

Lexical development norms for young children

PHILIP S. DALE

University of Washington, Seattle, Washington

and

LARRY FENSON

San Diego State University, San Diego, California

Developmental norms for young children's vocabularies have a number of applications in research design, assessment, and intervention, but have previously been very difficult to obtain. In the present study, month-by-month norms for comprehension and production of 396 words from 8 to 16 months, and production of 680 words from 16 to 30 months, were derived from a norming study of 1,789 children between the ages of 8 and 30 months using the MacArthur Communicative Development Inventories (Fenson et al., 1993). The norms are available in the form of a database program, LEX, for MS-DOS-based computers.

Normative information on the acquisition of individual English words has a number of important uses for both research and application. For example, norms permit investigators to construct stimulus sentences for assessment of syntactic comprehension with matched and appropriate lexical items, to test the role of hypothesized factors such as semantic complexity or prosodic pattern on lexical acquisition, and to develop sequenced curricula for language training programs on the basis of normal sequences of acquisition.

Normative information is difficult to derive from observational studies, however, because most words have such low frequency that they are unlikely to occur in a sample of 100–200 words. Parent report is much less subject to frequency biasing (Dale, Bates, Reznick, & Morisset, 1989). *The MacArthur Communicative Development Inventories* (CDIs; Fenson et al., 1993), which are based on parent report, provide a perspective on child language development that is highly complementary to other sources of evidence, such as language samples and standardized tests. In particular, the inventories have a very high "resolution" in the sense that they provide information about an extensive and representative list of words. Because the information is provided by parents, it is possible to obtain information on far larger samples of children than would be possible with language samples or standardized tests.

Fenson et al. (1993) summarized evidence concerning the validity of the CDI: Infants and the CDI: Toddlers.

Several studies (e.g., Dale, 1991) have directly correlated CDI data with laboratory and home assessments, with impressive results. Indeed, Dale (1991) argued that limitations of the criterion measures—a primary motivation for developing the inventories—may lead to an underestimation of the true validity of the CDIs. He noted that the two major methods of assessing early vocabulary have complementary weaknesses. Structured tests are primarily measures of concrete nouns, whereas measures of vocabulary diversity derived from a spontaneous language sample are particularly sensitive to high-frequency words such as verbs and functors. The multiple correlation between 24-month-old children's vocabulary as reported by the parent on the CDI: Toddlers was higher than either of the simple correlations with a structured vocabulary test or a measure of vocabulary derived from a spontaneous language sample. A multiple regression confirmed that each criterion measure was related to a distinct, significant portion of the variance in CDI vocabulary. Thus the parent report measure appears to assess a broader vocabulary range than either one of the criterion measures can do by itself. Further, as Bates et al. (1994) have shown, information derived from the CDIs concerning the broad composition of vocabulary (common nouns, verbs and adjectives, and functors) is highly consistent with other information on children's vocabularies. These and other findings lead us to conclude that the high resolution of the CDIs provides high-quality information in the details as well as in the overall scores.

In the norming study for the CDIs, almost 1,800 parents completed either the CDI: Infants (8–16 months; $N = 659$) or the CDI: Toddlers (16–30 months; $N = 1,130$). The CDI: Infants includes 396 words organized into 19 semantic categories; parents report whether their child comprehends and/or produces each word. The CDI: Toddlers includes 680 words organized into 22 semantic categories; production only is reported. At least 60 parents completed an inventory at each month of age.

The lexical norms reported here were derived from the norming study of *The MacArthur Communicative Development Inventories* (Fenson et al., 1993). The development of the inventories, the norming study, and the development of the LEX program were supported by grants from the MacArthur Research Network on Early Childhood Transitions. We are grateful to our colleagues on that project, E. Bates, J. Reilly, J. S. Reznick, and D. Thal for their contributions to this research, and to S. J. Pethick for his invaluable assistance in preparing the data for lexical analysis. Correspondence concerning this article should be addressed to P. S. Dale, Psychology Department Box 351525, University of Washington, Seattle, WA 98195 (e-mail: pdale@u.washington.edu).

Fenson et al. (1994) have extensively analyzed the normative data from this project on the basis of vocabulary totals for each child (along with other measures, such as gesture and sentence complexity). Appendix B of Fenson et al. (1994) includes for each word the first age at which parents reported at least 50% of all children to be comprehending and/or producing that word. However, more complete developmental information is of considerable interest. As illustrated in Figures 1–3, the rate of increase for specific words follows qualitatively different functions. Fenson et al. (1994) found that linear, quadratic, and cubic components accounted for nearly all the variance for most words. They suggested psychological interpretations of

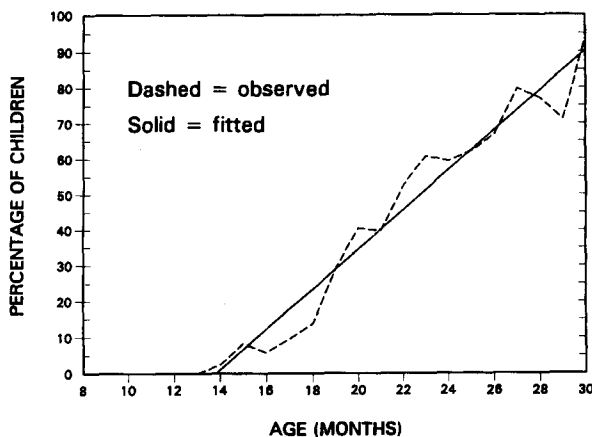


Figure 1. A sample word, *man*, whose growth is fitted by a function with a significant linear component only. From *Variability in Early Communicative Development*, by L. Fenson, P. S. Dale, E. Bates, J. S. Reznick, D. Thal, and S. J. Pethick, 1994, *Monographs of the Society for Research in Child Development*, 59. Copyright 1994 by the Society for Research in Child Development. Reprinted with permission.

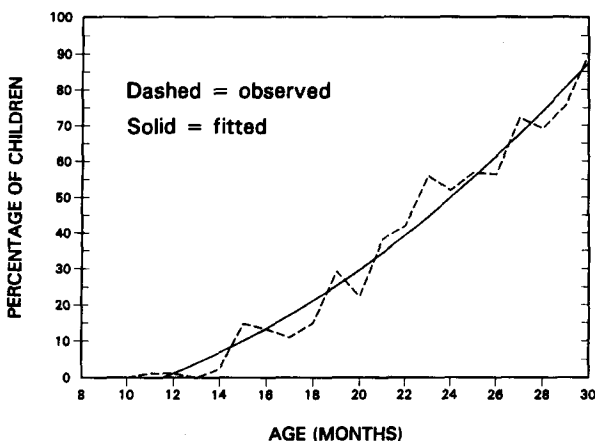


Figure 2. A sample word, *purse*, whose growth is fitted by a function with significant linear and quadratic components only. From *Variability in Early Communicative Development*, by L. Fenson, P. S. Dale, E. Bates, J. S. Reznick, D. Thal, and S. J. Pethick, 1994, *Monographs of the Society for Research in Child Development*, 59. Copyright 1994 by the Society for Research in Child Development. Reprinted with permission.

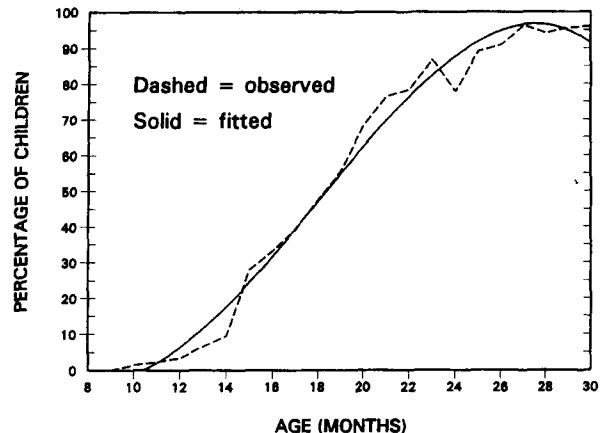


Figure 3. A sample word, *airplane*, whose growth is fitted by a function with significant linear, quadratic, and cubic components. From *Variability in Early Communicative Development*, by L. Fenson, P. S. Dale, E. Bates, J. S. Reznick, D. Thal, and S. J. Pethick, 1994, *Monographs of the Society for Research in Child Development*, 59. Copyright 1994 by the Society for Research in Child Development. Reprinted with permission.

these components. The capability of automated searching of the norms and of production of machine-readable output is also highly desirable. The full set of normative data in hardcopy form would be quite lengthy (containing 17,328 values) and difficult to query for alternative organizations of the data, such as by age, word type, or frequency. For these reasons, a flexible database program, LEX, has been constructed that contains the percentage of children at each age whose parents reported occurrence in comprehension and in production.

Users can query LEX in two modes. First, they can provide a single word, an existing CDI semantic category (such as food and drink words, or action words), or a user-defined list (e.g., internal state words such as *see/think/love/bad*, which fall in several CDI categories). The program then provides the month-by-month figures (from all three sources: Infant Comprehension, Infant Production, Toddler Production) for each word on the list. Second, users may provide a set of search criteria based on desired range of mastery at a particular age (e.g., the food and drink words that are produced by 50–75% of all subjects at 24 months). The program will locate the words that satisfy these criteria and provide month-by-month figures for each word. The selected words can be ordered in the output by semantic category, alphabetically, or in descending order of frequency. Output of this program can be provided in screen, hardcopy, dBase III or ASCII file, as desired. dBase III format for output files is particularly useful, since this format can be read by spreadsheets, SPSS, and many other programs.

Hardware Requirements and Program Availability

The LEX program runs on IBM-compatible computers, DOS version 3.3 or later, with a high-density 3.5-in. drive and approximately 1.4 Mb hard drive space. The program

will automatically detect the presence of a color monitor or LCD screen and will configure for best readability in either case. For occasional users of the program who wish to conserve hard disk space, a batch file is included that loads the program on the hard disk from a diskette, runs it, and deletes the program after use. The program is available without charge from P. Dale, Psychology Department Box 351525, University of Washington, Seattle, WA 98195.

REFERENCES

- BATES, E., MARCHMAN, V., THAL, D., FENSON, L., DALE, P., REZNICK, J. S., REILLY, J., & HARTUNG, J. (1994). Developmental and stylistic variation in the composition of early vocabulary. *Journal of Child Language*, **21**, 85-123.
- DALE, P. S. (1991). The validity of a parent report measure of vocabulary and syntax at 24 months. *Journal of Speech & Hearing Research*, **34**, 565-571.
- DALE, P. S., BATES, E., REZNICK, J. S., & MORISSET, C. (1989). The validity of a parent report instrument of child language at 20 months. *Journal of Child Language*, **16**, 239-249.
- FENSON, L., DALE, P. S., BATES, E., REZNICK, J. S., THAL, D., & PETHICK, S. J. (1994). Variability in early communicative development. *Monographs of the Society for Research in Child Development*, **59** (Serial No. 242).
- FENSON, L., DALE, P. S., REZNICK, J. S., THAL, D., BATES, E., HARTUNG, J. P., PETHICK, S., & REILLY, J. S. (1993). *The MacArthur communicative development inventories: User's guide and technical manual*. San Diego: Singular.

(Manuscript received October 27, 1994;
revision accepted for publication February 27, 1995.)