

# The Distributed Characteristic in Bilingual Learning

**D. Kimbrough Oller**

The University of Memphis

## 1. Vocabulary learning in bilingual children

For some time it has been known that children learning two languages early in life tend to show vocabulary deficiencies in each language if they are compared with monolingual children (Ben-Zeev, 1977, 1984; Rosenblum & Pinker, 1983). It is possible to be misled by this apparent deficiency into thinking that bilingual children have some fundamental handicap. But in fact the pattern appears to be merely a by-product of a property of bilingual learning that we have termed the 'distributed characteristic' (Oller & Pearson, 2002). Some of a bilingual's vocabulary tends to be available in one language without translation equivalent vocabulary being available in the other language. Thus it can be said that lexicalized concepts of the bilingual are 'distributed' across the two languages such that some concepts are lexicalized in one language but not the other and vice versa.

The distributed characteristic for vocabulary may be particularly strong for very young children learning two languages, and this apparent tendency, for which evidence will be reviewed below, is at the root of a fundamental problem for assessment of bilingual children's oral language capabilities. It proves to be utterly invalid to gauge bilinguals' vocabulary knowledge directly in terms of monolingual norms, precisely because the distributed characteristic makes some of the bilinguals' lexicalized world inaccessible to any single-language assessment (Abudarham, 1997).

A satisfying correction for the problem of oral language assessment in bilinguals will be difficult to implement because bilinguals come in so many different degrees of balance between the two languages. With so many degrees of balance, it appears we may need norming for groups to represent every degree. Methods have been suggested for coping with the problem, and these will be considered below.

## 2. Evidence for the distributed characteristic in bilingual children

During the 1990's a group of researchers at the University of Miami conducted a series of investigations providing substantial empirical evidence of the apparent vocabulary deficiencies in bilingual children, along with evidence that pointed to the distributed characteristic as the potential source. The first of the studies (Fernández, Pearson, Umbel, Oller, & Molinet-Molina, 1992) evaluated receptive vocabulary in four-year-olds who had been exposed to Spanish only or to Spanish and English in roughly equal amounts. The tests were based on instruments normed in both languages (Dunn & Dunn, 1981; Dunn, Padilla, Lugo, & Dunn, 1986). For children exposed to both Spanish only and Spanish and English, the Spanish vocabulary scores were lower by about a third of a standard deviation than would have been expected for monolingual Spanish speakers based on norms for the test, and in English the scores were lower by much greater amounts, two to three standard deviations.

An additional study evaluated the same tests with first graders (Umbel, Pearson, Fernández, & Oller, 1992), and again reported low scores with respect to the norming populations in both languages, especially in English. The children in these studies had been more strongly exposed to Spanish than to English to that point in life, even in the groups with Spanish and English at home, a fact which apparently accounts for the higher scores in Spanish. One suggestion about the source of the apparent deficiencies in vocabulary was possible to rule out empirically. Namely it was suggested that the children might have been from families of low socio-economic status, but in the Miami population studied, the expectation was not verified. In fact, based on empirical studies of the children's families, they were clearly of higher SES than the norming populations for the tests in both Spanish and

English. Consequently it was clear that the apparent vocabulary deficiencies were attributable to something else.

The critical test in the Umbel et al. study was the one that proved for the first time quantitatively that young bilingual children have 'singlet' vocabulary (that is, vocabulary for which translation equivalents are not present in the child's vocabulary in the other language). It is important to emphasize that the singlet vocabulary existed, as proven by the study, to a statistically significant degree in both languages. Consequently the study showed that the comparison of the bilingual children's vocabulary knowledge in either language in terms of the monolingual norms was invalid. In English the comparison failed to consider words the children had lexicalized in Spanish. In Spanish the comparison failed to consider words the children had lexicalized in English. The distributed characteristic had been proven to present a problem for assessment in both languages.

Additional groups of children, also subcategorized as having Spanish-only at home or English and Spanish at home, were tested in the 'Miami Project', a major effort funded by the National Institute of Child Health and Human Development in the mid 1990's, and reported on in a recent volume (Oller & Eilers, 2002). In this work, 704 bilingual children were tested. All the children had been born in the USA. The children were tested at Kindergarten, 2<sup>nd</sup> and 5<sup>th</sup> grade. At all grade levels deficiencies of receptive vocabulary in both languages were seen if comparisons were made in terms of monolingual norms (Cobo-Lewis, Pearson, Eilers, & Umbel, 2002a, 2002b). The same was true if comparisons were made with the 248 monolingual children who were matched with the bilinguals and tested similarly in the Miami Project. Further, the apparent deficiencies for English vocabulary applied even to children who were in the high SES category, and even to the bilingual children with Spanish and English at home. For Spanish vocabulary, the pattern was generally similar, with scores showing a general depression with regard to the monolingual norms.

If one considers scores on other vocabulary tests, the relative performance of bilingual children with respect to monolingual norms was not better. If we consider, for example, the picture naming (productive vocabulary) results from the Miami Project, the scores were actually more depressed with respect to monolingual norms than in the case of receptive vocabulary, and again the pattern applied to both languages and to all the subgroups of the design.

### **3. Magnitude of the apparent vocabulary deficiencies**

Results from the Miami Project indicated that differences between monolinguals and bilinguals diminished after a few years of schooling in English, but even in 5<sup>th</sup> grade the differences were not trivial from the standpoint of educational assessment and potential assignment of children to special education categories. The magnitude of the differences depended upon the grade level, the language background of the child, the type of schooling they received (two-way or English immersion), and the language being tested. High SES children did better than low SES children in both the monolingual and bilingual groups, so in attempting to assess the magnitude of differences it is important to keep that in mind.

All the bilingual groups showed the apparent deficiencies that appear to be attributable to the distributed characteristic. Even the bilingual subgroups showing the highest scores on vocabulary were notably lower than would have been expected for monolinguals. For example, even the bilingual children at 5<sup>th</sup> grade with Spanish and English at home and who were in the high SES subcategory had scores on the receptive vocabulary test in English that were four points below the norm expectation for monolingual children of mid SES (the children represented most strongly in the norming sample). This score is considerably below the one that would be expected for monolingual children of high SES. For monolingual English speakers in the Miami Project, high SES produced a twelve-point advantage on the receptive vocabulary test over low SES. Thus it is clear that the 5<sup>th</sup> grade bilinguals of high SES with Spanish and English at home were trailing their monolingual peers by considerably more than four points.

If we consider naming (productive) vocabulary, the apparent deficiencies were even greater. Bilingual children in many instances trailed the norms for monolinguals by more than a full standard deviation. Further, if we evaluate the scores specifically for children at Kindergarten and 2<sup>nd</sup> grade, the differences favoring the monolinguals were yet larger. In many instances, apparent deficiencies in

vocabulary scores of bilingual children in the Miami Project represented differences of more than two standard deviations with respect to monolingual norms – for example, at Kindergarten for children of low SES in both English and Spanish for the naming vocabulary test.

These are most certainly not trivial differences. They can produce false impressions about children's capabilities, and they can form the basis for misassignment of children in school. The fact that such tests are often treated as intelligence tests is disturbing in the case of bilingual children where it is clear any correlation between the vocabulary test and a test of general intelligence is fundamentally confounded with factors associated with the language histories of children.

#### **4. Possible corrections for the distributed characteristic**

One might imagine that the assessment problem implied by the apparent vocabulary deficiencies of bilinguals could be corrected by merely testing children in both languages. But since the problem applies to both languages, the result of merely testing in both offers nothing more than two views of the apparent deficiency. Bilingual children tend to test low in both languages. If we are correct, that the apparent deficiency is wholly or primarily the result of the distributed characteristic, then there should be means of correcting for it, but developing and implementing such means will require some thought, and will require supporting empirical work in order to be taken seriously.

A series of studies by Pearson and colleagues (Pearson, 1998; Pearson & Fernández, 1994; Pearson, Fernández, & Oller, 1993) have suggested that indeed the apparent deficiencies in vocabulary among young bilinguals may be largely or wholly attributable to the distributed characteristic. The work focused on early vocabulary measured primarily by the MCDI (Fenson et al., 1991) in both English and Spanish. Pearson defined the notion 'conceptual vocabulary' to refer to a combined vocabulary across languages. If an individual knows a word and a translation equivalent, the two are counted only once, as a doublet. Each singlet is also counted, so that the final conceptual vocabulary count reflects the number of concepts lexicalized in one or both languages. Using this notion, Pearson found that conceptual vocabulary size for monolingual and bilingual children was similar, even though vocabulary size in a single language favored the monolinguals strongly. Whether conceptual vocabulary was greater in monolinguals or bilinguals was not determinable based on the sample sizes involved, and the methods of determining conceptual vocabulary were difficult to implement in any case. A key demonstration of the work was that indeed the distributed characteristic accounted for substantial differences between monolingual and bilingual children's vocabularies in a single language, but still, even to assess the conceptual vocabulary of a single child is a major undertaking, requiring careful determination of what each word the child uses means to the child – no simple matter. The work is intriguing, but Pearson's method would be unwieldy for educational assessments.

In response to the need for a more practical method, and to the concerns about the invalidity of comparing bilingual vocabulary scores against monolingual normed values, a new method for testing English oral language has been proffered by which vocabulary items for Spanish-English bilinguals are tested first in English and then missed items are tested in Spanish (Muñoz-Sandoval, Cummins, Alvarado, & Ruef, 1998). If any retested item is known in Spanish, the score is increased accordingly. After adding points for words known in Spanish, children's scores can actually end up being higher than in monolinguals (Alvarado, 2000).

Unfortunately, this adjustment based on retesting missed items depends upon assumptions that are not themselves tested, and are not thus proven to be psychometrically valid. The problems of interpretation of the adjusted scores are various, but among them are problems that resemble those encountered by Pearson and colleagues, namely that the notion of translation equivalence is itself very slippery and full of potential hazards for score adjustment. Words in different languages do not always cover the same semantic domains. For example, in Spanish, the word for 'chin' is 'barba', but the same word in Spanish is also used to refer to a beard. The word for 'pitcher' is commonly 'jarro' but other options exist, and 'jarro' or 'jarra' can also mean a 'jug' or a 'pot'. Further 'pitcher' can also be translated as 'cántaro' or 'bocal'. There is no easy way to account for all these variations and mismatches of semantic domain in constructing a retest in Spanish for words missed in English. Perhaps even more importantly, there are not only dialect variations in uses of individual words to contend with (some of which are represented in the options indicated in the examples), but also

idiosyncratic interpretations of words that children may have learned from their idiosyncratic experience, or that they may have concocted. These facts complicate decisions about what an acceptable correct answer is in the second language tested. The psychometric problems with making an adjustment based on the method suggested by Muñoz-Sandoval et al. are bound up in these sorts of difficulties of interpretation and in the fact that the effects on norm comparisons when one adds a component to the test (retesting any sort of missed items whether in a different language or in the same one) cannot be gauged.

The only secure way to make an adjustment for bilingual children is to develop norms specifically for the bilinguals. Such work is underway in a project at Temple University under the direction of Aquiles Iglesias, who is developing new tests. At the same time there already exist extensive bilingual data on tests that are widely used. Raw scores that have been obtained from large samples of bilingual children (such as those in the Miami Project where over 200 bilingual children were tested at each of three grade levels) could be used as the basis for calculating new sets of bilingual-specific norms. Of course, the most reliable comparisons would take account of the published values for subgroups of SES and language spoken at home. The Miami data were taken on children born in the USA only, but additional data are being collected on the Woodcock-Johnson and PPVT/TVIP measures in other large-scale projects currently underway (see e.g., information about the the Delss network of projects at [www.cal.org/delss/](http://www.cal.org/delss/)) for children born in a variety of other countries.

In any attempt to utilize information from these studies to create new norms, it will be important to consider relative knowledge of the two languages. Sets of norms may need to be developed for several subgroups based on relative dominance in English or Spanish.

## **5. Circumstances that foster distributed vocabulary learning**

It is clear that it is indefensible to continue testing bilingual children and comparing their results to monolingual norms without substantial reinterpretation. In the years to come we are going to be required to account for the distributed characteristic of bilingual knowledge both in testing and in the fundamental attempt to understand language learning. Speculations about the reasons that the distributed characteristic occurs have focused upon apparent differences between the semantic contents of communications that are habitually available to children in one language or the other. For example, Li (1996) argues that there is domain specificity in language exposure for the two languages of many bilinguals, and that it is this domain specificity that primarily accounts for the apparent vocabulary deficiencies of bilinguals. In Miami, we speculated that bilingual children tended to know words in Spanish when they pertained to the home (kitchen implements, sewing items, bedroom furniture, etc.), and that they tended to know words that pertained to school in English (blackboards, writing implements, desks, etc.) precisely because their home language was Spanish and their school language was English.

If this speculation is correct, one might imagine that the apparent vocabulary deficiencies attributable to the distributed characteristic would tend to decline with time, as the learner experiences increased exposure to a variety of circumstances in both languages. Indeed, the only published investigation that I know of that offers a perspective on long-term effects of bilingual experience (Bahrick, Hall, Goggin, Bahrick, & Berger, 1994) suggests that eventually (after decades of exposure to both languages) vocabulary knowledge of the bilingual may outstrip that of monolinguals in both languages. The results suggest that advantages of the broad perspective afforded by bilingualism and the ability to reason about cognates (for language pairs that have cognates) may eventually outweigh time on task and domain specificity factors in vocabulary learning.

## **6. The distributed characteristic in other domains of language**

Vocabulary learning is a special kind of task in language learning, one that depends on item-by-item learning more than in the case of any other task. To illustrate the contrast between vocabulary and other aspects of language, consider phonology and syntax. In phonology, we learn a finite set of phonemic elements and contrasts, and the finite system by which these can be strung together in phonotactically well-formed sequences. Once the phonological system is in place, it is relatively fixed,

and it appears to be fixed in normal learners by early in life, perhaps because the number of basic elements (phonemes, features and rules) to learn is small. In syntax, similarly we learn a finite system by which sentences can be formed from strings of words. Once learned, syntax is also largely fixed, and changes only very gradually as dialects drift over the years. Children learn the basic syntax of their language fairly early in life, again, in part perhaps because the number of basic elements (parts of speech, constraints, and rules) to learn is small.

But lexical acquisition is completely open-ended. It happens every day, and continues throughout life. The lexicon's size must be counted in many thousands in any normal mature speaker, and the number continually expands. Lexical items have to be learned one by one, and in most cases there is little advantage in knowing one word (or perhaps one root word), when it comes to learning the next. Thus the individual who lives in two differentiable domains, and who speaks different languages in those two domains, may tend to learn words in one that simply never or perhaps rarely come up in the other. With syntax and phonology item-by-item learning appears to be much more limited because syntax and phonology involve elements that are more tightly bound together in a system where the elements are interdependent and tend to be learned in interwoven, generalizable groupings.

The item-by-item learning of (root) lexical items appears to create the possibility for domain specific learning and consequently, for the distributed characteristic. It is important however not to leave the impression that the distributed characteristic cannot have effects in other domains. Because lexical items incorporate syntactic information (subcategorization constraints, gender, etc.), it is logically necessary that an individual with weakness in vocabulary will also show a relative weakness in syntax with respect to those aspects of syntax that are bound to the lexicon. Consequently, bilingual children with vocabulary deficiencies with respect to monolinguals, can be expected to show deficiencies in syntax as well.

## **7. Profiles of bilingual speakers**

The strong tendency for the distributed characteristic to affect vocabulary scores, and the less prominent role that the distributed characteristic may play in other domains may help explain the very strong 'profile' effects that have been observed in bilingual children. Bilinguals tend to appear to be weak compared to monolinguals in some areas, but not in others. The profile results reported in the volume from the Miami Project (Cobo-Lewis et al., 2002a, 2002b), for example, showed relatively strong performance in both languages by bilingual children in phonics, very weak performance in picture naming vocabulary, and intermediate performance in reading comprehension. Such an outcome may be the result of differing degrees of effect of the distributed characteristic. In phonics, learning may occur predominantly in school for both monolingual and bilingual children and the task is of delimited size, so that the distributed characteristic may play little if any role. Vocabulary shows a very strong effect of the distributed characteristic, and reading comprehension, which may reflect a combination of factors including those that are involved in phonics and those involved in vocabulary, provides a mixed, and thus intermediate outcome.

These are current speculations about relative performance on tests normed on monolinguals and used for evaluation of bilinguals. However, it is clear from the consistency of vocabulary effects related to the distributed characteristic that the time has come for a new view of bilingual learning that does not treat bilingualism as a secondary phenomenon, always to be compared with a monolingual standard. The profiles of bilingual children with respect to monolinguals are strong and consistent precisely because bilinguals represent a different case of learning and knowledge. There is every reason to consider the bilingual case in its own right, as a natural language learning pattern. Indeed, it is ironic that bilingualism, with all its advantages in terms of flexibility of communication across communities and the ability to engage in multilingual commerce, would be treated as secondary in any science of language or practice of education. There has been considerable alarm expressed over the years in the USA and Britain in particular about the tendency of bilingual children to struggle in school in part because of their difficulty in catching up with monolingual peers in English vocabulary. We can only wonder if there will come a time when educators in these countries will take seriously the vocabulary limitations of monolinguals, about whom it would be just as reasonable to be alarmed that they have learned only about half as many lexical items as their bilingual peers.

## Acknowledgement

This work was supported by an endowment from the Plough Foundation. I also gratefully acknowledge the critical support for the development of this line of research through a grant from the National Institutes of Child Health and Human Development (5R01 HD30762, D. K. Oller, PI, R. E. Eilers, Co-PI).

## References

- Abudarham, S. (1997). The tenability of comparing the receptive lexical proficiency of dual language children with standardised monoglot norms. *Educational Studies*, 23(1), 127-143.
- Alvarado, C. G. (2000). A theoretical and empirical study of the English/Spanish bilingual verbal ability tests. (assessment, second-grade, third-grade). *Dissertation Abstracts International: Section A: Humanities and Social Sciences*, 60(7-A), 2336.
- Bahrick, H. P., Hall, L. K., Goggin, J. P., Bahrick, L. E., & Berger, S. A. (1994). Fifty years of language maintenance and language dominance in bilingual Hispanic immigrants. *Journal of Experimental Psychology: General*, 123(3), 264-283.
- Ben-Zeev, S. (1977). Mechanisms by which childhood bilingualism affects understanding of language and cognitive structures. In P. A. Hornby (Ed.), *Bilingualism* (pp. 29-55). New York: Academic Press, Inc.
- Ben-Zeev, S. (1984). Bilingualism and Cognitive Development. In N. Miller (Ed.), *Bilingualism and Language Disability* (pp. 55-80). San Diego: College Hill Press.
- Cobo-Lewis, A., Pearson, B. Z., Eilers, R. E., & Umbel, V. C. (2002a). Effects of bilingualism and bilingual education on oral and written English skills: A multifactor study of standardized test outcomes. In D. K. Oller & R. E. Eilers (Eds.), *Language and Literacy in Bilingual Children*. Clevedon, UK: Multilingual Matters.
- Cobo-Lewis, A., Pearson, B. Z., Eilers, R. E., & Umbel, V. C. (2002b). Effects of bilingualism and bilingual education on oral and written Spanish skills: A multifactor study of standardized test outcomes. In D. K. Oller & R. E. Eilers (Eds.), *Language and Literacy in Bilingual Children*. Clevedon, UK: Multilingual Matters.
- Dunn, L., & Dunn, L. (1981). *Peabody Picture Vocabulary Test-Revised*. Circle Pines, MN: American Guidance Service.
- Dunn, L., Padilla, E., Lugo, D., & Dunn, L. (1986). *Test de Vocabulario en Imágenes Peabody -- Adaptación Hispanoamericana [Peabody Picture Vocabulary Test -- Latin American adaptation]*. Circle Pines, MN: American Guidance Service.
- Fenson, L., Dale, P., Reznick, S., Thal, D., Bates, E., Hartung, J., Pethcik, S., & Reilly, J. (1991). *The MacArthur Communicative Development Inventories*. San Diego, CA: San Diego State University.
- Fernández, M. C., Pearson, B. Z., Umbel, V. M., Oller, D. K., & Molinet-Molina, M. (1992). Bilingual receptive vocabulary in Hispanic preschool children. *Hispanic Journal of Behavioral Sciences*, 14, 268-276.
- Li, D. C. S. (1996). *Issues in bilingualism and biculturalism: a Hong Kong case study* (Vol. 21). New York: Peter Lang.
- Muñoz-Sandoval, A. F., Cummins, J., Alvarado, C. G., & Ruef, M. L. (1998). *The Bilingual Verbal Ability Tests (BVAT)*. Itasca, IL: Riverside Publishing, Houghton-Mifflin.
- Oller, D. K., & Eilers, R. E. (2002). *Language and Literacy in Bilingual Children*. Clevedon, UK: Multilingual Matters.
- Oller, D. K., & Pearson, B. Z. (2002). Assessing the effects of bilingualism: A background. In D. K. Oller & R. E. Eilers (Eds.), *Language and Literacy in Bilingual Children* (pp. 3-21). Clevedon, UK: Multilingual Matters.
- Pearson, B. Z. (1998). Assessing lexical development in bilingual babies and toddlers. *The International Journal of Bilingualism*, 2, 347-372.
- Pearson, B. Z., & Fernández, S. (1994). Patterns of interaction in the lexical growth in two languages of bilingual infants and toddlers. *Language Learning*, 44, 617-653.
- Pearson, B. Z., Fernández, S., & Oller, D. K. (1993). Lexical development in simultaneous bilingual infants: Comparison to monolinguals. *Language Learning*, 43, 93-120.
- Rosenblum, T., & Pinker, S. A. (1983). Word Magic Revisited: Monolingual and Bilingual Children's Understanding of the Word-Object Relationship. *Child Development*, 53, 773-780.
- Umbel, V. M., Pearson, B. Z., Fernández, M. C., & Oller, D. K. (1992). Measuring bilingual children's receptive vocabularies. *Child Development*, 63, 1012-1020.

# **ISB4: Proceedings of the 4th International Symposium on Bilingualism**

edited by James Cohen, Kara T. McAlister,  
Kellie Rolstad, and Jeff MacSwan

**Cascadilla Press   Somerville, MA   2005**

## **Copyright information**

ISB4: Proceedings of the 4th International Symposium on Bilingualism  
© 2005 Cascadilla Press, Somerville, MA. All rights reserved

ISBN 978-1-57473-210-8 CD-ROM  
ISBN 978-1-57473-107-1 library binding (5-volume set)

A copyright notice for each paper is located at the bottom of the first page of the paper.  
Reprints for course packs can be authorized by Cascadilla Press.

## **Ordering information**

To order a copy of the proceedings, contact:

Cascadilla Press  
P.O. Box 440355  
Somerville, MA 02144, USA

phone: 1-617-776-2370  
fax: 1-617-776-2271  
sales@cascadilla.com  
www.cascadilla.com

## **Web access and citation information**

This paper is available from [www.cascadilla.com/isb4.html](http://www.cascadilla.com/isb4.html) and is identical  
to the version published by Cascadilla Press on CD-ROM and in library binding.