Cultural and Linguistic Fairness in the Assessment of Semantics

Jill G. de Villiers, Ph.D. ¹

ABSTRACT

In creating a semantic assessment that will be linguistically and culturally fair, it is important to avoid the bias of acquired vocabulary tests. This article describes techniques to assess children's processing of new words, their lexical organization and retrieval, which may be more significant than the number of words known. Special properties of the quantifier every also give an idea of the child's level of semantic ability. In addition, we examine how the child's performance on each item-type helps the speech-language pathologist determine the sources of children's problems in this domain.

KEYWORDS: Lexical organization, fast mapping (novel verbs), transitive structures, transfer structures, complement structures, acquired vocabulary, quantifiers, argument structure, learning from context

Learning Outcomes: As a result of this activity, the participant will be able to identify (1) how the semantic items address the issues of linguistic and cultural bias, and (2) how the proposed semantic tasks differ from semantic tasks on traditional language tests.

KEY CONCEPTS

Semantics is central to language acquisition; it is about the expression and understanding of meaning. It is usually construed as being about the mental dictionary of words, or lexicon, that the child has to acquire. Word learning is crucial for a language user, and it needs to be fast and efficient. Not only that, but once words are learned, they need to be stored in an organized way so retrieval is efficient. However, semantics is broader than the lexicon alone in that it also includes the special properties of how words such as quantifiers (every, all, some, none) interact in sentences.

Point 1: Words Depend on Input; Input Depends on Culture

How does the assessment of semantics impact on culture- and dialect-fairness? First, word
learning, more than any other domain, is highly subject to frequency in the input to the child. After all, words are arbitrary symbols, so there is no way the child can use innate knowledge to guess what a chair might be called. But the consequence is that children growing up in different households, in different conditions of input, will develop vocabularies that differ from each other, and grow possibly at different speeds. Some families may spend a lot of time reading picture books exposing the child to words outside of their usual experience, such as hippo, or octopus, or volcano. Another family may spend more time in discourse about family events, talking about relationships, emotions, or activities, which would expose the child to words such as great-grandma, college, promotion, and pride. Yet another family may engage their young children in chores such as shelling peas or shucking corn, in which case the child may develop vocabulary of a rich sort about a limited domain such as the garden: daddy long-legs, husk, pod, and compost. We were concerned that many previous studies had suggested that African American children reveal a more limited vocabulary on vocabulary tests, usually picture-based choice tests. These tests in general have been standardized on a sample representing very few such children. It is a fact of American life that although African American children may go to some of the same schools and live in some of the same neighborhoods as their white counterparts, there is nevertheless a rich culture that is distinctive of African American families. Participation in that different culture is likely to give the African American English (AAE)-speaking child exposure not only to a different vocabulary but a different emphasis on which words are central to childhood. The picture-based tests are often heavily biased toward nouns, because verbs are harder to capture in a still picture. Yet Blake has argued that verbs may form a more important part of the vocabulary for AAE-speaking children, just as they do for some languages such as Korean. Thus, the reduced vocabulary seen on standardized tests may be a biased estimate of AAE-speaking children’s competence in semantics.

The outcome on acquired vocabulary tests may tell us that a particular child is less likely than another to be prepared for schooling, especially if school requires the same type of vocabulary as the test. However, being culturally different is not the same as having a language disorder, as the American Speech and Hearing Association has been at pains to point out for several decades.

Point 2: Words Are Fast Mapped from Rich Contexts

Parents and teachers working with a very young child may use pointing and single word use to name a new object, such as train. However, once the child gets to be a toddler or older, much vocabulary is learned “on the fly” from conversational context. A new word is embedded in a sentence, and the child must identify the new word and use cues from the sentence to home in on what part of speech it is. Knowing that will allow the child to scan the context for something corresponding to a new action, or a new object or attribute. Brown showed that 3- to 5-year-old children were sensitive to the sentence context in picking out different meanings for sib across forms, such as:

Look, a sib!
Look, sibbing!
Look, some sib!

Recent work by Waxman and her colleagues has shown that children also can differentiate adjectives from nouns using linguistic context:

Look, a sib!
Look, a sib one!

Naigles and Gleitman have pointed out how significant the sentence context is in learning a verb, in that the child can guess what kind of event is being referred to from sentence context. An intransitive verb such as:

He’s sibbing

will suggest a solitary action by the agent, such as sneezing, or jumping. However, a transitive form such as:
suggestions an act being done by the agent on another, such as pulling, hugging, or catching. Verbs in particular depend heavily on sentence context for the meaning to be determined because the contexts in which they occur are so rich in potentially misleading referents. Gleitman and others\textsuperscript{10,11} emphasize that the process of learning the meaning of a new word begins with this fast mapping, a kind of informed guess about the word’s meaning, but that typically several contexts are required before a meaning can be fixed.

Point 3: Words Are Organized
What is it to “fix” the meaning of a word? Crucially, it is to place it in a matrix of fine contrasts. If you have guessed that \textit{plum} is a fruit, it is important to distinguish it from pears and apples and nectarines. If you have only a vague meaning attached to the word \textit{carry}, for example, you don’t yet have it set in contrast to slight variants such as \textit{lift} or \textit{hold}. Our lexicon seems to be organized into meaning clusters and contrasts, as well as elaborate networks of other associations, leading to fast and efficient retrieval. Furthermore, there is hierarchical organization to many areas, so we can call a single object by several names: a Golden Delicious, an apple, a fruit, food.\textsuperscript{12,13} The very young child may not have such networks and contrasts so well established, and it had been noted that sometimes young children resist accepting more than one word for an object: “That’s not an animal, it’s a dog!”\textsuperscript{14} However, Waxman and Hatch\textsuperscript{15} demonstrated that children as young as 3 years old do organize their nouns into hierarchical levels. In their task, they showed children a picture and asked the children to name it at different levels of description.

For example:

Examiner shows child picture of a rose:
E: That’s a dandelion.
Expected response: No it’s not, it’s a rose.
E: That’s a tree.
Expected response: No its not, it’s a flower.
E: That’s an animal.
Expected response: No it’s not, it’s a plant.

The children were able to shift their descriptions depending on the level of the prompt. Even the 3-year-old children were sensitive to the level of the prompting in providing a contrast. So with development, the lexicon gets increasingly refined, organized into clusters, contrasts, networks, and hierarchies, making for efficient retrieval. As children reach school age, increasing demands will be placed on this organization: “Think of an animal beginning with /b/.” “What’s the opposite of \textit{deep}?” “How many fruits can you tell me?”

Point 4: Semantics Goes Beyond Word Meaning
The domain of semantics is extraordinarily rich, with the meaning of individual words being just a tiny part of it. However, studies of the lexicon of young children have dominated semantic research in language acquisition until quite recently. In the last 10 or 15 years, an increasing interest has developed in the acquisition of the meaning of quantifiers.\textsuperscript{16–19} At first glance, quantifiers (words such as \textit{some}, \textit{all}, \textit{none}, \textit{every}, \textit{each}) may seem to pose no different a challenge than subtle contrasts in, for example, prepositions (\textit{in}, \textit{on}, \textit{under}, \textit{through}, \textit{between}). However, quantifiers introduce the interesting problem of “scope”; that is, they have influence over other parts of the sentence. For illustration of scope take a word such as \textit{only}:

\begin{itemize}
\item Only the man saw a robber in a mask
\item The man \textit{only} saw a robber in a mask
\item The man saw a robber in \textit{only} a mask
\end{itemize}

\textit{Only} “takes scope” over different parts of the sentence, and creates different meanings. Take for another example a sentence such as:

\begin{itemize}
\item Every boy is riding a horse.
\end{itemize}

The sentence is usually understood as meaning that there are several different boys and for each boy there is a horse that he is riding. But notice it could mean that one poor horse is taking the weight of the boys all together. The sentence is ambiguous, and the ambiguity comes from whether \textit{every} takes scope over \textit{a}, or \textit{a} takes scope over \textit{every}: either,
“For every boy there is a horse to ride,” or “There exists a horse, and every boy is riding it.”

Research has shown that young children can readily understand these readings, but that does not mean they have the whole system mastered. Studies have shown that children entertain alternative readings that adults don’t even consider,\(^ {16,18}\) and that they fail to limit their interpretations when readings become blocked for adults. As just one example, adults find it hard to get both readings for the sentence:

There is a horse that every boy is riding.

It seems to most adults that there must be only one beladen horse, but not to children.\(^ {20}\)

**WHAT CAN GO WRONG**

What about vocabulary learning in children with language disorders? A reduced vocabulary is a very typical sign of a language disorder, with the usual measure again being a standard picture-choice test or a naming task.\(^ {21}\) As a consequence, a language-disordered child will be at a disadvantage in everything from ordinary conversation to school readiness and reading. But why does the problem occur? Recent work has suggested that the problem may stem from several sources.

First, a child with a language disorder may have difficulty learning new words in a casual way, that is, from context. Research suggests that children who are language-delayed might have difficulties learning new words from such casual contexts, and may need more exposures to narrow down a word’s meaning.\(^ {22}\) Furthermore, it is evident that a child who has grammatical difficulty (with word order or inflections, for example) will have difficulty picking up the cues to the meaning of a word that sentence contexts provide.

This observation may shed new light on a phenomenon noted by Rice and Bode,\(^ {23}\) who found that language-delayed children overuse “all-purpose” verbs, called GAP verbs. That is, the children studied used many more verbs of a too-general nature, such as *make, do, put, let,* and *go,* where more specific verbs might be more appropriate, such as *draw, catch, fill,* *help,* or *drive.* This suggests one problem may lie in retrieving words that are specific enough to meet the communicative demands. It seems important to ask whether the difficulty stems entirely from inefficiency in fast mapping meanings in the first place, or whether there is an additional difficulty in the adequacy of the child’s lexical organization once the words have been learned. That is, how flexible is the retrieval of known words? Are they suitably organized into hierarchies and contrasts?

With regard to the learning of quantifiers and their scope, rather little has been done because the area is so new to the field of language acquisition. Yet an important start was made on this question in the work of Finneran.\(^ {24}\) With a small sample of language-delayed children, she studied the phenomena that were beginning to be explored in typically developing children. The surprising fact was that her children with language disorders were still exhibiting some of the same problems at 9 and 10 years of age. These results have since been extended in German by Penner and others.\(^ {25}\) Again, it is important to uncover which of the problems might stem from failure to understand the meaning of quantifiers, and which might stem from a failure to integrate that semantic knowledge with the syntax to get the right scope.

**ITEM GOALS**

In developing the specific probes for a linguistically fair assessment, we tapped only a tiny portion of the phenomena under the umbrella term of “semantics.” In traditional tests, semantics is construed mostly as word meaning, notably vocabulary. However, specific vocabulary is precisely what a dialect-sensitive assessment seeks to avoid. We argued that finding a vocabulary test that is fair to every child regardless of culture and circumstance is an impossible task. Existing vocabulary tests are testaments to that failure.

In the Semantics domain, we consider several aspects of semantic functioning that are neglected in existing tests, but show promise for distinguishing children with language disorders from those who differ only by dialect. In trying to minimize the impact that
having different vocabularies might have on children’s performance, we needed to find alternative strategies for distinguishing children who have language problems from those who do not. There are four properties of semantic functioning that we tap into in this domain—all are very different and unique to our work.

The core principles of the item types that we recommend for semantic assessment are as follows:

1. They avoid the bias of acquired vocabulary tests as too culturally dependent.
2. They look at process: Can the child learn a new word easily from context?
3. They look at lexical organization/retrieval, which may be more significant than size of vocabulary.
4. They explore in greater depth at least one element of complex semantics vocabulary: the logical properties and scope of the word every.

In the following text, each item will be discussed in turn.

Fast Mapping
The first subdomain to be discussed examines the process of learning a new word from context. The fast-mapping items capitalize on the essential aspect of semantic functioning that should be intact in the normal child regardless of circumstance: the ability to guess the likely meaning of a new word after just a couple of exposures. Verbs were the word type chosen, for several reasons mentioned above. First, verb learning is most heavily dependent on linguistic context for deriving a meaning. Second, verb vocabulary may be less tied to variations in culture than is noun learning. Third, language-disordered children seem to have weak verb vocabularies.

The assessment that we developed was based on recent experiments that have shown that children can use syntactic information to learn something about a new verb. Even with one exposure, children are able to gain a rough meaning of a verb. Fisher proposes that children can use something as vague as the number of arguments in a sentence to select between possible meanings of a verb that are carried by the transitive and intransitive frames in which the verb appears. Fisher showed 3- and 5-year-old children a video of an action. For example, one picture was of Person B sitting on a swivel stool, being spun by Person A pulling off a scarf wrapped around the waist of Person B. This action was labeled with either a transitive (a), or intransitive (b):

a. She’s mooping her over there, or
b. She’s mooping over there.

If the label for the action was (a), the action described must be the action of pulling on the scarf, and Person A must be doing the mooping to Person B. If the subject heard (b), the action described is most likely, though not inevitably, the action of spinning on the stool, and this time Person B is mooping. After three presentations, the experimenter brought out a still picture from the video and asked the child to point to the one performing the mooping. When given a transitive sentence (a), all of the children performed almost perfectly, selecting the agent as the subject. Children have an ‘‘agency bias,’’ but when given an intransitive sentence (b), even 3-year-old children could override this bias to some degree, quite often selecting the patient (e.g., the one sitting on the stool) as the one doing the mooping.

Johnson extended the basic methodology of such studies to ask whether other sentence frames would also be used to establish which action to attend to in a complex scene. She used not only transitive and intransitive verbs, but also verbs in dative constructions:

c. The boy temmed the flowers to the girl.

d. The girl temmed the boy to send the flowers.

Notice that the first can only be some kind of “transfer” meaning, like pass, or give, and the second can only be some kind of communication or desire-type meaning, such as asked or wanted. The children saw three scenes in a series, as in Figure 1, in which a girl was clearly signaling to a boy with some flowers across a
gap, then the boy floats the flowers across to her in a basket hanging from a balloon. Obviously (c) draws attention to the balloon carrying or transfer act and (d) to the signaling. The argument structure sets limits on the verb meaning. In Johnson’s study, one group of children would receive one form, whereas the other group received the other form. In that way, it could be determined whether there was any inherent bias in the scenario leading a child to pick one action more often than the other as the referent. The different choices should just be governed by the argument structure.

Unlike Fisher’s study, because the events were complex it was not always possible to simply point to the action being named. In the case of the complement-taking verb, the most likely meaning is a mental state or communication: You cannot easily point to wanted or asked. So, the task is designed to ask children questions about the event, and their answers reveal which action/event type they have associated with the novel verb. For example, the child is asked, “Which one is temming”? or “Which one got temmed”? However, the problem occurs if the child fails to answer the questions correctly because of the morphology or syntax of those questions, not because of the novel verb meaning? The test is introduced and modeled first with real verbs that the children know, and they are asked parallel questions about those verbs. This ensures that they have practice with the task with familiar verbs first, and also that the examiner can tell if the questions cause differential difficulty with the novel or familiar verbs.

Johnson’s subjects were 60 AAE- and MAE-speaking 4- to 6-year-old children, and the task helped to reveal how much children could learn from different structures about verb meanings on limited exposure. Furthermore, the children from both dialects did equivalently well, suggesting this might be an unbiased methodology suitable for a linguistically fair test.

Johnson’s methods and pictures were adapted for the items in the fast-mapping probe. The argument structures used were transitive, dative, and infinitival complement forms, each presented with a different novel verb.

**Figure 1** Fastmapping novel verbs by Dialect.
Because it was not desirable to have two forms of the task, arbitrary choices were made to select one picture to associate with the transitive and another with the dative, for example. Johnson’s study had shown that each action/event was equally available in the scenarios, and the child’s choice was governed by the sentences used. The novel verbs are preceded by items with real verbs so the child learns the task, and so the examiner can tell if the questions cause problems. In addition, to break up response sets and provide some variety, foil questions are interspersed, such as, “Which one is wearing a green dress.” More details of procedure, design, and results are provided in Johnson, de Villiers, Seymour, and Roeper (manuscript in preparation).

For time constraints, there was further selection of the materials to sample all the types of argument structures and questions in the most efficient manner, reducing redundancy while keeping the task understandable to the youngest children. As in all of the subdomains, it was necessary to balance constraints of time with concerns about breadth and diversity of tasks. In addition, it was important that some questions would tax the skills of 9-year-old children, yet not discourage the 4-year-old children because the task was too difficult. These are constraints that test-makers are familiar with but experimenters rarely encounter.

Lexical Organization

To tap lexical organization, we were inspired by the work of Waxman and Hatch. However different the child’s experiences and consequent vocabulary are, the child must have words organized to make efficient use of them. Words are organized in hierarchies of increasing specificity like

animal → dog → poodle
food → vegetable → lettuce

They are also organized as oppositions, such as hot/cold, tall/short, nice/nasty, and as synonyms or related words, such as big/tall/large/huge, and so forth. We knew from existing work that children begin this process of organization at about age 3 years, and their ability to relate words flexibly is a developmental process continuing into the school years. However, much of the previous research work was with noun hierarchies, and nouns are perhaps the most subject to cultural variation. What if a child has not grown a garden, or been to a zoo or a dog show, or had a varied diet?

A pilot study by Pearson, Wagner, Asplin, and de Villiers, asked if similar information could be gathered by using the verb domain instead of nouns. Verbs are a more culturally neutral domain to examine because the majority of common verbs do not require specific cultural experiences. Children raised in poverty, therefore, may have relatively more equal opportunities to acquire a structured vocabulary of verbs than they do of nouns. Furthermore, there is suggestive evidence that children with language impairments may have reduced verb vocabularies, with an increased use of all-purpose verbs. Thus verbs are a promising domain in which to differentiate between language disorder and difference.

Verbs show less hierarchical organization than do nouns but it is still present in some sets: move → walk → stagger

That is, walking is a type of moving; staggering is a specific manner of walking. In the verb contrasts subdomain we ask whether a child can provide an appropriate contrast, at the appropriate “level” in the hierarchy, in naming some actions in flexible ways. We created a verb antonym task that was modeled after the Waxman and Hatch study of noun hierarchies. Linguistic research has suggested that verb hierarchies are relatively “flat.” Nevertheless, an antonym task with verbs does seem to tap lexical subcategories, such as manner of motion, direction of motion, means of creation, etc.

For example, to a picture of a girl licking a popsicle, we might say

“This girl is not chewing the ice cream, she’s...”

an appropriate response is “licking,” but not “eating,” even though that is also true of the picture. However, given the prompt:
“She’s not drinking the ice cream, she’s…”

then “eating” is a better response than “licking.” To a picture of a man hammering a toy airplane together, we would say:

“He’s not gluing the airplane, he’s…”

and expect, for example, “hammering it,” not “fixing it.” But if we then said,

“He’s not breaking the airplane, he’s…”

we expect, for example, “fixing it,” not “hammering it.” To succeed at this task, children need to have some minimum number of verbs in their vocabulary, and, equally importantly, they must have those verbs organized into appropriate subcategories and contrasts.

The task consisted of 10 pictures of a range of common verb classes, such as motion, grooming, and breaking, that have at least partial hierarchies. A single practice item was used to demonstrate the task above, with three different prompts per picture: (a), (b), and (c), differing in what they picked out. Often, but not always, the prompts increased in generality from (a) to (c). If the child failed to produce a new verb to a further prompt, the experimenter said, “You already said that—think of something else,” and repeated the prompt. Only one such repeated prompt was given at each step.

We coded the adequacy of the child’s answer; namely, was it at the right level of description? We also coded the novelty of the answer: Did the child just repeat the same answer or give three different answers? In this work we attempted a refined scale, with 0 being inadequate; 1 being on target, but too specific or general; and 2 being just right.

**SAMPLE TOO-SPECIFIC ANSWERS**

In answer to:

“He’s not breaking the car apart, he’s…”

A child might say, “painting the door.”

“Shes not hating the popsicle, she’s…”

“A child might say, “licking it.”

**SAMPLE TOO-GENERAL ANSWERS**

In answer to:

“He’s not gluing the airplane, he’s…”

A child might say, “holding it,”

“He’s not breaking the car apart, he’s…”

A child might say, “working”.

A score of 2 was given for a good antonym of the verb provided in the prompt, such as licking not chewing, or painting not gluing.

The preliminary work with 30 AAE-speaking children aged 4 to 6 years suggested the feasibility of the task and measures, and that there was development throughout that age range. A very small pilot study with four language-disordered children suggested the task might discriminate well between normal and disordered children.

The pictures and prompts were adapted for the experimental test, with a larger sample of items initially that was reduced for the reasons of time and efficiency. Instead of three prompts per picture, which sometimes strained even our verb lexicons, pictures were chosen that were suitable for two clear prompts. Prompts needed to have distinct answers, because in a production task it is often surprising how much variability there is and how much ambiguity of scoring can then result (especially over this very large sample and age range). The fine gradations were abandoned for consistency and ease of scoring, so the new scoring is simply 1 or 0 for each item depending on its adequacy as a contrast.

The probe was successful enough that we sought to use the same task design to look at another domain of semantic contrasts, namely prepositions. Prepositions are also in a matrix of oppositions, though again without much hierarchy. A child must know the subtle differences between prepositions to arrive at the right description of a scene. Work on prepositions had revealed that although prepositions are used early, preposition contrasts emerge over the course of development. Furthermore, even closely related European languages set up
the contrasts among spatial prepositions in different ways, so the child’s task is not so easy. In addition, not all prepositions refer to space: some are abstract or purely grammatical in function. For example, why do we listen to something, look at something, go on vacation? The challenge for the child is to learn how the language uses prepositions in both concrete and abstract ways.

The preposition task mirrors the verb contrast task in providing prompts for the child, two per picture, that require the child to fix or supply a preposition to describe what is happening. For example, the child is shown a picture of a girl riding on a horse.

The prompts would be:

“She’s not riding to a horse, she’s riding...”

The expected response would be “on a horse”

“She’s not sitting behind the saddle...”

The expected response would be “in the front seat”

The adequacy of the child’s responses is fairly easy to judge. We avoided cases where we had reason to believe AAE might introduce a difference, though we may not have succeeded entirely because so little is known about dialect differences in children’s use of prepositions. Orr reports a use of at in AAE in place of Mainstream American English (MAE) to that we may not have been sufficiently sensitive to in the current scoring; for example, “She’s going at Grandma’s.”

The Meaning and Scope of the Word Every

We argued above that semantics is not just about the lexicon, it also involves the way that quantification works in sentences. Quantifiers are terms like less, more, some, any, all, every, each, none, and the way these terms interact in sentences to give meaning is quite intricate.

The work leading up to the design of the items in these subdomains is quite new and its interpretation is still controversial. For example, Philip and Drozd have both found that children make systematic misinterpretations of a sentence such as

“Is every man on a horse?”

If the scene contains an extra horse, children say “No, not that one,” pointing to the horse. It has been suggested and contested that children treat “every” as if it means “always” (very roughly speaking); that is,

“Is it true in every case that a man is on a horse?”

The tantalizing idea is that every at first is not anchored to the noun phrase “man,” but somehow takes scope over the whole sentence. In addition, previous work had shown that children did not restrict the scope of “every” to elements within its sentence, but sometimes allowed it to cross sentence boundaries. Take for example, the following sentences:

The man watched every cat. He scratched his ear.

For an adult, be cannot possibly refer to every cat. This is not because the pronoun is singular, because if a singular form (e.g., his, or a) occurred within the same sentence as every, it would be fine to give it that multiple or “variable” interpretation:

The man watched every cat scratch his ear.

So the child must learn the meaning of specific quantifiers; the differences between all, some, none, and every. The child must recognize that these modify nouns, not whole events; for example, every is not the same as always. The child must also learn the limits on how every works within a sentence, not across sentences:

The man watched every boy play the drum, is not the same as:

The man watched every boy. He played the drum.

Finneran had tested a variety of different quantifier and wh-question tasks that involve these subtle semantics on a small group of
MAE-speaking language-disordered children, and had evidence for significant delay in their understanding even at age 9 and 10 years. However, we had no guarantee that AAE development might work the same way as MAE, and several reasons to suspect it might not. AAE has properties such as negative concord, which can be construed as a form of quantification that takes a very different form syntactically. We therefore included several different tests of the quantifier every to determine if the tests could be both fair across dialects and discriminating of difference.

The quantifier items are restricted to evaluating children’s understanding of the word every and how it works in sentences. In the simplest of cases, the child is asked about a picture in which all but one man, for example, would be driving a car:

**Is every man driving a car?**

The answer is “no” if the child understands that every does not mean, for example, some or any.

More interesting is a picture in which several monkeys are eating bananas, and a rabbit is eating a carrot. The question is:

**Is every monkey eating a banana?**

The answer is “yes,” if the child understands that every is only attached to monkey, and not to other cases like the rabbit. If the child thinks that the sentence means that every subevent had to be one of a monkey eating a banana, the child would answer “no.”

Finally, the child sees a pair of contrasting pictures that give two alternative readings for sentence sets such as:

**The man watched every boy. He played the drum.**

The child must choose the picture in which the man, not every boy, is playing the drum. If the child chooses the correct picture, he or she understands that every cannot cross sentence boundaries for be.

Children can fail in several ways on the quantifier tasks. For example:

1. They can miss the basic meaning of every and answer “yes” instead of “no” to the car picture.
2. They can take every to apply also to the rabbit in the monkey picture, not limiting it to the noun phrase monkey, and thus say “no” instead of “yes.”
3. They can fail to discern the condition on the scope of every imposed by a sentence boundary, and think he refers to every boy in the case of “The man watched every boy. He played the drum.”

**HOW CHILDREN PERFORM ON THE SEMANTICS SUBDOMAINS**

How Children Perform on Fast Mapping

The fast-mapping task produces intriguing data at the level of the individual child, question, and construction type that are too elaborate to describe here; the reader is referred to Johnson et al. (manuscript in preparation) for the full details. Instead, what follows is a summary of the composite data across the different age, dialect, and disorder-status groups. The child is given a point for a correct answer judging by the adult grammar, and the points are tallied across constructions and questions.

Figure 2 shows that the data from this study replicate Johnson’s initial work. Even with this much larger sample, and larger age range (4 to 9 instead of 4 to 6 years), there is no measurable difference in the performance of AAE- and MAE-speaking children. [Novel Verbs: Age, \( F(5, 1002) = 15.141, p < 0.0001; \) Dialect, \( F(1, 1002) = 0.415, p = 0.520; \) Age by Dialect, \( F(5, 1002) = 1.150, p = 0.332. \)]

Furthermore, there is a very reliable age trend that continues throughout the age range, making this a suitable item-type for examining development. However, Johnson’s methods and procedures were selected primarily because they suggested (and now we confirm) the lack of bias in the test. If the test shows no reliable difference between normally developing and disordered subjects, it will fail the other requirement of our test.

Fortunately, Figure 3 shows that the test is much more challenging for language-disordered children at every age. [Novel Verbs:
In our fuller account of these data, we discuss where the differences are located (Johnson et al, manuscript in preparation).

How Children Perform on Verb Contrasts

The verb contrast task lived up to the expectations generated by the pilot study, over the larger age range. Normally developing children who spoke AAE or MAE proved increasingly adept with age at providing an appropriate level contrast. In addition, there was no significant difference in the performance of the two dialect groups. (See Figure 4) Age, $F(5, 1002) = 57.116, p < 0.0001$; Dialect, $F(1,1002) = 6.885, p = 0.009$; Age by Dialect, $F(5, 1002) = 0.838, p = 0.523.$) Note that the main effect of Dialect is almost entirely due to the 4-year-old children, and is nonsignificant once they are removed [$F(1, 796) = 3.576, p = 0.06.$] Mistakes, especially for the younger children, consisted of saying things such as, “I don’t know,” but more commonly giving an answer that was too specific or too general to be the right level for the prompt. This could suggest that young children’s verb vocabularies either are still incomplete or weakly organized.

However, it is very clear from Figure 5 that the problems that young children have are amplified in language-disordered children, who differ from their normal age-mates at all ages. In analyses of variance: Age, $F(5, 1002) = 65.505, p < 0.0001$; Clinical Status, $F(1, 1002) = 85.616, p < 0.0001$; Age by Clinical
The disordered children use more vague expressions, such as all-purpose verbs (do, try, and put) or they omit the verb. They have difficulty with the right contrasts after the prompt, and have difficulty saying two different things about the same picture.

Thus, the subdomain discriminates well between children with language disorder and children who are normally developing, and confirms previous reports that the verb vocabulary may be vulnerable in language-disordered children.

**How Children Perform on Preposition Contrasts**

The preposition contrast task did not have the same level of pilot work behind it to back up the hints in the literature that it might prove revealing. Nonetheless, the data from the field testing (Fig. 6) showed good age trends over ages 4 to 9 years, and equivalent levels of performance from MAE- and AAE-speaking children [Age, $F(5, 1002) = 18.379, p < 0.0001$; Dialect, $F(1, 1002) = 1.077, p = 0.30$; Age by Dialect, $F(5, 1002) = 0.682, p = 0.637$]. Mistakes consisted of confusing prepositions or sometimes omitting them altogether. These mistakes, especially omissions, were much more prevalent in the disordered children, who can be seen clearly to lag behind their normally developing age-mates in Figure 7. [Age, $F(5, 1002) = 20.187, p < 0.0001$; Clinical Status, $F(1, 1002) = 35.288, p < 0.0001$; Age by Clinical Status, $F(5, 1002) = 0.914, p = 0.471$.] The disordered children sometimes use odd forms, such as, “down the chair” instead...
of “under the chair,” or “up the cat” instead of “up to the cat.” We have not detected any particular difference between spatial and abstract prepositions, but the sample of the latter is tiny.

**How Children Perform on the Quantifier Subdomain**

First, consider the items about the meaning of every within a sentence. The results of the field testing research reveal that the findings of earlier studies are borne out. Children do make the kinds of mistakes we have highlighted above, and there is a clear developmental time course persisting until age 7 or 8 years even in normally developing children (Fig. 8). We still need to explore these data in detail, but for the moment, these items appear to fulfill our requirement not to create bias against AAE speakers. In the analyses of variance, the values for these items followed the general pattern of the other item-types: Age, \( F(5, 1002) = 33.051, p < 0.0001 \); Dialect, \( F(1, 1002) = 2.225, p = 0.136 \); Age by Dialect, \( F(5, 1002) = 0.877, p = 0.496 \).

In addition, the subdomain is highly successful in separating out disordered children from either dialect group. It is clear that the language-disordered children are considerably delayed in mastering this quantifier (Fig. 9) (Age, \( F(5, 1002) = 31.766, p < 0.0001 \); Clinical Status, \( F(1, 1002) = 27.675, p < 0.0001 \); Age by Clinical Status, \( F(5, 1002) = 0.351, p = 0.882 \)).

With regard to the sentence boundary items, Figure 10 reveals a good match in the developmental time course for AAE and MAE children (Age, \( F(5, 1002) = 44.981, p < 0.0001 \);
Dialect, $F(1, 1002) = 0.051, p = 0.821$; Age by Dialect, $F(5, 1002) = 0.946, p = 0.450$.

Figure 11 shows the expected lag for language-delayed children, at least after age 4 years [Age, $F(5, 1002) = 42.180, p < 0.0001$; Clinical Status, $F(1, 1002) = 34.711, p < 0.0001$; Age by Clinical Status, $F(5, 1002) = 3.408, p = 0.005$].

**HOW DO THESE FACTS HELP US IN FULL DIAGNOSIS AND REMEDIATION?**

To an experimenter, test making is a peculiar process. Experiments demand many examples to test the robustness of an effect; tests require the minimum that will provide group separation, for efficiency and time. Experiments demand keeping separate those types that are significant theoretically; tests lump items together to give general descriptions. Ironically, however, once the testing is done and the child has a “diagnosis” of a language disorder, the same questions should occur to the therapist as occur to the experimenter. Just why did the child fail this item? What does this pattern of responses mean? How do I intervene and fix the problem if I only know there is a general one?

Thus, although a test requires giving points for different sections and adding them up, we think it is important to examine the details of what the child did on each item-type to determine the source of the problem.

**Fast Mapping**

Consider fast mapping. When the child fails, it is important to know several things. Did the
child fail on the real verbs as well as the novel verbs? If so, what kinds of questions were difficult? A substantial failure on the real verb questions is more suggestive of a problem with syntax or morphology of those questions than it is of a failure to learn a novel verb. Suppose the child passes most of the real verb questions, but fails on the novel verbs. Does the child tend to fail on the more complex argument structures, such as datives and complements? Again, the suggestion is that the child may have trouble learning new verbs if the sentence structure in which they appear is too complex. How is the child’s performance on the syntax item-types? Perhaps the child has trouble across the board with novel verb learning. Is the presence of a novel verb affecting his or her memory? How does he or she perform on nonword repetition in the phonology domain of the Screener, Part 2?

**Verb Contrasts**

Next consider verb contrasts. How does the child’s performance on verb contrasts compare with his or her performance on fast mapping? Is the child showing evidence of being a slow learner from context? Is the failure a genuine problem with the size of his or her verb vocabulary? What kinds of errors occur? If the verbs used are too general, perhaps the child has a limited verb vocabulary. If they are too specific, then the child may have a disorganized lexicon. If the child “misses the mark” of the pictures, does this show up in the preposition task or on the pragmatics tests, where that skill
of getting the meaning from a picture is at a premium?

**Preposition Contrasts**
In the preposition task if the child omits prepositions, how is his or her morphosyntax on the screener? How is his or her memory for phonological material in phonology nonword repetition? Is there evidence from the article task (syntax) that the child has a general difficulty with functors, that is, small grammatical words? If the child provides odd prepositions, does this parallel a problem in verb contrasts?

**Quantifiers**
In the quantifier subdomain, does the child tend to say “yes” to everything? Does the child show a basic understanding of the meaning of every on the car picture? If not, perhaps the basic meaning is still absent, and work needs to be done to contrast all, some, every, and one. Does the child insist that the rabbit is relevant in the monkey/banana picture? If so, perhaps the meaning of every is not yet linked to its noun phrase. How do such children perform on the double **wh**-questions in **wh**-syntax and **wh**-production? Might the child have a problem with the notion of scope or variables? In the sentence boundary items, does the child show a position bias in choosing pictures? That would be the most primitive error. If not, does the child get the single sentences right? If so, does the child take every to cross the sentence boundary? How does the child’s problem relate to his or her understanding of other complex syntax; for example, passives or **wh**-questions?

**CONCLUSION**
This article has demonstrated some ways it might be possible to develop relatively unbiased methods for assessment of the process of vocabulary learning and the efficient organization of vocabulary as alternatives to the measurement of the size of acquired vocabulary. It has also shown that semantic knowledge extends beyond word meaning conceived simply in terms of content words, into areas such as quantifier scope. This latter knowledge may be especially sensitive to language disorders, and provide a window into larger aspects of the grammar as a whole. There is still some distance to go in translating the discoveries into careful diagnostics that will allow efficient and precise remediation, but the prospect is an exciting one.

**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 Roeppe 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 **DELV** is the name trademarked by The Psychological Corporation of Harcourt Assessments, 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 **DELV-ST**, or “screener,” or the **DELV-CR**, **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, Roeppe, de Villiers, de Villiers, Pearson, and Ciolli). Questions about the tests themselves should be addressed to the Project Leader at The Psychological Corporation of Harcourt Assessment, Inc.; Lois Ciolli, Senior Research Director.

**REFERENCES**


16. Philip W. Event Quantification in the Acquisition of Universal Quantification [dissertation]. University of Massachusetts, Amherst, MA; 1995


18. Drozd K, van Loosbroek E. Weak quantification, plausible dissent, and the development of children’s