Applied Psycholinguistics 12 (1991), 23–45
Printed in the United States of America

1221

The relation of utterance length to grammatical complexity in normal and language-disordered groups

HOLLIS S. SCARBOROUGH
Brooklyn College, City University of New York

LESLIE RESCORLA
Bryn Mawr College

HELEN TAGER-FLUSBERG
University of Massachusetts, Boston

ANNE E. FOWLER
Haskins Laboratories

VICKI SUDHALTER
Institute for Basic Research on Developmental Disabilities

ADDRESS FOR CORRESPONDENCE
H. Scarborough, Department of Psychology, Brooklyn College, CUNY, Brooklyn, NY 11210

ABSTRACT
Mean length of utterance (MLU) in morphemes was examined as a predictor of the grammatical complexity of natural language corpora of normal preschoolers and of children and adolescents with delayed language, Fragile X syndrome, Down syndrome, and autism. The Index of Productive Syntax (IPSyn) served as the measure of syntactic and morphological proficiency. For all groups, a strong curvilinear association between measures was found across the MLU range from 1.0 to about 4.5. Correlations were weaker when MLU exceeded 3.0 than during earlier stages of language development, however, confirming previous suggestions that MLU becomes less closely associated with grammatical development as linguistic proficiency increases. For the language-disordered groups, moreover, the curves relating the two measures differed from the curves for the normal preschoolers because MLU frequently overestimated actual IPSyn scores. The results are discussed with respect to the use of MLU in conjunction with other measures of syntactic complexity in the study of atypical language development.

Mean length of utterance (MLU) is frequently used to describe individual differences and developmental changes in linguistic proficiency. Despite its wide acceptance as a general index of grammatical development in research and practice, there remain some questions about how well MLU serves as a reliable estimate of syntactic complexity and, consequently, about how

© 1991 Cambridge University Press 0142-7164/91 $5.00 + .00

confidently one can assume that children matched for utterance length will have equivalent grammatical abilities. In the present study, the relation of MLU to scores on a new measure of grammatical complexity, the Index of Productive Syntax (IPSyn), was examined for children and adolescents with delayed language, Fragile X syndrome, Down syndrome, and autism, and also for normal preschool children. Our central question was a methodological one: Does the same relation between utterance length and grammatical complexity hold for language-disordered speakers as for normal preschoolers?

THE LENGTH–COMPLEXITY RELATION

Normal preschoolers

MLU has been used widely to describe the early stages of syntax from the onset of two-word combinations to the genesis of complex syntax in normally developing preschool children (e.g., Bloom, 1970, 1973; Brown, 1973; de Villiers & de Villiers, 1973; Klima & Bellugi, 1966; Nelson, 1973; Newport, Gleitman, & Gleitman, 1977; Shipley, Smith, & Gleitman, 1969; Wells, 1985). Although MLU simply reflects the mean number of morphemes per utterance in a speech sample, the measure has been found to be associated with a number of grammatical competencies, including the expression of semantic relations (Bloom, Lightbown, & Hood, 1975), the number and diversity of grammatical categories such as verb, pronoun, or article (Newport et al., 1977; Shipley et al., 1969), productive mastery of grammatical inflections and functors (Brown, 1973; de Villiers & de Villiers, 1973), and the emergence of increasingly complex negative and interrogative constructions (Klima & Bellugi, 1966).

Nevertheless, MLU has sometimes been criticized with respect to its reliability (Klee & Fitzgerald, 1985; Kramer, James, & Saxman, 1979; Rondal, Ghiotto, Bredart, & Bachelet, 1987) and validity (Crystal, 1974; Klee & Fitzgerald, 1985). In particular, MLU appears to be most useful primarily for evaluating the earliest stages of development, beyond which it appears to become less reliable and less closely tied to grammatical development. For example, in a study of 18 preschoolers with MLUs between 2.54 to 3.82 morphemes, Klee and Fitzgerald (1985) found that the frequency and diversity of syntactic constructions at different developmental levels did not correspond closely to MLU. On the other hand, strong associations were observed between utterance length and grammatical skill in a group of 21 younger children (with MLUs from 1.05 to 3.06) studied by Rondal, Ghiotto, Bredart, and Bachelet (1987), although indeed the reliability of MLU declined somewhat with age in that study as well. Accordingly, Rondal and colleagues concluded that MLU is a valid index of development until a mean length of approximately 3.0 morphemes is achieved, but that the association between syntactic development and increased utterance length is weaker beyond that point.

Atypical language samples

Given the ease with which MLU is calculated and the association it has often been shown to have with grammatical development in normal preschoolers, a number of investigators have found MLU to be a useful window through which to examine the emerging grammars of children and adolescents with atypical language development. To date, the focus has been on children with language delay (Johnston & Kamhi, 1984; Johnston & Schery, 1976; Morehead & Ingram, 1973) and with Down syndrome (Fowler, 1985; Fowler, Gelman, & Gleitman, 1980; Rondal, 1978; Rondal, Ghiotto, Bredart, & Bachelet, 1987). In light of the criticisms that have been advanced regarding the reliance on MLU for the evaluation of normal preschoolers, however, it is important that the relation between MLU and grammatical complexity be directly examined for atypical groups. Of particular concern is the possibility that the aforementioned weakening of this association with age seen in normal samples may become exacerbated in language-impaired children, whose chronological and mental ages are often sharply divergent from those of the normally developing toddlers with whom they may be compared.

In this article, we examine the association between MLU and grammatical complexity in normally developing preschoolers and in samples of children and adolescents with early language delay, Fragile X syndrome, Down syndrome, and autism. Although the language data were collected by different investigators for different purposes, identical coding and scoring procedures were applied to all groups so that the results could be legitimately compared. As our criterion for validation, a new measure of grammatical complexity, the IPSyn (Scarborough, 1990) was used.

We asked two different questions about the association between MLU and grammatical complexity in these atypical language samples. First, we sought to determine whether the overall relation between measures is as strong, and whether the strength of the relation weakens as MLU increases as observed in normally developing children. With respect to cases with Down syndrome, Rondal and colleagues (1988) indeed obtained a high correlation between MLU and scores on the Language Assessment, Remediation, and Screening Procedure (LARSP; Crystal, Fletcher, & Garman, 1976) in a study of 21 children with MLUs from 1.0 to 3.5, and similar results were obtained in Fowler's (1988) detailed grammatical analyses of a few longitudinal cases. Comparisons of the strength of the MLU–complexity association at different linguistic levels and for other diagnostic groups have not been previously examined, however.

The second question, of interest only if consistent associations are in fact observed between MLU and grammatical complexity in children with atypical language development, concerns the similarity of the predictions derived from MLU about syntactic complexity in different groups. For instance, even if equally strong associations are obtained for each diagnostic group, the predicted level of grammatical skill might be consistently lower (or higher) for different groups at a given MLU level. In other words, in some groups or individuals, MLU might consistently *over*estimate syntactic abilities, while in others it might *under*estimate them. In Rondal and colleagues' (1988) Down syndrome sample, for instance, the slopes of the regression lines relating complexity scores to MLU were less steep for eight of nine analyses relative to the reported slopes for a normal sample studied by these investigators (Rondal et al., 1987). Also, in comparisons of Down syndrome samples with MLU-matched normal samples, observed syntactic and morphological production abilities have been generally similar, but some greater grammatical limitations have been evident for children at higher MLU levels (e.g., Fowler, 1985; Fowler et al., 1980; Rondal, 1978; Wiegel-Crump, 1981). Similarly, while most comparisons of language-delayed and normal preschoolers of equivalent MLU have revealed similar patterns of syntactic production, there is also some evidence that MLU may overestimate morphological abilities of children with language delays (Johnston & Kamhi, 1984).

THE INDEX OF PRODUCTIVE SYNTAX

As noted by Klee and Fitzgerald (1985), the examination of the length-complexity relation has been hampered by a lack of consensus in defining and measuring the grammatical complexity of children's language production. Grammatical profiles based upon levels of grammatical organization defined by the LARSP (Crystal et al., 1976) were used as the basis for evaluating syntactic proficiency in the two aforementioned studies of the validity of MLU in normal preschoolers (Klee & Fitzgerald, 1985; Rondal et al., 1987). To derive such profiles requires the preparation of a comprehensive frequency distribution of all grammatical productions by the child, which are classified by type and by developmental stage so that frequency and diversity values can be computed. Such a comprehensive analysis was simply not feasible for examining the large number of corpora needed to carry out the group comparisons in the present study, in which a total of 273 language samples (105 normal, 168 disordered) from 95 subjects (45 normal, 50 disordered) were analyzed. Instead, a recently developed measure, the IPSyn (Scarborough, 1990), was used.

The IPSyn was developed as a means of scoring the grammatical complexity of child language corpora so as to evaluate individual differences in syntactic development in a relatively efficient manner for research purposes. IPSyn scores are derived from 100-utterance language corpora by reviewing the utterances for the presence of 56 syntactic and morphological forms. Noun and verb phrase elaborations, question and negation forms, and sentence constructions are scored, and the items encompass a range of complexity that is typically observed in the language production of preschool children, from "noun," "verb," "intonational question," and "two-word combination" at the low end of the scale, to such items as "past tense copula," "tag question," and "relative clause" at the high end of the scale. Up to two instances of each form may be credited (i.e., 0, 1, or 2 points are

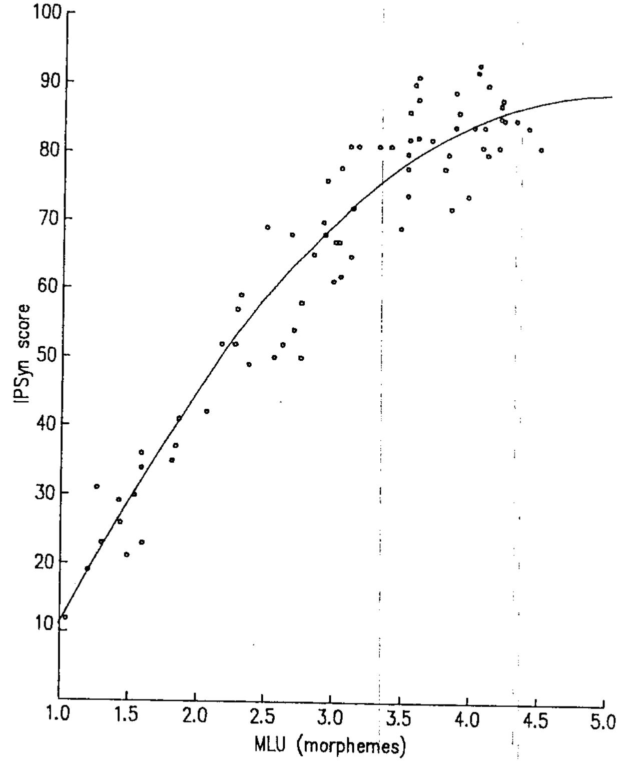

Figure 1. Relation of IPSyn scores to MLU in a group of 15 preschoolers studied longitudinally.

awarded per item), and points are summed across items to provide a summary score. A brief description of the 56 IPSyn items is provided in the Appendix.

Good reliability and age differentiation were demonstrated for the IPSyn within a longitudinal sample of 15 normal preschoolers whose speech was recorded during mother–child play sessions at ages 24, 30, 36, 42, and 48 months (Scarborough, 1990). Both IPSyn and MLU scores were derived from these 75 corpora. Mean agreement between scorers was 94% for the IPSyn, and correlations were high between successive 100-utterance (r = .99) and 50-utterance (r = .96) corpora drawn from the same transcripts. IPSyn scores, but not MLUs, were found to increase significantly over the last three observations (i.e., between 36–42 months and 42–48 months), suggesting that the IPSyn scale provided greater age differentiation than did utterance length.

The bivariate distribution of the two measures in that study is shown in Figure 1. MLU values ranged from 1.01 to 4.48 across the 75 observations, and IPSyn scores from 12 to 93 were obtained. The overall correlation between MLU and IPSyn was high (r = .94). As was suggested by Rondal and colleagues (1987), better prediction of grammatical complexity from utterance length was obtained at earlier periods of language development. For the 33 corpora with MLU less than 3.0, a correlation of .93 was obtained with IPSyn, whereas for the 42 corpora with MLU greater than 3.0, a correlation of only .55 was obtained. A multiple regression analysis of data for the entire age range revealed that the relation of IPSyn to MLU was curvilinear, with the quadratic term accounting for an additional 4% of total variance (R = .96). The curve described by the regression equation,

$$\text{IPSyn} = 49.202(\text{MLU}) - 4.94(\text{MLU}^2) - 33.566$$

is shown in Figure 1.

In the analyses to be reported, the relation of IPSyn to MLU was analyzed in a similar manner for a new, cross-sectional sample of normal preschoolers, as well as for the language-disordered groups. Many aspects of the methodology were common to all or most of the groups, but the studies also differed in several respects. Unless noted otherwise, all language samples were transcribed from recorded play sessions lasting 20 minutes or more, and 100 successive utterances by the subject were identified within each transcript.[1] No unintelligible utterances, immediate exact or reduced imitations or self-repetitions, or routines (counting, singing, and so forth) were included in these corpora. Each corpus was then scored for IPSyn according to the procedures described by Scarborough (1990) and for MLU according to conventional guidelines for counting morphemes (Brown, 1973; MacWhinney & Walter, 1987; Miller & Chapman, 1985).

For each group, we computed correlations between MLU and IPSyn scores: first across the entire MLU range and then separately for corpora in which MLU was less than 3.0 and for those in which MLU exceeded 3.0 morphemes. Multiple regression analyses were used to describe the shape of the relation between measures, and each group's curve was compared to that for the longitudinal normal group (presented in Figure 1). Finally, we examined the magnitude and direction of differences between observed IPSyn scores and scores predicted by the regression equation derived for the normal preschoolers.

STUDY 1: NORMAL REPLICATION

Subjects and language samples

Thirty normal preschoolers (13 boys, 17 girls) were selected from the larger research project from which the 15 longitudinal subjects had previously been drawn (Scarborough, 1989), so that there were: (a) 6 children at each age (24, 30, 36, 42, 48 months); (b) 15 children from lower-middle class and 15 from upper-middle class backgrounds, on a scale based on parental education and occupational prestige (Hollingshead & Redlich, 1958); (c) no children with family histories of reading problems in the sample; and (d) no children who had previously been included in the longitudinal sample. In all other respects, the selection of the subjects from those available was random. All children were from homes in which only Standard English was spoken, and all consistently earned normal IQ scores in preschool assessments.

From available transcripts of mother–child play sessions, 100-utterance corpora were defined as described earlier and were scored for MLU and IPSyn. Because these measures were derived using the same procedures and the same scorers as for the longitudinal sample, interscorer reliability was not reevaluated.

Results

MLU ranged from 1.40 to 4.81 in the sample, and IPSyn scores ranged from 26 to 94. The simple correlation between the two measures was .92, and the linear relation was again stronger for the 13 corpora with MLU below 3.0 (r = .93) than for the 17 corpora with higher MLUs (r = .58), as was observed for the longitudinal normal group. As before, a multiple regression analysis revealed a significant quadratic component (R = .95), yielding a regression equation very similar to that for the longitudinal group,

$$\text{IPSyn} = 46.444(\text{MLU}) - 4.611(\text{MLU}^2) - 28.705.$$

This curve and the bivariate distribution of the two measures in the cross-sectional group are shown in Figure 2, in comparison to the regression function for the longitudinal sample (from Figure 1).

The mean difference between observed and expected (defined by the curve for the longitudinal sample) IPSyn scores for the 30 cross-sectional observations was only −0.2 points, confirming the excellent agreement between the two estimates of the relation of complexity to length in normal preschool children aged 2–4 years. Seventy percent of observed scores were within 5 or fewer points of the expected value, and there were only two differences greater than 10 points: one a 12-point underestimation, and the other a 19-point overestimation.[2] In sum, the relation of IPSyn scores to MLUs was very similar in the cross-sectional and longitudinal normal samples.

STUDY 2: EARLY LANGUAGE DELAY

Subjects and language samples

Two groups of preschoolers with early language delay were studied. The *longitudinal* group included 5 children (4 boys, and 1 girl) of normal IQ who, alone among a larger group from a study of the antecedents of reading disability (Scarborough, 1989), were not combining words in a regular manner at age 30 months. Language samples collected at their subsequent examinations at ages 36, 42, and 48 months were also analyzed for the present report.[3] These subjects, 4 of whom later developed reading problems, are described fully elsewhere (Scarborough & Dobrich, 1985, 1990).

The *cross-sectional follow-up* group of language-delayed preschoolers

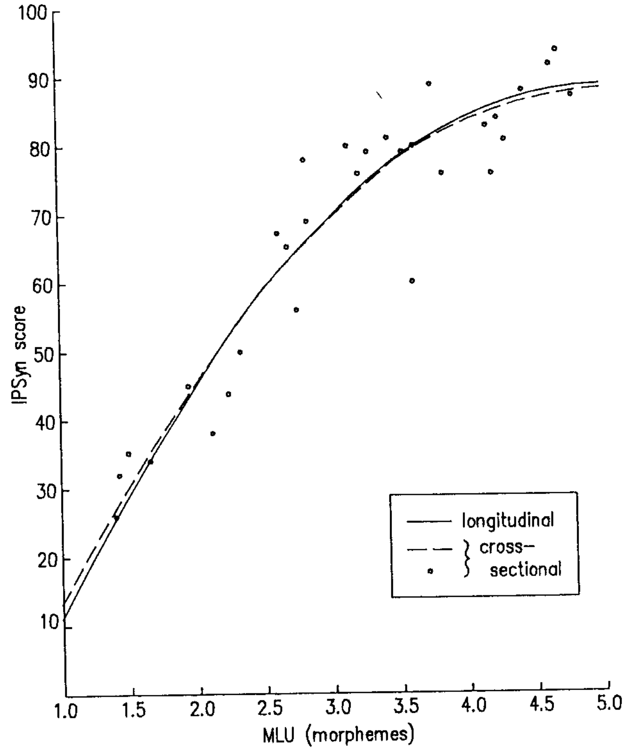

Figure 2. Relation of IPSyn scores to MLU in a cross-sectional sample of preschoolers, aged 24–48 months, in comparison to the regression curve for the longitudinal normal group (from Figure 1).

included 15 boys whose scores were within 3 months of chronological age on the Reynell Receptive Language Scale (Reynell, 1985), but at least 6 months delayed on the Expressive scale. On average, their Receptive scores were .25 *SD* above the mean for their age, and their Expressive scores were 1.70 *SD* below the norm. By maternal report at ages 24–31 months, most were producing fewer than 30 vocabulary words and/or no (or very few) word combinations (Rescorla, 1989). All scored above 85 on the Bayley Scales of Infant Development (Bayley, 1969). Language samples collected for each child at a single follow-up examination between ages 36–48 months were analyzed. Rescorla and Schwartz (1990) provided additional information about this group.

Interscorer reliability

Five (26%) of the longitudinal corpora were independently scored for IPSyn and MLU by two coders, who agreed on 94% of their coding judgments, on average, for each measure. All of the cross-sectional language

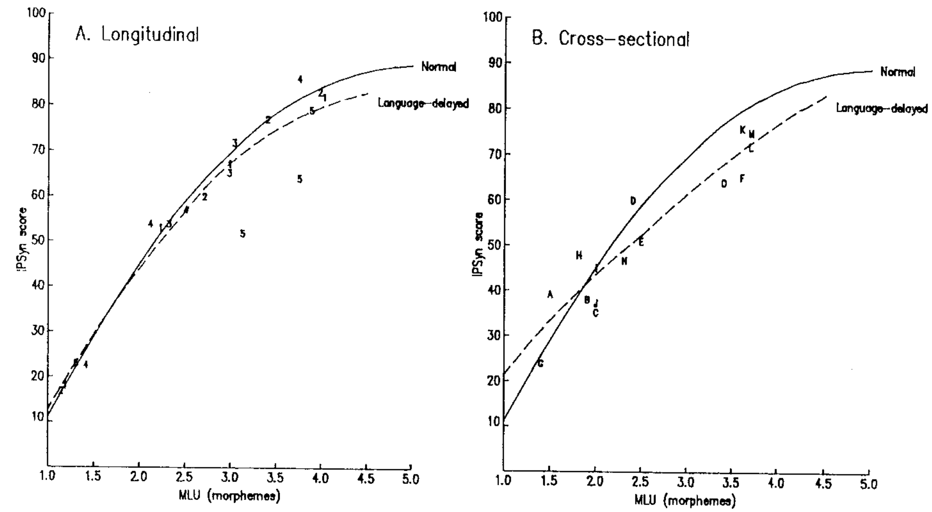

Figure 3. Relation of IPSyn scores to MLU in language-delayed preschoolers from the longitudinal and cross-sectional groups, in comparison to the regression curve for the longitudinal normal group.

samples were independently coded for IPSyn, and 96% agreement between scorers was obtained; reliability was not assessed for MLU, which was computer scored using the CHAT system (MacWhinney & Walter, 1987).

Results

MLU ranged from 1.14 to 4.03 in the longitudinal group and from 1.40 to 3.70 in the cross-sectional group, and IPSyn ranges were 17 to 84 and 24 to 76, respectively; these ranges were thus somewhat narrower than those of the two normal groups. Nevertheless, the overall correlation between measures was .92 for the longitudinal group and .94 for the cross-sectional group. For the former there was a decidedly stronger correlation between scores for the 11 corpora with MLU less than 3.0 (r = .98) than for the 8 corpora with higher MLUs (r = .64), as in the normal groups; for the cross-sectional group, these values were more equal (r = .77 and .69). Of course, these results should be interpreted cautiously in light of the small numbers of observations and restricted ranges in these subsets of the data.

The bivariate distributions for the two groups of language-delayed preschoolers, and the quadratic curves describing them, are shown in Figure 3 in comparison to the regression curve for the longitudinal normal group (from Figure 1). The mean difference between observed and expected IPSyn scores was −2.3 points for the longitudinal and −4.1 points for the cross-sectional language-delayed sample. Prediction errors were larger when MLU exceeded 3.0 (−4.8 and −9.1 points, respectively, for the two groups) than when MLU was below 3.0 (−0.5 and −1.6, respectively). As can be seen in Figure 3, these errors of prediction are primarily attributable to four instances (cross-sectional cases F and O, and two successive observations of longitudinal case 5) whose IPSyn scores were 12 to 21 points below the curve for the normal preschoolers. These findings suggest that for most of the language-delayed children, a good estimate of grammatical development could be made on the basis of the relation of MLU to IPSyn in normal preschoolers, but that for some cases (12% of the present samples) who were beyond the earliest stages of language acquisition, utterance length severely overestimated actual syntactic production.

STUDY 3: FRAGILE X SYNDROME

Subjects and language samples

This group included 19 males, aged 5–19 years, who were studied cross-sectionally. Their standard scores on the Vineland Adaptive Behavior Scales (Sparrow, Balla, & Cicchetti, 1984) ranged from 20 to 79. These subjects were drawn from a larger group of children and adolescents with Fragile X syndrome, a chromosomal abnormality that is a common cause of inherited mental retardation. Other subjects from that group were excluded from the present analyses because their language samples contained fewer than 50 utterances, or because they earned Communication Age Equivalent scores below age 3 years on the Vineland Scales. Some aspects of the speech and language of the subjects of the present study and of other children and adolescents with Fragile X syndrome have been described elsewhere (Sudhalter, Cohen, Silverman, & Wolf-Schein, 1990; Wolf-Schein et al., 1987).

The subject's conversational partner was usually one of his parents, but in several cases, in order to obtain language samples of sufficient length for analysis, utterances during conversation with a psychologist (who entered the room 20 min after the recording began) were included in the language samples. As was not the case for the other groups studied in this investigation, moreover, the subjects' parents were asked to elicit specific information (about color knowledge and comprehension of prepositions) during the play sessions. These language corpora were thus somewhat less naturalistic than those for the other groups.

Interscorer reliability

Five (26%) of the language samples were independently coded for IPSyn and for MLU by two scorers, who agreed on 96% of coding judgments for IPSyn and 95% for MLU.

Results

MLUs ranged from 1.33 to 4.83 and IPSyn scores ranged from 25 to 81 for the 19 cases with Fragile X syndrome. The correlation between measures

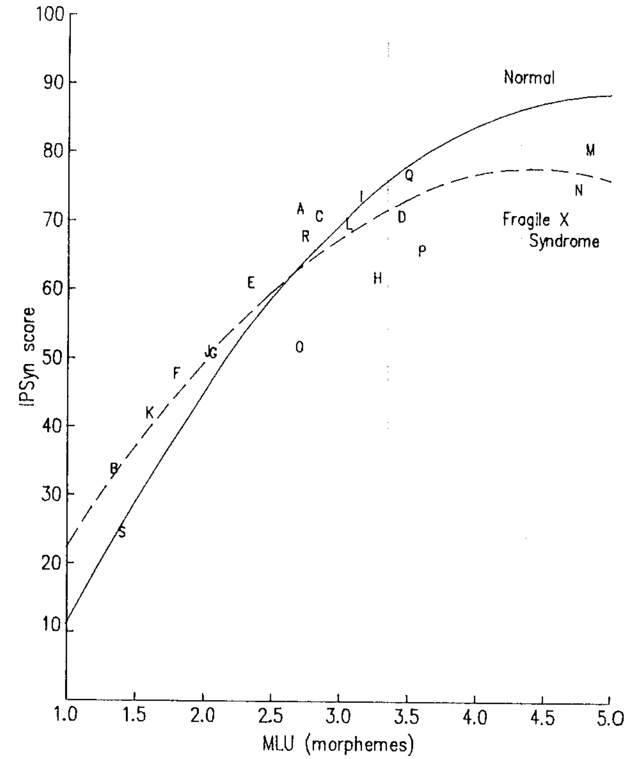

Figure 4. Relation of IPSyn scores to MLU in a group of 19 children and adolescents with Fragile X syndrome, in comparison to the regression curve for the normal preschool group.

was high overall ($r = .86$) and was higher for the 10 cases with MLU below 3.0 ($r = .92$) than for the 6 cases with higher MLUs ($r = .59$).

The regression curve and bivariate distribution of the two measures for the Fragile X syndrome group, with cases identified by the letters A to S in order of increasing chronological age, are shown in Figure 4 in comparison to the regression curve for the normal longitudinal preschool group. The mean difference between observed and expected IPSyn scores was only -0.1 points overall, with underestimation occurring at lower MLU values ($M = 4.8$ points for the 10 observations with MLU less than 3.0), and overestimation at higher MLU values ($M = -6.7$ points for the remaining 9 observations). For 3 (16%) of the 19 cases (H, N, and P in Figure 4), expected IPSyn scores exceeded actual scores by 12 points or more. In sum, the relation of complexity to length in this group was generally similar to the normal pattern, but for some individuals MLU overestimated IPSyn scores quite severely, as was seen in the language-delayed groups.

STUDY 4: DOWN SYNDROME

Subjects and language samples

This group included five cases (1 boy, 4 girls) with Down syndrome, the other common cause of congenital mental retardation. Each child was first examined between ages 51-67 months, at which time each was producing at least some two-word combinations in spontaneous speech. These subjects were subsequently seen regularly (often monthly) over a period of 43-63 months. For each child, two observations per year were selected to provide the subset of 51 corpora analyzed for this report. Their IQ scores on the Stanford-Binet Intelligence Scale (Terman & Merrill, 1973) at age 7 years ranged from 38 to 57. Fowler (1988) provided detailed information about the larger data base from which the corpora for this group were drawn.

Play sessions were transcribed from audiorecordings and notes taken during the taping. For all observations, conversational partners were researchers who became familiar to the subjects over the course of the study.

Interscorer reliability

MLU was computer scored according to Brown's (1973) guidelines for counting morphemes, and two scorers completely reviewed each corpus for IPSyn. Because their coding judgments were not made independently, however, reliability could not be assessed.

Results

MLUs from 1.26 to 4.22 and IPSyn scores from 26 to 83 were obtained for the Down syndrome group. The correlation between measures was .93 overall, .84 for the 37 instances in which MLU was less than 3.0, and .60 for the 14 observations involving higher MLUs.

In Figure 5, the regression curve and the bivariate distribution of the two measures for the Down syndrome group are shown in comparison to the regression curve for the normal preschool group. In the multiple regression analysis, only the linear component accounted significantly for variance in IPSyn scores, yielding the regression equation,

$$\text{IPSyn} = 18.40(\text{MLU}) + 3.998.$$

The mean difference between observed and expected IPSyn values was −5.1 points for the Down syndrome group. Larger errors of prediction occurred when MLU exceeded 3.0 ($M = -6.7$ points), but even when MLU was lower, the differences were not negligible ($M = -4.8$ points), a pattern that was not observed in any other group in the present study. Actual IPSyn scores exceeded predicted scores by 12 points or more for 10 of the 51 observations (20%), and these severe overestimates were most characteristic of the child with the highest MLU (case E), for whom 5 of

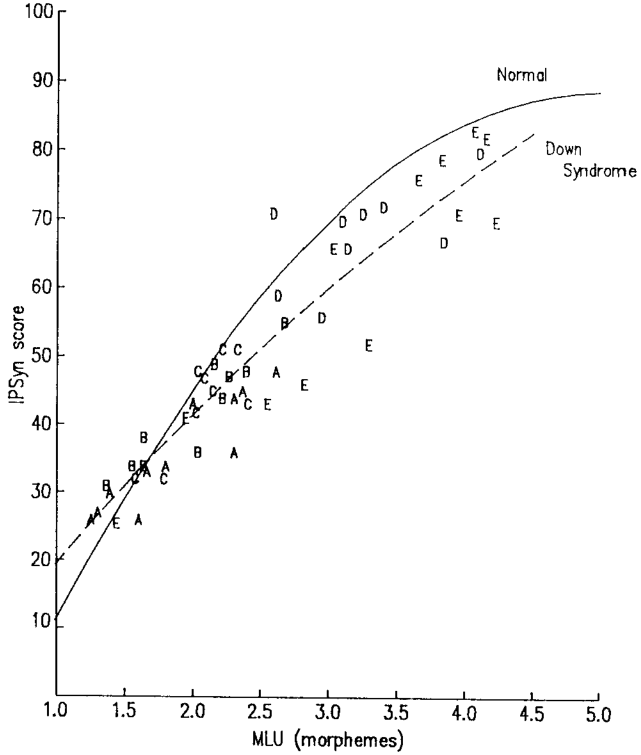

Figure 5. Relation of IPSyn scores to MLU for five cases with Down syndrome (identified with the letters A through E), in comparison to the regression curve for normal preschoolers.

the 12 observations (42%) were more than 12 points below the curve for the normal preschoolers.

STUDY 5: AUTISM

Subjects and language samples

This group included 6 autistic boys who were studied longitudinally. Their scores ranged from 61 to 108 ($M = 89$) on the Leiter International Performance Scale. Their ages at the time of the initial observations were from 40-91 months, and each child was seen bimonthly for 15-26 months, yielding 8-13 observations per child and a total of 64 language samples. Following Rutter (1978) and more recent proposals for defining autism (Cohen, Paul, & Volkmar, 1986, 1987; Denckla, 1986), the cases were identified by the presence or definite history of all of the following characteristics: onset prior to 30 months; gross and sustained impairments in socialization and social relations; delays and deficits in language and communicative development; and pervasive repetitive activities. Children medicated with behavior-controlling drugs were excluded from the study. Additional information about this group is reported elsewhere (Tager-Flusberg et al., 1990).

Interscorer reliability

Forty-nine (75%) of the corpora were independently coded for IPSyn by two scorers, who agreed on 97.9% of their coding judgments, on average. MLU was computer scored (Miller & Chapman, 1985).

Results

MLUs of the autistic children ranged from 1.16 to 4.12, and their IPSyn scores ranged from 19 to 75. Although the correlation between measures was .87 across the full MLU range, and .84 for the 44 observations with MLU below 3.0, the association of length with complexity was very weak for the 20 observations with higher MLUs ($r = .21$).

The regression curve and bivariate distribution of IPSyn and MLU scores for the autistic children, in comparison to the regression curve for the normal preschoolers, are shown in Figure 6. The following regression equation was derived for the autistic group,

$$\text{IPSyn} = 37.124(\text{MLU}) - 4.451(\text{MLU})^2 - 13.296.$$

As can be seen in Figure 6, although the ranges of the two measures were similar in the two groups, the relation between IPSyn and MLU was clearly different, with much lower IPSyn scores earned by the autistic than by the normal subjects at higher MLU levels.

The mean difference between observed and expected IPSyn scores was −6.1 points overall, and the degree of overestimation was much larger for the 20 observations with MLU greater than 3.0 morphemes ($M = -15.4$ points) than for the 44 observations with lower MLU ($M = -1.8$ points). For the two least linguistically advanced children (cases A and F, with mean MLUs of 1.48 and 1.62, respectively), only 1 overestimation by more than 12 points was obtained in 21 observations (5%), whereas for the children with the longest utterances (cases B and D, with mean MLUs of 3.02 and 3.69, respectively), a total of 12 of 20 observations (60%) involved such large differences. For the cases with intermediate MLU levels (cases C and E, with MLU 2.48 and 2.52, respectively), there were 6 discrepant observations out of 23 (26%). This was not merely a between-subjects effect, however, since for each case individually differences increased as MLU rose ($r = -.50$ to $-.88$).

DISCUSSION

What can the mean utterance length in a sample of natural language tell us about its grammatical complexity? The present findings suggest that

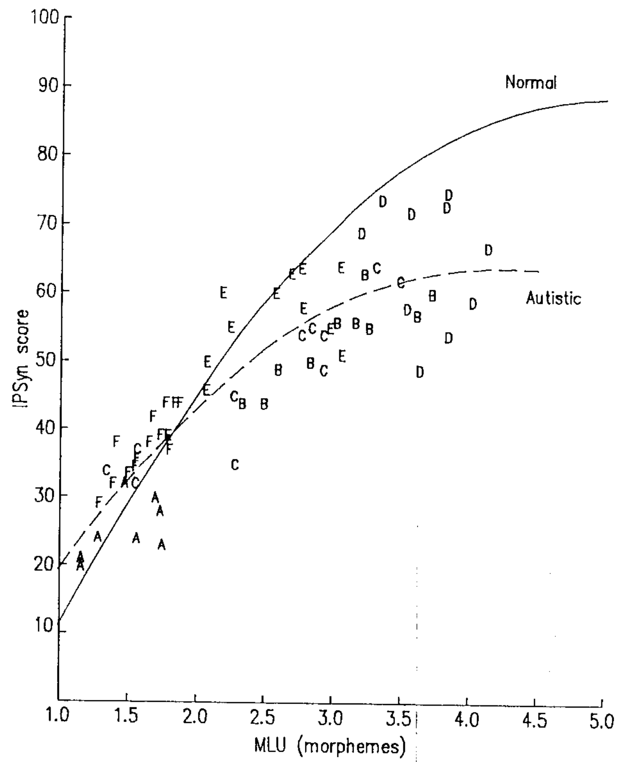

Figure 6. Relation of IPSyn scores to MLU for six autistic children (identified with the letters A through F), in comparison to the regression curve for normal preschoolers.

the answer to that question depends on the range of MLU levels across the language corpora to be analyzed and on the diagnostic group to be studied.

With respect to normal 2- to 4-year-old children, very similar results were obtained in the present cross-sectional sample as were found previously in a longitudinal investigation (Scarborough, 1990). For both of these normal preschool groups, across the entire range of MLUs from about 1.0 to 4.5 morphemes a strong curvilinear relation of MLU to IPSyn scores was seen. Previously, on the basis of two recent studies of the complexity-to-length relation in smaller samples with more restricted MLU ranges (Klee & Fitzgerald, 1985; Rondal et al., 1987), Rondal and colleagues (1987) concluded that utterance length provides a good estimate of syntactic ability for MLUs up to 3.0 morphemes, but that the association between grammatical complexity and utterance length is not as strong or reliable for MLUs above 3.0 morphemes. The present findings, which were based on a different measure of grammatical complexity, provide further support for that conclusion, since higher correlations between measures for lower than for higher MLU ranges were obtained in both normal preschool groups.

Despite differences in age and diagnosis, some patterns of results were quite similar across all language-disordered groups, as summarized in Table 1. First, the overall association between grammatical complexity and utterance length was very strong, but the strength of this association was consistently weaker at higher MLU levels. In these respects, the results resemble those for the normal preschoolers quite closely.

Second, MLU was found to overestimate complexity scores more often for the disordered than for the normal subjects, particularly for children beyond the earliest stages of language development. For MLUs greater than 3.0, mean differences between observed and predicted IPSyn scores were consistently negative for the language-disordered groups, ranging from − 4.8 to − 15.4 points. Consequently, the best-fit regression curves of these groups overlapped the curves for the normal groups quite closely at lower MLU values but tended to fall below the normal curve at higher MLU values, as summarized in Figure 7(A). Moreover, from 11–30% of the individual observations in the five language-disordered groups were characterized by large discrepancies (12 points or more) between expected and observed complexity scores.

The results also suggest that the extent to which MLU overestimates grammatical complexity may not be the same for different diagnostic groups. In particular, this problem appeared to be greatest for the autistic group and least frequent for the children with specific early language delays. In view of the relatively small numbers of subjects per group, however, a firm conclusion on this question cannot be reached on the basis of the evidence from the present study. A partial replication of the apparent difference between the autistic and Down syndrome groups, however, has been provided by Tager-Flusberg, who looked at IPSyn and MLU scores in both an autistic group (included in the present analyses) and a Down syndrome group (Tager-Flusberg et al., 1990). The regression curves derived for the two groups in that study, shown in Figure 7(B), cannot be directly compared to the results of the present analyses because the language corpora from which MLU and IPSyn scores were derived had been defined somewhat differently (i.e., by excluding all *yes/no* responses, single-word proper names, and delayed imitations). Nevertheless, it can be seen that, like Fowler's Down syndrome cases in the present study, Tager-Flusberg's Down syndrome group exhibited greater grammatical complexity at the highest MLU levels than did the autistic children. This not only demonstrates that the present findings probably did not arise merely because the corpora for autistic and Down syndrome cases were analyzed by different coders in different laboratories, but also suggests that this apparent difference between diagnostic groups may prove to be a reliable, albeit localized, one in future research.

Even when diagnostic groups appear similar in the extent to which MLU overestimates grammatical complexity, they may not necessarily do so for

Table 1. *Comparison of groups with respect to linear correlations between MLU and IPSyn, and differences between observed and expected IPSyn scores*

	Normal	Delay: Longitudinal	Delay: Cross-Sectional	Fragile X syndrome	Down syndrome	Autism
All corpora						
Observations	30	19	15	19	51	64
MLU range	1.4–4.8	1.1–4.0	1.4–3.7	1.3–4.8	1.3–4.2	1.2–4.1
r	.92	.92	.94	.86	.93	.87
Mean difference	− 0.2 (6.7)	− 2.3 (6.4)	− 4.1 (7.1)	− 0.1 (8.4)	− 5.1 (7.1)	− 6.1 (10.4)
Severe overestimates*	3%	11%	13%	16%	20%	30%
MLU < 3.0						
Observations	13	11	10	10	37	44
MLU range	1.4–2.8	1.1–3.0	1.4–2.5	1.3–2.8	1.3–3.0	1.2–3.0
r	.93	.98	.77	.92	.84	.84
Mean difference	0.4 (6.6)	− 0.5 (2.8)	− 1.6 (7.0)	4.8 (6.2)	− 4.1 (7.2)	− 1.8 (8.4)
MLU > 3.0						
Observations	17	8	5	9	14	20
MLU range	3.1–4.8	3.0–4.0	3.4–3.7	3.0–4.8	3.0–4.2	3.0–4.1
r	.58	.64	.69	.59	.60	.21
Mean difference	− 0.7 (6.9)	− 4.8 (9.0)	− 9.1 (4.6)	− 6.7 (6.2)	− 7.5 (6.6)	− 15.4 (8.1)

Note: Standard deviations are shown in parentheses.
*Expected IPSyn score exceeds actual score by 12 points or more.

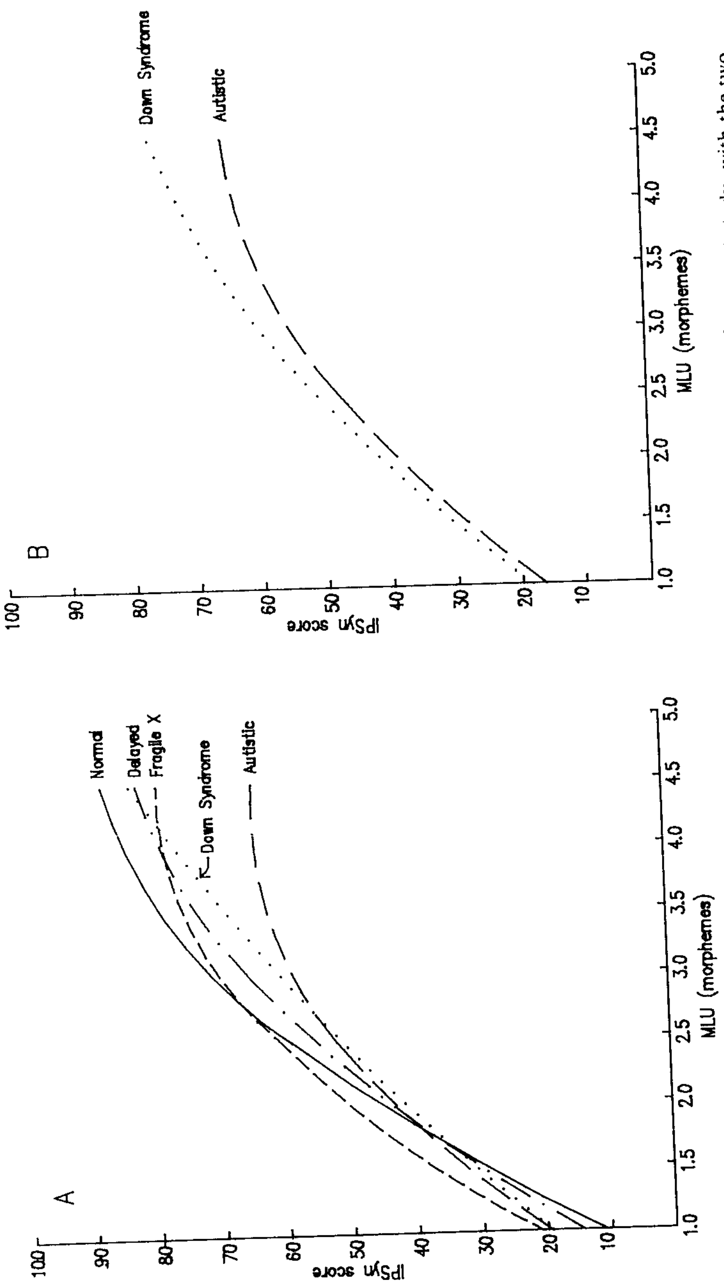

Figure 7. Regression curves showing the relation of IPSyn to MLU for (a) the groups analyzed in the present study, with the two language-delayed samples combined, and (b) autistic and Down syndrome groups whose language corpora were defined somewhat differently than in the present study (Tager-Flusberg et al., 1990).

the same reasons, since the relation of complexity to length may be determined by a variety of linguistic, cognitive, and communicative factors. Children with Down syndrome and Fragile X syndrome, for example, might in fact have severe syntactic limitations, such that (with increasing mental age) their improving cognitive abilities might outstrip their growth in grammatical skill, enabling them to string together longer sequences of words but in structurally simple ways. Evidence from both production and comprehension measures indeed suggests that syntax and morphology may be impaired to a greater degree than other linguistic and cognitive abilities in children with Down syndrome (Fowler, 1990).

It is possible that the autistic children, too, might have syntactic limitations of this sort. Alternatively, the distorted relation of complexity to length in these cases may be better understood by taking into account their autistic propensities. In particular, these children characteristically engage in "repetitive activities" (Baron-Cohen, 1989), and this tendency pervades their attentional and motor performance as well as their language production. Accordingly, these children may be inclined to use only a narrow range of grammatical constructions during the period of any given observation, and this conversational rigidity may thus limit the extent to which their grammatical knowledge is displayed in naturalistic contexts.

SUMMARY AND CONCLUSIONS

MLU has been shown to be more closely tied to grammatical complexity under some circumstances than others. The relation of IPSyn scores to utterance length was very strong during the earliest stages of language acquisition, regardless of diagnostic classification, but tended to weaken as MLU increased. While many IPSyn scores obtained by language-disordered children and adolescents were equivalent to those of normal preschoolers with the same MLU, there were also many cases for whom MLU provided a severe overestimate of grammatical complexity. This was especially problematic when MLU exceeded 3.0 morphemes and occurred most frequently in the autistic group.

It is clear, therefore, that beyond the earliest stages of language development, MLU alone often does not provide a very accurate estimate of the child's production of specific syntactic and morphological forms as assessed by the IPSyn or the LARSP. Only when convergence between MLU and such scores is seen, we would argue, can conclusions about the grammatical production capabilities of the individual be made with confidence, and can strong claims be made as to the matching of individuals or groups for syntactic level.

APPENDIX

A listing of the 56 IPSyn items is shown here. Scarborough (1990) provided additional examples and more detailed information about scoring criteria.

Noun phrases

N1 proper, mass, or count noun: *dog, milk, Mommy.*
N2 pronoun or prolocative: *I, here, who, mine.*
N3 modifier (adjective, possessive, quantifier): *his, hot, any, that* boy, *baby's* cup.
N4 two word NP (nominal preceded by article or modifier): *my doll, more milk, the one.*
N5 article, used before a noun: *a* dolly, *the* saw, *an* apple, *a* corn, *a* orange, *the* something.
N6 two word NP (as in N4) after verb or preposition: drive *a car*, see *two cats*, is *a wrench*, It for *my Daddy*.
N7 plural suffix: *cats, pages, mans, milks.*
N8 two word NP (as in N4) before verb: *This one* goes, *My turn*'s next.
N9 three word NP (Det/Mod + Mod + N): *a bad boy, my new car.*
N10 adverb modifying adjective or nominal: *too* hot, *pretty* good, *only* me, *really* cute, what *else, right* here.
N11 other bound morpheme on adjective or noun: tall*est*, a fix*er*, *un*happy, lock*ies*, baby*ish*.
N12 other

Verb phrases

V1 verb: *fall, is, help.*
V2 particle or preposition: *up*, put *back*, this *on*.
V3 prepositional phrase: *into the truck, at school, of them.*
V4 copula: He *is* okay. We *is* monsters. *Are* you sure?
V5 catenative before verb: *haf(ta)* eat, *s'pose(da)* do, *gon(na)* be, *wan(t)(a)* go, *better* get, *got(ta)* find.
V6 auxiliary be/do/have: She *is* coming. Where *does* it go?
V7 progressive suffix: *Sleeping*. It's *turning*. We're *soldiering*.
V8 adverb: *now, tomorrow, too, hardly, already, yet.*
V9 modal: I *can* do it. *Will* you help me?
V10 third-person singular present-tense suffix: *goes, haves.*
V11 past-tense modal: *Could* I try? You *should*n't hit. It *might* fall.
V12 regular past-tense suffix: *cried, bringed.*
V13 past-tense auxiliary: He *was* talking. What *did* you say? *Did*n't you hear? I *had*n't gotten enough.
V14 medial adverb: I *just* need two. It's *still* turning.
V15 emphatic/elliptical copula, modal, or auxiliary: Yes it *can*. I *did* say please. No she *isn't*.
V16 past-tense copula: We *were* tired. It *was* fun.
V17 other (e.g., bound morpheme on verb or to create an adverb): *un*screw, brave*ly*, *re*cut.

Questions and negations

Q1 intonationally marked question: *Mine? Want some?*
Q2 *wh-* pronoun alone or routine do/go/existence/name question: *What dis? Where he going? What that called? Why? Who else?*
Q3 simple negation: *No talk. Not yours. Can't find. Don't know.*
Q4 *wh-* pronoun followed by verb: *Who made it? What she wants?*
Q5 negative morpheme between subject and verb: *I not going. You can't have* it.
Q6 *wh-* question with inversion: *Where was it? What can I do?*
Q7 negation of copula, modal, or auxiliary: It *doesn't work, Aren't* you ready? It*'ll never fit* in there.

Q8 inverted yes/no question: *Can I have it? Is she hungry?*
Q9 *Why, When, Which*, or *Whose* question-introducer.
Q10 tag question: I'll be the mom, *okay*? This is fun, *isn't it*?
Q11 other (e.g., negative inverted questions): *Why won't she play?*

Sentence structures

S1 two word sequence: *Lookit Mom. Here hammer.*
S2 subject–verb sequence: *Monkey fall. It goes* over there.
S3 verb–object sequence: *Help me*. Gonna *drink coffee*.
S4 subject–verb–object: *I need that. He likes milk.*
S5 any conjunction: *And* me. *Or* a caboose. *Then* my turn.
S6 two verb sentence: I *like play* this. *See fall* down.
S7 conjoined phrases: *red and blue, you or me, runs and jumps.*
S8 infinitive with *to*: I like *to swim*. Want *to feed* her?
S9 let/make/help/watch forms: *Let's play this. Let me look. Help him stand up. Make her be quiet.*
S10 adverbial conjunction: *'Cause. But* not this one. *If . . . then, after, unless, until, as soon as.*
S11 propositional complement: I know *you broke it*. D'ya think *she's wet*?
S12 conjoined sentences: *I'll start it and then you can finish. Do I get to keep this when you go home?*
S13 *wh-* clause: Here's *where it went*. Tell me *if it hurts*. Look at *what's under here*.
S14 bitransitive predicate: Read *me the book*. I pour *tea for you*.
S15 sentence with three or more VPs: She *told* them to *hurry* so they *ran* home. We*'re* cold and we *want* to *come* in.
S16 relative clause: Find the one *that fits here*. It's like the kind *we have in school*.
S17 complex infinitive: I need *you to help*. Want *me to fix it*?
S18 gerund: This is *for hammering. Hammering* is fun. I make noise *(by) hammering*.
S19 fronted or center-embedded subordinate clauses: The one *I like best* is the Daddy doll. *After it rains* there's puddles.
S20 other (e.g., passive constructions, tag comments, intrusions): This guy *is killed by* the dragon. He's the prince, *I imagine*. And this, *I guess*, goes on top.

ACKNOWLEDGMENTS

This research was supported, in part, by grants HD 18571 and HD 22351 (Scarborough), HD 22355 (Rescorla), HD 18833 (Tager-Flusberg), and HD 06543 (Fowler) from the National Institute for Child Health and Human Development. Fowler's work was also supported by the National Down Syndrome Society. Support for the research on the Fragile X syndrome group was received from the New York State Office of Mental Retardation and Developmental Disabilities, and from grant MH 38201 (to I. Cohen) from the National Institutes of Mental Health.

NOTES

1. When corpora of fewer than 100 child utterances were analyzed, IPSyn scores were adjusted for differences in corpus length according to the conversion table provided by Scarborough (1990).

2. Differences of 12 points or larger will henceforth be considered "severe" estimation errors. Since the standard error (*SE*) obtained in the regression analysis for the normal group was 6.1 points, such severe discrepancies are approximately 2 *SE* or more above or below expected values.
3. As discussed elsewhere (Scarborough & Dobrich, 1990), a language sample for case 3 at age 30 months was not available, so a total of 19 corpora were analyzed for the longitudinal language-delayed group.

REFERENCES

Baron-Cohen, S. (1989). Do autistic children have obsessions and compulsions? *British Journal of Clinical Psychology, 28*, 193–200.
Bayley, N. (1969). *Bayley Scales of Infant Development*. New York: Psychological Corp.
Bloom, L. (1970). *Language development: Form and function in emerging grammars*. Cambridge, MA: MIT Press.
(1973). *One word at a time*. The Hague: Mouton.
Bloom, L., Lightbown, P., & Hood, L. (1975). Structure and variation in child language. *Monographs of the Society for Research in Child Development, 40*(2).
Brown, R. (1973). *A first language: The early stages*. Cambridge, MA: Harvard University Press.
Cohen, D. J., Paul, R., & Volkmar, F. (1986). Issues in the classification of pervasive and other developmental disorders: Toward DSM-IV. *Journal of the American Academy of Child Psychiatry, 25*, 213–220.
(1987). Issues in the classification of pervasive developmental disorders and associated conditions. In D. J. Cohen & A. M. Donnellan (Eds.), *Handbook of autism and pervasive developmental disorders* (pp. 20–40). New York: Wiley.
Crystal, D. (1974). A review of Brown's *A first language*. *Journal of Child Language, 1*, 289–307.
Crystal, D., Fletcher, P., & Garman, M. (1976). *The grammatical analysis of language disability: A procedure for assessment and remediation*. New York: Elsevier.
Denckla, M. B. (1986). New diagnostic criteria for autism and related behavioral disorders: Guidelines for research protocols. *Journal of the American Academy of Child Psychiatry, 25*, 221–224.
de Villiers, J. G., & de Villiers, P. A. (1973). A cross-sectional study of the acquisition of grammatical morphemes in child speech. *Journal of Psycholinguistic Research, 2*, 267–278.
Fowler, A. E. (1985). Language acquisition in Down syndrome children: Production and comprehension (Doctoral dissertation, University of Pennsylvania, 1984). *Dissertation Abstracts International, 46*(1-B), 324.
(1988). Determinants of rate of language growth in children with Down syndrome. In L. Nadel (Ed.), *The psychobiology of Down syndrome*. Cambridge, MA: MIT/Bradford.
(1990). Language abilities in children with Down syndrome: Evidence for a specific syntactic delay. In D. Cicchetti & M. Beeghly (Eds.), *Down syndrome: The developmental perspective* (pp. 302–328). New York: Cambridge University Press.
Fowler, A., Gelman, R., & Gleitman, L. (1980, October). *A comparison of normal and retardate language equated on MLU*. Paper presented to the Boston University Conference on Child Language Development, Boston.
Hollingshead, A. B., & Redlich, F. C. (1958). *Social class and mental illness*. New York: Wiley.
Johnston, J., & Schery, T. (1976). The use of grammatical morphemes by children with communication disorders. In D. Morehead & A. Morehead (Eds.), *Normal and deficient child language* (pp. 239–258). Baltimore: University Park Press.
Johnston, J. R., & Kamhi, A. (1984). The same can be less: Syntactic and semantic aspects of the utterances of language impaired children. *Merrill-Palmer Quarterly, 30*, 65–86.
Klee, T., & Fitzgerald, M. D. (1985). The relation between grammatical development and mean length of utterance in morphemes. *Journal of Child Language, 12*, 251–269.

Klima, E., & Bellugi, U. (1966). Syntactic regularities in the speech of children. In J. Lyons & R. Wales (Eds.), *Psycholinguistic papers*. Edinburgh: Edinburgh University Press.
Kramer, C. A., James, S. L., & Saxman, J. H. (1979). A comparison of language samples elicited at home and in the clinic. *Journal of Speech and Hearing Disorders, 44*, 321–330.
MacWhinney, B., & Walter, J. (1987). CHAT manual. *Transcript Analysis, 4*, 1–52.
Miller, J. F., & Chapman, R. S. (1985). *Systematic analysis of language transcripts: User's guide*. Madison: University of Wisconsin Language Analysis Laboratory.
Morehead, D., & Ingram, D. (1973). The development of base syntax in normal and linguistically deviant children. *Journal of Speech and Hearing Research, 16*, 330–352.
Nelson, K. (1973). Structure and strategy in learning to talk. *Monographs of the Society for Research in Child Development, 38* (Serial No. 149).
Newport, E. L., Gleitman, H., & Gleitman, L. (1977). Mother I'd rather do it myself: Some effects and non-effects of maternal speech style. In C. E. Snow & C. A. Ferguson (Eds.), *Talking to children: Language input and acquisition*. Cambridge: Cambridge University Press.
Rescorla, L. (1989). The language development survey: A screening tool for the identification of language delay in toddlers. *Journal of Speech and Hearing Disorders, 54*, 587–599.
Rescorla, L., & Schwartz, E. (1990). Outcome of specific expressive language delay. *Applied Psycholinguistics, 11*, 393–407.
Reynell, J. K. (1985). *Reynell developmental language scales*. Windsor, UK: NFER-Nelson.
Rondal, J. A. (1978). Developmental sentence scoring procedure and the delay-difference question in language development of Down's syndrome children. *Mental Retardation, 16*, 169–171.
Rondal, J. A., Ghiotto, M., Bredart, S., & Bachelet, J. (1987). Age-relation, reliability and grammatical validity of measures of utterance length. *Journal of Child Language, 14*, 433–446.
Rutter, M. (1978). Diagnosis and definition. In M. Rutter & E. Schopler (Eds.), *Autism: A reappraisal of concepts and treatment* (pp. 1–25). New York: Plenum.
Scarborough, H. S. (1989). Prediction of reading disability from familial and individual differences. *Journal of Educational Psychology, 81*, 101–108.
(1990). Index of Productive Syntax. *Applied Psycholinguistics, 11*, 1–22.
Scarborough, H. S., & Dobrich, W. (1985). Illusory recovery from early language delay. *Proceedings of the Symposium on Research on Child Language Disorders, 6*, 90–99.
(1990). Development of children with early language delay. *Journal of Speech and Hearing Research, 33*, 70–83.
Shipley, E. F., Smith, C., & Gleitman, L. R. (1969). A study in the acquisition of language: Free responses to commands. *Language, 45*, 322–342.
Sparrow, S. S., Balla, D. A., & Cicchetti, D. V. (1984). *Vineland Adaptive Behavior Scales*. Circle Pines, MN: American Guidance Co.
Sudhalter, V., Cohen, I. L., Silverman, W., & Wolf-Schein, E. G. (1990). Conversational analyses of males with Fragile X, Down syndrome, and autism: A comparison of the emergence of deviant language. *American Journal on Mental Retardation, 94*, 431–441.
Tager-Flusberg, H., Calkins, S., Nolin, T., Baumberger, T., Anderson, M., & Chadwick-Dias, A. (1990). A longitudinal study of language acquisition in autistic and Down syndrome children. *Journal of Autism and Developmental Disorders, 20*, 1–21.
Terman, L. M., & Merrill, M. A. (1973). *Stanford-Binet Intelligence Scale: 1972 norms edition*. Boston: Houghton Mifflin.
Wells, G. (1985). *Language development in the preschool years*. Cambridge: Cambridge University Press.
Wiegel-Crump, C. A. (1981). The development of grammar in Down's syndrome children between the MA's of 2-0 and 6-11 years. *Education and Training of the Mentally Retarded, 6*, 24–60.
Wolf-Schein, E. G., Sudhalter, V., Cohen, I. L., Fisch, G. S., Hanson, D., Pfadt, A. G., Hagerman, R., Jenkins, E. C., & Brown, W. T. (1987, July). Speech-language and the Fragile X syndrome: Initial findings. *American Speech-Language Hearing Association*, 35–38.