD, typically developing; LI, language impaired; MAE, Mainstream American English.

**Characteristics of the Sample**

Given that the original goal of the project was to find a means to identify AAE-speaking children at risk for language impairment as they entered school, the focus of the tryout research was children ages 4 to 6 years, who comprised approximately 65% of the sample. The items were also tested on children ages 7 through 12 years to ensure that ceilings were reached for the different target behaviors. It was also important to test proportional numbers of children of each gender living in different regions of the country to ensure that any patterns found were not limited to only one gender or region.

**DISCRIMINATION**

Equally important was to discover which items were capable of discriminating disorder through the age range, so the same items were tried out on 250 African American children diagnosed with and receiving services for language impairment. (There was also a group of 147, including 16 additional children not in the main study, who had been identified as phonologically impaired to test the noncontrastive, diagnostic phonology items.) This aspect of the developmental milestones data collection allowed the selection of items that followed a different path of development in the typically developing (TD) and LI children. In the end, a set of items was found that showed steady development through age 9 years (ceiling values were reached at 10 years for the majority of them, even among LI children).
Because the prevalence of AAE speakers and the density of their dialect was thought to be greatest at lower economic and educational levels, the greatest portion of the children (79%) were from families where the parents had only a high school education or less. Finally, given that the items included many innovative elements never before included on a language test, they were also tried out on a comparison set of 475 MAE-speaking children, both TD and LI. These children were matched as closely as possible to the AAE sample with respect to economic and educational level. As in the AAE group, about one third of the children were diagnosed with and receiving services for language impairment. Items that showed different performance between the two dialect groups were eliminated from the pool of candidate items.

There were 49% females overall, more or less evenly distributed throughout the age ranges. (One exception was that the LI children were 61% male; to keep the general balance, the TD groups had more females than males, 55% and 45%, respectively.) The regional distribution of participants was generally based on the distribution of African Americans in the different parts of the country according to the Current Population Survey of the Census Bureau of October 2000 (Table 2).

RESULTS
The broad results of the milestone research are stunning. They show that typically developing AAE and MAE speakers perform similarly with respect to many sophisticated syntactic and semantic aspects of their grammar.

In analyses of variance, for example, *wh*-questions (WH), articles (AR), and fast mapping (FM), graphed in Figures 1–3, showed extremely large age effects, but no effect of dialect or interaction of age and dialect. [WH: Age, $F(5, 1002) = 46.427, p < 0.0001$; Dialect, $F(1, 1002) = 0.380, p = 0.538$; Age by Dialect, $F(5, 1002) = 0.385, p = 0.859$. AR: Age, $F(5, 1002) = 41.268, p < 0.0001$; Dialect, $F(1, 1002) = 0.006, p = 0.937$; Age by Dialect, $F(5, 1002) = 0.441, p = 0.820$. FM: Age, $F(5, 1002) = 29.600, p < 0.0001$; Dialect, $F(1, 1002) = 0.361, p = 0.548$; Age by Dialect, $F(5, 1002) = 0.603, p = 0.670$.]

In addition, from the point of view of pragmatics in communicative role-taking (CR) and question asking (QA), both groups show that they are developing an age-appropriate sense of how language is used to give and receive specific information (Figs. 4 and 5). [CR: Age, $F(5, 1002) = 46.901, p < 0.0001$; Dialect, $F(1, 1002) = 0.025, p = 0.875$; Age by Dialect, $F(5, 1002) = 0.620, p = 0.685$.]

![Figure 1 Wh-question scores by Dialect group. AAE, African American English; MAE, Mainstream American English.](image-url)
Figure 2  Article scores by Dialect group. AAE, African American English; MAE, Mainstream American English.

Figure 3  Fast mapping novel verbs by Dialect group. AAE, African American English; MAE, Mainstream American English.

Figure 4  Question asking scores by Dialect group. AAE, African American English; MAE, Mainstream American English.
The milestone data also make it clear that there is a small set of predictable morphosyntactic and phonological elements, the language variation score (LVS), that distinguish TD AAE speakers from MAE speakers. In this case (Fig. 6), the strong dialect effect diminishes slightly with age and so one sees a just barely significant interaction. [LVS: Age, $F(5, 696) = 15.748, p < 0.0001$; Dialect, $F(1, 707) = 337.615, p < 0.0001$; Age by Dialect, $F(5, 696) = 2.274, p = 0.046$.] This is not to say that the AAE grammar is characterized only by this small set of features. Rather, these features pattern with the deeper aspects of AAE grammar—negative concord and the rich aspectual and phonotactic systems—so the surface features can be used as proxies for the deeper aspects when it is useful to clarify quickly and efficiently whether the individual is an MAE speaker or not.

Finally, the milestones for these same tasks are significantly different between LI and TD children in all four domains (Figs. 7–10): Syntax (SYN), Pragmatics (PRG), Semantics (SEM), and Phonology (PHO; the comparison for phonology between children with phonological impairment and those with no impairment) (Figs. 7–10). [SYN: Age, $F(5, 1002) = 89.140, p < 0.0001$; Clinical Status, $F(1, 1002) = 114.77, p < 0.0001$; Age by Clinical Status, $F(5, 1002) = .765, p = 0.575$. PRG: Age,
Figure 7  Syntax domain by Clinical Status.

Figure 8  Pragmatics domain by Clinical Status.

Figure 9  Semantics domain by Clinical Status.
MAJOR FINDINGS

The developmental milestones/tryout research described above provided a wealth of data on the language patterns of the four groups of children studied: TD AAE, LI AAE, TD MAE, and LI MAE. The findings on the 4- to 6-year-old children are summarized in the April 2002 report to the National Institutes of Health, portions of which are available from the authors. Additional patterns are being investigated in articles in preparation by Roeper, Strauss, Jackson, Johnson, de Villiers, Pearson, and Vellemann, as well as others.10,11,53

The major findings from the research we conducted and report on in the articles in this issue can be summarized as follows:

1. A set of 10 contrastive morphosyntactic and 5 contrastive phonological features reliably distinguish AAE from MAE speakers among TD children aged from 4 to 12 years.

2. A set of 17 noncontrastive items (7 morphosyntax, 4 wh-comprehension, and 6 nonword repetition) reliably distinguish 4 levels of risk for language delay for children aged 4 to 9 years, regardless of the child’s dialect.

3. A set of 25 noncontrastive Phonology items reliably identify risk for speech disorder, regardless of the child’s dialect.

4. One hundred sixteen noncontrastive items (46 in Semantics, 28 in Syntax, 17 in Pragmatics, plus the 25 Phonology items) reliably distinguish typical development from language disorder, regardless of the dialect of the child.

ACKNOWLEDGMENTS

This work was funded in part by National Institutes of Health (NIDCD) under Contract # N01 DC8-2104 and Grant # R01 DC 02172-04 to Harry Seymour, Principal Investigator, at the University of Massachusetts Amherst, with Thomas Roeper and Jill de Villiers at the University of Massachusetts and Smith College, as co-investigators. It was accomplished in conjunction with The Psychological Corporation of Harcourt Assessment, Inc., San Antonio, TX.

The tests that are the products of this research collaboration are the Diagnostic Evaluation of Language Variation (DELV) assessments, the DELV Screening Test, DELV Criterion-Referenced edition, and the DELV Norm-Referenced edition. The phrase “evaluating language variation” refers generally to the assessment processes discussed in this issue.
The term \textit{DELV} is the name trademarked by The Psychological Corporation of Harcourt Assessment, Inc., and refers to the specific tests that are the outcome of the extensive research described in this article. The specific tests are referred to as the \textit{DELV-ST}, or “screener,” or the \textit{DELV-CR}, \textit{DELV-NR}, or the “full diagnostic test,” as appropriate. Questions about the principles underlying the tests can be referred to the authors of this issue (Seymour, Roeper, de Villiers, de Villiers, Pearson, and Cioll). Questions about the tests themselves should be addressed to the Project Leader at The Psychological Corporation of Harcourt Assessment, Inc.; Lois Cioll, Senior Research Director.

\section*{REFERENCES}

32. Jakubowicz C. About the nature of the linguistic deficit in SLI. Paper presented at: Symposium on the Acquisition of Functional Categories in French across Learner Contexts, IASCL, University of Wisconsin, Madison, WI; July 16–21, 2002
41. Philip W. Event quantification in the acquisition of universal quantification. Dissertation, University of Massachusetts, Amherst, MA; 1995