# Using a corpus of school children's writing to investigate the development of vocabulary diversity

N. CHIPERE, D. MALVERN, B. RICHARDS and P. DURAN
School of Education, University of Reading, UK

**Abstract**

The paper shows how a corpus-based approach can be used to investigate the development of vocabulary diversity during the school years. The theoretical and pedagogical motivations for the investigation are outlined and advantages of using a corpus-based approach are discussed. The problem of the text-length effects on type-token ratios is presented, followed by a description of a recent mathematical solution to the problem (Richards and Malvern, 1997). Data for the investigation consisted of a corpus of 899 narrative essays from school children aged 8 to 15. The essays were grouped into three three age categories (Key Stages 1, 2 and 3) and within each Key Stage, essays were further categorised according to a maximum of eight possible levels of writing ability as defined by the British National Curriculum for English. An analysis was carried out for each Key Stage to determine the relationship between level of linguistic ability and vocabulary diversity. The paper presents results from the analysis and discusses some theoretical and pedagogical implications. Future applications of the mathematical model to the investigation of diversity in categories of linguistic form other than vocabulary are discussed.

## 1 Introduction

Computer-based corpus analysis can reveal meaningful patterns in first language development during the school years. Results are reported here from an analysis of vocabulary diversity in the writing of school children aged eight to fifteen. The analysis is part of a three-year funded research project which seeks to identify quantitative measures of the written language skills of school children. The paper begins by outlining some key theoretical and pedagogical motivations for studying children's writing. A discussion of some of the methodological issues involved is then followed by a report of the study and its results.

## 2 Theoretical and pedagogical motivations for the study

Recent policy changes in first language pedagogy in England are compelling linguists to reconsider long-held assumptions about language development. It has been assumed that procedural grammatical competence is attained without instruction by age five or soon thereafter. This assumption is congruent with previous educational practice which excluded the teaching of first language grammar. However, the recent 'National Literacy Strategy' now requires the grammar of English to be taught in English schools. This policy shift, brought about by public concerns with standards of literacy, makes an implicit but clear theoretical statement about grammatical development: that it is not complete by age five and that it can be facilitated by formal instruction.

The British government has therefore adopted a view of grammatical development which can loosely be described as empiricist. The empiricist view is that language is learned with feedback from experience (e.g. Salzinger, 1975 and Sampson, 1999). However, it is a nativist axiom that competence in language is attained without assistance very early in life (e.g. Pinker, 1994 and Marcus, 1994) and that, as far as the purely *grammatical* component of linguistic competence is concerned, the role of experience is merely to trigger innate knowledge. While empiricists might therefore regard the attempt to teach first language grammar as fully justified, nativists might see it as misguided and redundant.

The theoretical and pedagogical importance of settling the question needs no emphasis. However, it is not easy to tease empiricist and nativist claims apart on the basis of pre-school language acquisition data. In view of the evidence that grammatical development continues during the school years (e.g. Karmiloff-Smith 1986 and Perera 1986), it is necessary to extend the investigation to cover this period as well. It is also necessary to use both experimental and observational methods of data collection. For instance, one of the authors reports variations in the procedural grammatical competence of 18 year-old native English speakers and relates these variations statistically to differences in academic ability (Chipere, in press and Chipere, 1999). The literature reviews in the just cited work indicate that levels of procedural grammatical competence are also statistically related to levels of formal education. While this evidence supports the empiricist view, the necessarily tight focus of experimental studies restricts the range of subjects and materials. Corpus-based analyses can complement experimental studies by providing a wider coverage of subjects and materials.

Corpus-based analyses can also produce data which has pedagogical applications. For instance, corpus analysis might reveal linguistic features which characterise good versus poor writing among school children. It is known that skilled writers use more diverse vocabulary than less skilled writers. Corpus analysis techniques can make it possible to measure vocabulary diversity scores for pieces of children's writing and these scores could then enable teachers to decide how much time and effort a given pupil should spend in improving vocabulary knowledge. Vocabulary diversity scores might also be useful for assessment purposes. There is a growing interest in the possibility of computerised assessment of writing and vocabulary diversity scores might inform automated assessment of lexical richness.

**3 Framework of analysis**

Thus there are both theoretical and pedagogical motivations for studying language development during the school years. Some of the methodological issues surrounding such an investigation now need to be discussed. The framework of analysis which has been adopted for the current project is that of Biber (1988). This framework 'is based on the assumption that strong co-occurrence patterns of linguistic features mark underlying functional dimensions.' (Biber, 1988:12). For instance, conversational interaction represents a functional dimension of language which is characterised by a given pattern of co-occurring linguistic features. This pattern differs markedly from the pattern which chacterises, for instance, the delivery of technical information. Biber's analytical procedure therefore involves calculating the frequencies of a large number of selected linguistic features and then deriving a set of functional dimensions through factor analysis. Biber proposes that one application of his framework is in composition research. It might be the case, for instance, that good and poor writing are marked by different co-occurrence patterns. Grabe and Biber (1987, cited in Biber, 1988) found only small differences between good and poor essays in their study. However, it is possible that large differences will be found if a) the sample represents the whole range of writing ability and b) texts are analysed for the whole range of linguistic features known to mark language development.

These two considerations inform the current project. It is intended to analyse the writing produced by children who represent a wide range of writing ability. Their writing will be then be analysed in terms of quantitative measures of language development reported in the literature (e.g. Johnson, 1944; van der Geest, Gerstel, Appel and Tervoort 1973; Barnes, Gutfreund, Satterly and Wells, 1983; Fletcher and Peters, 1984; Bennett-Kastor, 1988; Klee, 1992 and Snow, 1996). This literature identifies the ratio of different types of words to the total number of words in a text, or type-token ratio (TTR) as one of the most important indicators of language development. TTR is taken as a measure of vocabulary diversity and it is usually expected that TTRs will be positively correlated with other measures of language development.

**4 The type-token ratio**

A serious flaw in the calculation of TTR, however, has been identified by several writers from as early as Chotlos (1944) to Biber (1988). Richards and Malvern (1997) provide an extensive discussion of the issues and propose a solution which will be described presently. The following paragraphs summarise the key points in that discussion.

TTR is calculated by dividing the number of different types (V) with the total number of tokens (N) in a text. Many researchers have mistakenly assumed that the ratio is constant over a given text. Richards (1987) shows that the ratio is closely related to text length with the following simple demonstration. Consider a simple case of a two-word text in which the same word occurs twice. In that case, TTR = 1 type divided by 2 tokens = 0.5. Now consider the case of a three-word text in which the same word occurs three times. Then, TTR = 1 type divided by 3 tokens = 0.3. In a four-word text with the same word occurring 4 times, TTR = 1 type divided by 4 tokens = 0.25. Finally, in a five-word text with the same word occurring five times, TTR = 1 type divided by 5 tokens = 0.2. Thus five texts which have exactly the same range of vocabulary yield five different values of TTR. Additionally, longer texts will tend to produce smaller TTRs.

Richards and Malvern show how failure to recognise this flaw has resulted in contradictory research findings in the child development literature. Examples are cases where the text length effect produces results which indicate a) no differences in TTRs taken from transcripts of children at different levels of language development; b) lower TTRs for more advanced versus less advanced children and c) a lack of correlation between TTR and other measures of language development. Several researchers have long been aware of the problem, however, and have tried to correct it. Solutions have taken the form of controls on text length or transformations of TTR. As shown below, all these solutions are either inherently flawed or subject to practical limitations which make them inappropriate for the analysis of children's writing.

**5 Controls on text length**

In the child language development literature, Stickler (1987) proposes standardising text length by using 50 utterances taken from the middle of a transcript. However, this method does not eliminate the text length effect, since more advanced children produce more words per utterance than less advanced children. Thus, not only will the measure of lexical diversity be distorted, but there is a possibility that the TTR values of more advanced children will be smaller than those of less advanced children. A demonstration of this anomaly is provided in Richards and Malvern (1997: 26).

Another solution is to standardise the number of tokens. While this solution does eliminate text length effects, there are practical problems. Standardising text length is practicable for a given corpus but it is difficult to arrive at a standard text length which can be applied to all corpora. Thus the standard text lengths have varied from 1000 tokens (Wachal and Spreen, 1973 and Hayes & Ahrens, 1988) to 400 tokens (Biber, 1988 and Klee, 1992) to 350 tokens (Hess et al, 1986) to 50 tokens (Stewig, 1994). This variation is problematic because TTRs which are calculated on the basis of shorter text lengths will be higher than those calculated from longer text lengths. It is therefore not possible to compare the two sets of TTRs. The difficulty of arriving at a universal standard cannot be solved simply by consensus because of wide variations in the lengths of transcripts from different sources. For instance, many transcripts of child language data are much shorter than those from adult language data. Standards based on the length of child language transcripts will therefore involve wasting a considerable amount of adult language data and possibly reducing the reliability of the measure. The crux of the problem is that TTR continues to fall with increasing text length and measuring TTR at any one point is inherently unsatisfactory.

The final solution based on standardising text length to be discussed here is the Mean Segmental Type Token Ratio or MSTTR (Johnson, 1944). This measure involves calculating the mean TTR for consecutive equal-length segments of text. The advantage of this method over standardising the number of tokens is that a) the size of the smallest transcript in a corpus can be used as the size of the segment and b) nearly all the data are used. However, a problem remains in that it is not possible to compare cross-corpus MSTTRs based on different-sized segments, since MSTTRs based on short segments will be higher than those based on longer segments. Thus, while MSTTR might appear to be an elegant solution, it too does not fully overcome the problem of text length.

**6 Transformations of TTR**

Attempts to transform TTR in various ways also fail to eliminate the effect of text length. Guiraud (1960) divides the number of types by the square root of the number of tokens to derive root type token ratio or RTTR. Herdan (1960) divides the logarithm of the number of types by the logarithm of the number of tokens to obtain the Bilogarithmic TTR. Carroll (1964) divides the number of types by twice the square root of the number of tokens to derive corrected type token ratio or CTTR. Ultimately, however, none of these transformations overcome the effect of text length, since "any apparent reduction of the relationship with sample size is an artefact of the change in scale and will be accompanied by a reduction in the sensitivity of the measure due to the use of smaller units" (Richards and Malvern, 1997: 33).

**7 Mathematically modelling diversity**

Most of the solutions discussed above, with the exception of MSTTR, fail to eliminate text length effects because they do not utilise the fact that diversity in a text is better represented in terms of a curve described by values of TTR taken at successive points along the length of the text rather than in terms of a single value taken at one point. A more successful class of solutions has focussed on mathematically modelling the way TTR falls with increasing token counts. A detailed discussion of the development of these models is provided in Richards and Malvern (1997). They present an equation which describes the family of curves obtained when TTR values are plotted against token counts. These curves lie between the two extremes of total diversity and zero diversity. In the case of total diversity, the number of types equals the number of tokens throughout the text and TTR = 1 at each successive point along the abscissa, resulting in a straight line with a zero slope. In the case of zero diversity, the total number of types is 1 throughout the text and TTR = 1/(N)umber of tokens for increasing values of N. The result is a curve which falls steeply from an initial value of 1 along the ordinate and then gradually flattens as it asymptotically approaches the abscissa. The TTR-Token curves of different texts will therefore lie between the two extremes with increasing lexical diversity represented by increasingly shallower slopes and decreasing diversity represented by increasingly steeper slopes. The equation found by Malvern and Richards to describe this family of curves is:

Fig. 1

$$\mathrm{TTR} = \frac{D}{N}\left[\left(1 + 2\frac{N}{D}\right)^{\frac{1}{2}} - 1\right]$$

where TTR = type-token ratio, N = number of tokens and D is a constant with serves as the index of diversity.

**8 Implementation and Validation of D**

An algorithm for computing values of D from transcripts is described in McKee, Malvern and Richards (2000). Points for a TTR-Token curve are obtained by calculating type-token ratios for increasing values of N from N = 35 to N = 50. Each point is averaged from 100 sub-samples drawn randomly from the text without replacement. D is then obtained through a curve-fitting procedure. The algorithm has been implemented in a C program called *vocd*, also described in McKee et al, which runs on UNIX, PC and Macintosh platforms as part of the CLAN suite of programs (McWhinney, 2000). D has been validated in a number of analyses on corpora containing data from first and foreign language learning and academic writing (Malvern and Richards, 2000). The current project seeks to extend the application of D to the analysis of first language development during the school years. A description of a study carried out as part of the project now follows. The study was concerned with discovering patterns in the development of vocabulary diversity in school children aged between 8 and 15.

**9 Background to the study**

There are well-established differences in the written language abilities of school-going children. These differences are found between and within age groups. Between-group differences suggest that language development continues during the school years while within-group differences suggest individual differences in language development. It is interesting to find out how such differences might be related to objective measures of language development. Children’s writing in England has been assessed by a government body called the Qualifications and Curriculum Authority. QCA mark schemes for the assessment of writing focus on three major aspects of written language: Purpose and Organisation, which is concerned with discourse level aspects; Style, which is concerned with sentence structure and vocabulary and Punctuation, which is concerned with the use of punctuation conventions and spelling. These criteria are applied by markers who give a global score for each of the three aspects. The three scores are then added up and used to classify each script in terms of eight levels of writing ability. The assessment is therefore qualitative in nature. However, QCA has recently developed a quantitative instrument consisting of a set of coding frameworks for writing.

The frameworks are used by trained markers to count the frequencies of selected features in 100-word samples of scripts. These features include correct and incorrect uses of various punctuation marks, types of spelling errors, word tokens belonging to various word classes, subordinate and co-ordinate clauses and so on. The fact that the coding is done manually limits the size of the sample per child and the variety of features which can be studied. The current project grew out of an attempt by one of the authors to automate the coding process. The study reported here measured vocabulary diversity as a first step towards automated analysis.

**10 Aims of the study**

The primary aim was to analyse the lexical diversity of school children and gauge the extent to which it is sensitive to age and ability level.

**11 Materials**
899 narrative essays at least 50 words long were analysed. The essays were obtained from various schools in England. They cover a cross section of three age groups referred to as Key Stages 1, 2 and 3 in the English education system (i.e. age groups 8, 11 and 14 years) and seven (out of a possible eight) levels of writing ability. All the students were asked to write a narrative essay beginning with the sentence ‘The gate was always locked, but on that day it was open …’.

**12 Data Preparation**
10 markers assigned a score to the scripts on the basis of National Curriculum Level descriptors (see QCA 2000 for instance). The markers were unaware of the age or ability level of the pupils. Scores for each essay were assigned separately by least two markers and later averaged to obtain the final score. In a few cases, scores from different markers diverged considerably and the final score was decided through negotiation. The final score was then used to assign each script to one of eight National Curriculum Levels of writing ability. A breakdown of the numbers of scripts in each Key Stage and Ability Level are shown on Table 1 in terms of the percentage of the total number of scripts in each category. The table needs some explanation. Data for each Key Stage is presented in a column which is further divided into two columns. The first of these columns, which should be read vertically, shows the percentage of pupils in that Key Stage who were assigned to the different levels. The second column, which should be read horizontally, shows the percentage of pupils in each Level who belonged to different Key Stages.

**Table 1 % of scripts in each Key Stage and Level**

| | **Key Stage 1** | | **Key Stage 2** | | **Key Stage 3** | | **% of all scripts** |
|---|---|---|---|---|---|---|---|
| | *% KS1* | *% Level* | *% KS2* | *% Level* | *% KS3* | *% Level* | |
| **Level 1** | 11 | 92 | 1 | 8 | 0 | 0 | 4 |
| **Level 2** | 68 | 76 | 19 | 21 | 2 | 2 | 31 |
| **Level 3** | 19 | 30 | 30 | 47 | 15 | 23 | 23 |
| **Level 4** | 2 | 3 | 36 | 57 | 25 | 40 | 22 |
| **Level 5** | 0 | 0 | 13 | 33 | 27 | 68 | 12 |
| **Level 6** | 0 | 0 | 1 | 4 | 24 | 96 | 6 |
| **Level 7** | 0 | 0 | 0 | 0 | 7 | 100 | 2 |
| **% of all scripts** | 32 | - | 43 | - | 25 | - | 100 |

The graded scripts were converted into machine-readable form by a typist. It was necessary to correct spelling errors in order to prevent spelling errors from being treated as different types and thereby inflating the token counts of poor spellers. Spelling errors were corrected using a special utility program. The procedure for correcting spelling errors was as follows.

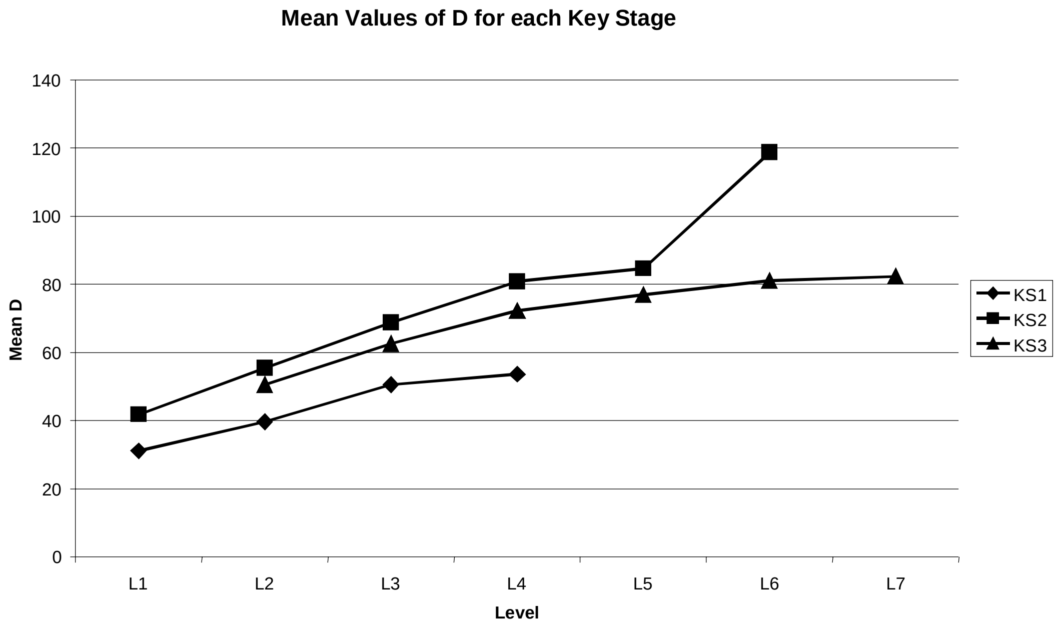
Firstly, a list of all the words in the scripts was compiled. Any words found in the list which were not also found in a dictionary list were considered as potential spelling errors by the program. It was then up to the human editor to decide if a specific word was indeed a spelling error and if so, what the correct spelling ought to be. Instances of spelling error in the corpus were then sought, found and edited through a search and replace dialogue box. This method had the advantage of speed over manual correction involving reading through the scripts. In addition, the method provided assurance that all the spelling errors flagged by the program were accounted for. However, there are at least two disadvantages with the method. Firstly, cases where correctly spelled words were used incorrectly, such as homophones, were missed. However, the margin of error thus incurred was deemed acceptable. Secondly, the scripts were altered in such a way that the original spelling could not be recovered. Refinements will be made to the utility program in future to overcome both problems.

**13 Analytical Procedure**
Essays were analysed using *vocd* (Malvern and Richards, 2000) via the CLAN interface. Memory limitations in the software meant that the scripts could only be analysed in batches of fifty at a time. After all the scripts had been processed, the output files from *vocd* were concatenated into one file and another utility program was used to extract values of D for each essay and produce a spreadsheet of all the results.

**14 Results**

D values from all 899 scripts were subjected to a 2-way ANOVA with Key Stage and Level as the independent variables. Main effects were obtained for both Key Stage $F(2) = 221.8$, $p<0.001$ and Level $F(2) = 92.965$, $p<0.001$. Mean values of D are plotted on Figure 2 and standard deviations of values D are shown on Table 2 by Key Stage, Level, by Key Stage collapsed over Level and by Level collapsed over Key Stage.

**Figure 2**




There was no significant interaction between Key Stage and Level $F(7) = 1.344$, $p<0.226$ and Level appeared to have a greater effect size than Key Stage, judging from the mean square error generated by each factor: the mean square error due to Level = 419.866 while that due to Key Stage = 454.406.

**Table 2 Standard Deviations by Key Stage and Level**

| | Key Stage 1 | Key Stage 2 | Key Stage 3 | Level stdev |
|---|---|---|---|---|
| **Level 1** | 12.376 | 8.006 | - | 12.351 |
| **Level 2** | 15.762 | 21.927 | 6.522 | 18.764 |
| **Level 3** | 16.426 | 20.833 | 15.809 | 20.448 |
| **Level 4** | 19.344 | 25.835 | 16.335 | 24.012 |
| **Level 5** | - | 23.535 | 15.693 | 19.969 |
| **Level 6** | - | 40.931 | 17.985 | 21.38 |
| **Level 7** | - | - | 14.370 | 14.37 |
| **Key Stage stdev** | 16.517 | 25.866 | 17.753 | - |

A surprising result was that Key Stage 2 consistently obtained higher mean values of D than Key Stage 3. However, the difference between Key Stage 2 and 3 was not significant. A post-hoc Tukey test shows that while there are significant differences between Key Stage 1 and Key Stage 2, $p<0.001$ and between Key Stage 1 and 3, $p<0.001$, there are no significant differences between Key Stage 2 and 3, $p>0.5$. Additionally, while there are significant differences at the $p<0.01$ level between each of Levels 1, 2, 3 and all other levels, there were no significant differences between Levels 4, 5, 6 and 7. (NB. There are only 3 data points contributing to the very high mean of D in Key Stage 2 Level 6 and this high value should not therefore be treated as a significant trend).

## 15 Discussion

The results indicate that vocabulary diversity is significantly related to age and writing ability. Of the two factors, writing ability appears to be have a greater effect size. Rather surprisingly, KS2 obtained higher scores of diversity than KS3. The reasons for this are not clear. One possibility is a dip in performance which is known to occur at Key Stage 3. This dip has been variously ascribed to the trauma of moving from primary school to secondary or school or to maturational influences. However, there might also be a linguistic explanation. It may be that older pupils produce more coherent discourse which requires a greater level of lexical repetition and therefore results in lower diversity scores. If that possibility were true, then it might indicate a limit in the sensitivity of vocabulary diversity as a measure of writing ability. However, these comments are all purely speculative. To explain the result more fully, it is necessary at least to analyse Key Stage 4 scripts and obtain a wider view of the trend.

The effects of writing ability and age do not in themselves discriminate between different theories of language development, since few would deny that vocabulary grows during the school years and that rates of growth may differ across individuals. However, these effects do raise significant questions for deeper investigation. It has been proposed that vocabulary growth during the school years can be accounted for in terms of the development of derivational morphology (Anglin, 1993). By analysing the co-occurrence of morphologically related forms in a text, a future study might determine whether increasing knowledge of derivational morphology accounts for the observed increases in vocabulary diversity. If lexical diversity is found to be an effect of morphological productivity, then the observed effects of age and writing ability would suggest that morphological knowledge grows gradually during the school years but at different rates for different individuals. That observation might have both theoretical and pedagogical implications which would need to be explored.

A second question for further research is also suggested. Inasmuch as considerable grammatical information is now stated in lexical terms, current linguistic theories posit a close relationship between the lexicon and sentence level grammar. This relationship is not a purely descriptive artefact. Bates and Goodman (1997) have shown that there is a close relationship between lexical and grammatical development in first language acquisition, aphasia and on-line sentence processing. It has also been shown, by means of both corpus and on-line sentence processing data, that syntactic structures are often intimately associated with specific lexical properties (eg. MacDonald, 1994 and Trueswell 1996). The question which arises is whether the differences in lexical diversity reported here might be related in some way to differences in sentence level grammatical development. A syntactically annotated corpus would make it possible to address the question. If such a relationship does exist, there would be theoretical and pedagogical implications to explore.

## 16 Conclusion

The paper has shown that the study of language development during the school years is a rich area for corpus-based investigation. It can help to discover objective developmental trends and it can also yield results of potential pedagogical value. It also offers methodological challenges for the corpus-based approach. It was shown how an apparently simple measure, vocabulary diversity, requires rather sophisticated modelling. The use of mathematical modelling to facilitate quantitative analysis of corpora will no doubt increase as linguistics develops further as a quantitative discipline. The question arises whether the model described here can be generalised to the study of diversity in forms of language other than vocabulary. For instance, syntactic annotation would allow the measurement of syntactic diversity while pragmatic annotation would allow the measurement of the diversity of pragmatic functions in a text. If these kinds of analysis yield useful results, the analysis of diversity could well provide a general method of measuring linguistic productivity for various pedagogical, clinical and other applications.

## References


Anglin J 1993 Vocabulary development: a morphological analysis. *Monographs for the Society for Research in Child Development.* Serial No. 238, 58(10).

Barnes S, Gutfreund M, Satterly D, Wells CG 1983 Characteristics of adult speech which predict children's language development. *Journal of Child Language* 3(10): 65-84.

Bates E, Goodman J 1997 On the inseparability of grammar and the lexicon: Evidence from acquisition, aphasia and real-time processing. In Altmann G (Ed.), Special issue on the lexicon, *Language and Cognitive Processes,* 12*(5/6):* 507-586.
Bennett-Kastor T 1988 *Analyzing children's language: methods and theories.* Oxford, Blackwell.
Biber D 1988 *Variation across speech and writing*. Cambridge, CUP.
Carroll J 1964 *Language and thought.* Englewood Cliffs, Prentice-Hall.
Chipere N (in press) Variations in Native Speaker Competence: Implications for First Language Teaching. *Language Awareness*.
Chipere N (1999) *Processing Embedding in Humans and Connectionist Models.* Unpublished PhD thesis. Cambridge University.
Chotlos J 1944 Studies in language behavior: IV. A statistical and comparative analysis of individual written language samples. *Psychological Monographs* 56, 75-111.
Fletcher P, Peters J 1984 Characterizing language impairment in children: an exploratory study. *Language Testing* 1, 33-49.
Guiraud P (1960) *Problemes et methodes de la statistique linguistique*. Dordrecht, D. Reidel.
Grabe P, Biber D 1987 Freshman student writing and the contrastive rhetoric hypothesis. *Paper presented at SLRF 7*, University of Southern Carlifornia. Cited in Biber D 1988 *Variation across speech and writing*. Cambridge, CUP.
Hayes D, Ahrens M 1988 Vocabulary simplification for children. Journal of Child Language 15(2): 395-410.
Herdan,G 1960 *Type-token mathematics: A textbook of mathematical linguistics.* The Hague, Mouton.
Hess C, Sefton K, Landry R 1986 Sample size and type-token ratios for oral language of preschool children. *Journal of Speech and Hearing Research* 32: 536-40.
Johnson W 1944 Studies in language behavior: I. A program of research. *Psychological Monographs* 56: 1-15.
Karmiloff-Smith A (1986) Some fundamental aspects of language develoment after five. In Fletcher P, Garman M (eds) *Language Acquisition*. Cambridge, CUP, pp 455-474.
Klee T 1992 Developmental and diagnostic characteristics of quantitative measures of children's language production. *Topics in Language Disorders* 12(2): 28-41.
McDonald M (1994) Probabilistic Constraints and Syntactic Ambiguity Resolution. *Language and Cognitive Processes* 9: 157-201.
McKee G, Malvern D, Richards B 2000 Measuring vocabulary diversity using dedicated software. *Literary and Linguistic Computing* 15(3): 323-338.
MacWhinney B (2000) *The CHILDES Project Vol 1: Tools for Analysing Talk – Transcription Format and Programs*. New Jersey, Lawrence Erlbaum.
Malvern D, Richards B (2000) A new method of measuring lexical diversity in texts and conversations. *TEANGA* 19: 1-12.
Marcus G (1993) Negative evidence in language acquisition. *Cognition* 46: 53-85.
Perera K (1986) Language acquisition and writing. In Fletcher P, Garman M (eds) *Language Acquisition*. Cambridge, CUP, pp 494-518.
Pinker S 1994 *The language instinct: How the mind creates language*. New York, Morrow.
QCA (2000) *English Tests: Mark Schemes*. London, HMSO.
Richards B Malvern D 1997 *Quantifying lexical diversity in the study of language development.* The University of Reading, The New Bulmershe Papers.
Richards B (1987) Type/token ratios: what do they really tell us? *Journal of Child Language* 14: 201-9.
Salzinger K (1975) Are Theories of Competence Necessary? In Aaronson D, Rieber R (eds) *Developmental Psycholinguistics and Communication Disorders. Annals of the New York Academy of Sciences* 263: pp 178-196.
Sampson (1999) *Educating Eve*. London, Continuum International.
Snow C 1996 Change in child language and child linguists. In Coleman H, Cameron L (eds) *Change and language*. Clevedon, BAAL in association with Mulitlingual Matters, pp 75-88.
Stewig J 1994 First graders talk about paintings. *Journal of Educational Research* 87(5): 309-16.
Stickler K 1987 *Guide to analysis of language transcripts*. Eau Claire, Thinking Publications.
Trueswell J 1996 The role of lexical frequency in syntactic ambiguity resolution. *Journal of Memory and Language* 35: 566-585.
van der Geest T, Gerstel R, Appel R, Tervoort B 1973 *The child's communicative competence: language capacity in three groups of children different social classes*. The Hague, Mouton.
Wachal R, Spreen O 1973 Some measures of lexical diversity in aphasic and normal language performance. *Language and Speech* (16): 169-81.