

1980

**The language of 5 1/2-year-old children from homes where
Macedonian, Vietnamese and an Aboriginal tribal language are
used as the language of the home**

A. L. McGregor

Follow this and additional works at: <https://ro.ecu.edu.au/ecuworks>



Part of the [Bilingual, Multilingual, and Multicultural Education Commons](#)

McGregor, A. (1980). *The Language of 5 1/2-year-old children from homes where Macedonian, Vietnamese and an Aboriginal tribal language are used as the language of the home*. Mount Lawley, Australia: Mount Lawley College.
This Report is posted at Research Online.
<https://ro.ecu.edu.au/ecuworks/7041>

Edith Cowan University

Copyright Warning

You may print or download ONE copy of this document for the purpose of your own research or study.

The University does not authorize you to copy, communicate or otherwise make available electronically to any other person any copyright material contained on this site.

You are reminded of the following:

- Copyright owners are entitled to take legal action against persons who infringe their copyright.
- A reproduction of material that is protected by copyright may be a copyright infringement. Where the reproduction of such material is done without attribution of authorship, with false attribution of authorship or the authorship is treated in a derogatory manner, this may be a breach of the author's moral rights contained in Part IX of the Copyright Act 1968 (Cth).
- Courts have the power to impose a wide range of civil and criminal sanctions for infringement of copyright, infringement of moral rights and other offences under the Copyright Act 1968 (Cth). Higher penalties may apply, and higher damages may be awarded, for offences and infringements involving the conversion of material into digital or electronic form.

MOUNT LAWLEY COLLEGE SECOND LANGUAGE DEVELOPMENT RESEARCH PROJECT RESEARCH REPORT NO. 1

**THE LANGUAGE OF 5½-YEAR-OLD CHILDREN
FROM HOMES WHERE MACEDONIAN, VIETNAMESE
AND AN ABORIGINAL TRIBAL LANGUAGE ARE USED
AS THE LANGUAGE OF THE HOME**

COMPILED BY A.L.McGREGOR



**MOUNT LAWLEY COLLEGE, 1980
2 BRADFORD STREET,
MOUNT LAWLEY,
WESTERN AUSTRALIA 6050,
AUSTRALIA**

**This project was supported by a grant from the Education Research and
Development Committee, Commonwealth of Australia.**

MOUNT LAWLEY COLLEGE SECOND LANGUAGE DEVELOPMENT RESEARCH PROJECT RESEARCH REPORT NO. 1

COMPILED BY A.L.McGREGOR



**MOUNT LAWLEY COLLEGE, 1980
BRADFORD STREET,
MOUNT LAWLEY,
WESTERN AUSTRALIA 6050,
AUSTRALIA**

**This project was supported by a grant from the Education Research and
Development Committee, Commonwealth of Australia.**

FOREWORD AND ACKNOWLEDGMENTS

Few areas in the educational world have been more subject to the winds of change sweeping over them during the last three decades than the fields of language acquisition, language learning and language teaching, particularly second language acquisition, learning and teaching. In their turn the behaviourists, the formal linguists and psycholinguists have had a dominating influence on the models proposed for language acquisition and learning processes, giving way, it would seem, in the present day to the sociolinguists. Language teachers have therefore been presented with a bewildering array of approaches to their task based on audio-lingualism, structuralism, situationalism, wholistic or global learning, and, more recently, the so-called functional/notional approach.

The resultant confusion in the minds and actions of practitioners - for all of these approaches are still very much alive and with us - has led to a situation in which excitement is intense, theories abound, recriminations are frequent but basic and stable facts are exceedingly hard to come by. As McLaughlin (1978, p. 208), after presenting his most helpful treatment of the literature on second language acquisition in childhood, says:-

'It is, unfortunately, easier to point to what is not known or proven than to spell out what can be said about second-language acquisition in children'.

The often anguished cry has therefore been for basic data on what is happening when a child learns and/or acquires a second language. It is perfectly possible, and some would say it is even highly likely, that there could be many different answers to that question based on the multiplicity of variables affecting the acquirers/learners, the contexts, the teachers and the languages themselves. This research is an attempt to supply basic data on these processes which can be examined from many different angles.

It finds its place in that very common context in which it is, practically, very difficult to distinguish between language acquisition and language learning. In Australia, as in many other parts of the world, migrant children and adults, and Aboriginal children and adults are both acquiring English through daily contact with native speakers, and learning it in schools and other institutions. The study presents data which is open to phonological, semantic, structural and sociological analysis with, to date, only the beginnings of a semantic and structural examination included. It is the hope of the research team that the data will be used by, and prove valuable to, other researchers and teachers in their particular areas of interest.

From the very large number of helpers who have willingly assisted in making this study possible five groups to whom the team is particularly indebted must be mentioned:-

- a) The Mount Gravatt College of Advanced Education Language Research Project has through its processes for gathering and examining language data provided the basic methodology for the present study. We are deeply indebted to Drs. N.W.M. Hart and R. Walker and other members of their research team for their open-handed generosity in making their programs, data and experience available and adding to this warm encouragement. Attention is particularly drawn to the fact that as one of the primary purposes of the study is to compare the language data gained from second language learners with that gained from the Mount Gravatt first language speakers, the

procedures used have as far as possible duplicated their procedures, the data is presented in the same or very similar form, and this report attempts to follow the same structure as their report for the corresponding age group with explanations and instructions being kept, where possible, identical. Changing circumstances have, of course, made exact duplication impossible but the research team believes that a large measure of comparability has been maintained.

- b) Financially this study has been made possible by the generous assistance of the Education Research and Development Committee (\$4453 for the year 1979). Their continuing support and encouragement in 1980 is greatly appreciated. It is the hope of the Research Team that the committee will be able to feel that their grant, though it no doubt represents only a small percentage of their total funding, is proving to be a valuable investment.
- c) The education authorities in Western Australia, ranging from the Director of Education through Superintendent of Research to other superintendents, heads of schools, teachers and clerks have given warm cooperation in the study, often in the face of not a little disruption to their schools and programmes. Amongst these it would be impossible to fail to mention:-

Mr Tinsley-Beck, then acting Superintendent of Research, W.A.
Education Department.

Mr Kevin Dinsdale, Supervisor of the Child Migrant Resources
Centre, W.A.

Mr Colin Moundsey, Superintendent, Dr T Metcalfe and Miss May O'Brien
of the Aboriginal Child Education Department, W.A.

Mr H F Creeper, Headmaster, and the teachers of Highgate Primary
School.

Mr Peter Mills, formerly Headmaster, Mr Barry Sivewright,
Headmaster, and the teachers of the Graylands Primary School.

Mr Len Heywood, Headmaster, and the teachers of the Wiluna Primary
School.

Mrs Allsworth, formerly of the Graylands Migrant Reception Centre,
W.A.

- d) The team has been continually gratified at the level of cooperation offered by the children who were selected as subjects for the study, and their parents. While the anonymity guaranteed makes the mention of names impossible we wish to record our genuine appreciation of their welcome into their homes and families and the willingness with which they supplied information and facilities for the recording of their children's language.
- e) Finally, though all members of the team are listed on the following pages a particular note of thanks must be made regarding -

Mr T Watson, Coordinator and Mr T Watt of the Computer Centre, Mount Lawley College, for their magnificent work in the re-writing and extension of the computer program used in this study and for their patient endurance of the number of changes requested throughout the year not to mention number of hours of computer time demanded by the project. Without them there would be no study.

Mr Neville Green of Intercultural Studies, Mount Lawley College, for his arrangements and supervision of the trip to Wiluna to gather the Aboriginal language data; his tireless work made that part of the project not only possible but enjoyable and profitable.

Particular thanks are due (along with regrets for damage to the eyes!) to our secretary Mrs Paula Pettit, to Mr Graham Curtis of Printing Department, Mount Lawley College, and to Mr John McQuillan for the purchase, maintenance and careful control of the equipment used in the project

Our sincere hope is that through the continuation of such cooperation a real contribution may be made to our knowledge about the processes of second language acquisition and learning.

A L McGREGOR

CHIEF INVESTIGATOR

MOUNT LAWLEY COLLEGE

SECOND LANGUAGE DEVELOPMENT RESEARCH PROJECT

THE RESEARCH TEAM (1979)

CHIEF INVESTIGATOR

Dr A L McGregor

CO-INVESTIGATORS

Dr I G Malcolm

Mr J L Sherwood

LECTURERS

Mr N Green (Intercultural Studies)

Mr B Hird (English)

Mr S Jongeling (Psychology & Education)

Mr J Rainford (Psychology & Education)

Dr E Vaszolyi (Intercultural Studies)

Dr R Williams (English)

EQUIPMENT MAINTENANCE, ETC.

Mr J McQuillan

SECRETARY

Mrs P Pettit

COMPUTER PROGRAMMING AND ENTRY

Mr T Watson

Mr T Watt

PRINTING

Mr G Curtis

STUDENT RESEARCH ASSISTANTS

Ian Atkinson

Susan Bowles

Pam Dettman

Judith Gauntlett

Lenore Hanley

Rose Kakulas

John Latham

Joyce McCarthy

Julianne Molloy

James Peterkin

Monica Reid

Denise Shaw

Mary Tarrant

Marlene Tonkes

Pat Attrill

Helena Coleman

Yvonne Eikelboom

Konstantin Grimm

Sheryle Innes

Erika Kalaitzis

Teresa Leahy

Sue Marville

Tim O'Keefe

Anne Pether

Carmela Rossi

Sheryle Stevens

Judith Thomas

Gerry Walsh

Barbara Zahari-Noh

Vera Apai

Liana Da Re

Linda Galton-Fenzi

Allison Horgan

Val Johnson

Joan Knowles

Frances Letizia

John Miller

Dorothy Parin

Eleanor Pree

Maria Robinson

Sheila Strauch

Pat Thompson

Peter Westall

CONTENTS

<u>FOREWORD AND ACKNOWLEDGMENTS</u>		(i)
<u>THE RESEARCH TEAM (1979)</u>		(iv)
<u>CONTENTS</u>		(v)
<u>LIST OF TABLES</u>		(vi)
<u>CHAPTER 1</u>	<u>INTRODUCTION</u>	1
<u>CHAPTER 2</u>	<u>OUTLINE OF PROCEDURES</u>	8
	(1) <i>INITIAL OBJECTIVES</i>	8
	(2) <i>PILOT PROJECT</i>	8
	(3) <i>SELECTION OF CHILDREN FOR THE STUDY</i>	9
	(4) <i>RECORDING OF LANGUAGE</i>	10
	(5) <i>TRANSCRIPTION AND CODING OF DATA</i>	11
	(6) <i>COMPUTER PROCESSING OF LANGUAGE</i>	11
	(a) <i>Capacity of Program</i>	
	(b) <i>The computer print out</i>	
	(7) <i>BACKGROUND DATA ON SUBJECTS</i>	15
	(8) <i>LIMITATIONS OF STUDY</i>	15
<u>CHAPTER 3</u>	<u>LANGUAGE USAGE OF FIVE AND A HALF YEAR OLD MIGRANT AND ABORIGINAL CHILDREN FOR WHOM ENGLISH IS A SECOND LANGUAGE</u>	17
	(1) <i>SINGLE WORD FREQUENCIES</i>	17
	(a) <i>Calculation of Index</i>	
	(b) <i>Most Frequent Single Words</i>	
	2. <i>LANGUAGE SEQUENCES</i>	20
	(a) <i>Selection of Sequences</i>	
	(b) <i>Calculation of Indices</i>	
	(c) <i>Most frequent two and three morpheme sequences</i>	
	(d) <i>Brief discussion</i>	

REFERENCES

28

APPENDICES

<u>APPENDIX 1</u>	<i>EQUIPMENT USED FOR RECORDING LANGUAGE</i>	30
<u>APPENDIX 2</u>	<i>GUIDELINES FOR RECORDING PRAGMATIC INFORMATION (HART AND WALKER, 1976)</i>	31
<u>APPENDIX 3</u>	<i>EXAMPLE OF PAGE OF TRANSCRIPTION</i>	34
<u>APPENDIX 4</u>	<i>INSTRUCTIONS FOR CODING</i>	35
<u>APPENDIX 5</u>	<i>MOUNT LAWLEY COLLEGE SECOND LANGUAGE CONCORDANCE PROGRAM: PROGRAM DESCRIPTION</i>	42
<u>APPENDIX 6</u>	<i>SAMPLE PAGE FROM A LANGUAGE CONCORDANCE PRINT OUT FOR FIVE AND A HALF YEAR OLD CHILDREN (MACEDONIAN SAMPLE)</i>	49
<u>APPENDIX 7</u>	<i>INTERVIEW SCHEDULE FOR BACKGROUND INFORMATION (MIGRANT SAMPLE)</i>	50
<u>APPENDIX 8</u>	<i>INTERVIEW SCHEDULE FOR BACKGROUND INFORMATION (ABORIGINAL SAMPLE)</i>	52
<u>APPENDIX 9</u>	<i>ONE, TWO AND THREE WORD SEQUENCES OF THE TOTAL SAMPLE OF UPPER FIVE YEAR OLD ABORIGINAL AND MIGRANT CHILDREN IN ORDER OF INDEX VALUE</i>	56
<u>APPENDIX 10</u>	<i>ONE, TWO AND THREE WORD SEQUENCES OF THE ABORIGINAL SAMPLE OF UPPER FIVE YEAR OLD CHILDREN IN ORDER OF INDEX VALUE</i>	101
<u>APPENDIX 11</u>	<i>ONE, TWO AND THREE WORD SEQUENCES OF THE MACEDONIAN SAMPLE OF UPPER FIVE YEAR OLD CHILDREN IN ORDER OF INDEX VALUE</i>	113
<u>APPENDIX 12</u>	<i>ONE, TWO AND THREE WORD SEQUENCES OF THE VIETNAMESE SAMPLE OF UPPER FIVE YEAR OLD CHILDREN IN ORDER OF INDEX VALUE</i>	136

LIST OF TABLES

<u>TABLE I</u>	<i>ORDER IN WHICH SELECTED STRUCTURES OF ENGLISH ARE ACQUIRED BY YOUNG SPEAKERS OF ENGLISH VERSUS CANTONESE, SPANISH, TAGALOG, AND ILOKANO (MACE-MATLUCK, 1979)</i>	CHAPTER 1	2
<u>TABLE II</u>	<i>LANGUAGE DATA HELD AT MT. GRAVATT RESEARCH AND LEARNING CENTRE (HART AND WALKER, 1976)</i>	CHAPTER 1	7
<u>TABLE III</u>	<i>A COMPARISON OF THE TWENTY SINGLE WORDS MOST FREQUENTLY USED BY THE MT. LAWLEY AGE 5½ MIGRANT AND ABORIGINAL SAMPLE, AND THE MT. GRAVATT (FIRST LANGUAGE) AGE 5½ SAMPLE</i>	CHAPTER 3	18
<u>TABLE IV</u>	<i>A COMPARISON OF THE TWENTY TWO-MORPHEME SEQUENCES MOST FREQUENTLY USED BY THE MT. LAWLEY AGE 5½ MIGRANT AND ABORIGINAL SAMPLE, AND THE MT. GRAVATT (FIRST LANGUAGE) AGE 5½ SAMPLE</i>	CHAPTER 3	22
<u>TABLE V</u>	<i>A COMPARISON OF THE TWENTY THREE-MORPHEME SEQUENCES MOST FREQUENTLY USED BY THE MT. LAWLEY AGE 5½ MIGRANT AND ABORIGINAL SAMPLE, AND THE MT. GRAVATT (FIRST LANGUAGE) AGE 5½ SAMPLE</i>	CHAPTER 3	24

CHAPTER 1

INTRODUCTION

Throughout the world migration patterns and changing attitudes towards education and other cultures have led over the past three or four decades to a remarkable growth in the learning of second or foreign languages. Within this area the learning of English by speakers of other languages takes up the largest numbers and is to be found in every part of the globe.

Nevertheless to date comparatively little is known of processes and patterns in second language development. Evidence on phonological and semantic development is exceedingly skimpy. Hernandez (in Ervin-Tripp 1970), Malmberg (1945), Wode (1976) and Ervin-Tripp (1974) have examined the order of the development of sounds and differences between older and younger, and literate/non literate children in sound-system processing. This work does not so far seem to have been carried much further. Ervin-Tripp (1974) is also one of the few to have worked on the semantic aspect, noting which words appear to be learned first and differences between age groups in semantic processing in the second language.

More attention has been paid to syntactic processing with a fair number of studies having been carried out with individual children, in general supportive of the hypothesis that the strategies involved in second language learning are similar to those used in first language acquisition. This position is supported by Ravem (1974) and Milon (1974) though Politzer (1974) has expressed some doubts. Cancino, Rosansky and Schumann (1974 1975) and Hakuta (1975) have to some extent supported Politzer by confirming considerable individual variation in the way learners acquire structures in the new language. These could be seen as relatively recent studies in a series that began near the beginning of this century. These earlier studies and quite a few recent ones, were carried out with individuals or very small numbers of subjects. Only within the last decade have researchers begun to work with larger samples of children. Thus Price (1968) studied the development of Welsh word-order in 21 English speaking children. An exceedingly influential series of studies was conducted by Dulay and Burt (1973, 1974a, 1974b) comparing and contrasting the order of acquisition of such functors as verb inflections, articles, possessives, prepositions such as 'in' and 'on', etc. They were able to show that the order of acquisition of these functors by Spanish-speaking children learning English was different from their order in first language acquisition yet quite consistent within the sample of Spanish speaking children even though they came from different areas, had different lengths of exposure to English and were of different ages. A later study which incorporated Chinese-speaking children learning English appeared to show that the order of acquisition of these functors was strikingly similar to other second language learners even for children from different language backgrounds (Dulay and Burt, 1974b).

This finding has been largely confirmed in the study by Mace-Matluck (1978) which showed with over four hundred students from Cantonese, Spanish, Tagalog and Ilokana speaking backgrounds that the order of mastery of functors was very similar, whatever the backgrounds, adding the interesting suggestion that this acquisition by second language learners occurs not in a linear fashion as in first language learning but rather in a global manner. Thus it is suggested that learners gain some control of most morphemes in the early stages of learning and gradually refine their mastery of them.

This significant study serves as an interesting example of the methodology of such studies and is summarised* by Mace-Matluck (1979) in the following way:-

"Four hundred forty-two students, kindergarten through grade four, from two school districts in the state of Washington were involved in the study. All the students were tested bilingually with the MAI-SIA-CAL Oral Proficiency Tests (Matluck and

*Warm thanks are extended to Dr. Mace-Matluck for permission to quote this succinct summary of her research study.

Mace-Matluck 1974). From the language data obtained from the English edition of these tests, 26 test items containing 10 grammatical morphemes for which L1 acquisition order had been established previously (Brown 1973) were isolated for study.

The degree of control of each of the morphemes, based on the percentage of correct responses for the morpheme, was determined for each child. Mean scores for the total group of children and for various subgroups of children were calculated for each morpheme. The 10 morphemes were then rank ordered, from highest to lowest on the basis of the mean scores for each group of children as defined by the variable categories of home language, grade, and English-language proficiency. The value of the Spearman rho rank difference correlation coefficient was computed for two sets of ranks:-

- 1) with first-language acquisition order and;
- 2) with each home language. Where significant correlations were found, an analysis was made of the nature of the similarities and differences between the two sets of ranks.

TABLE 1

Order in which selected structures of English are acquired by young speakers of English versus Cantonese, Spanish, Tagalog, and Ilokano

English as L2, rank order					
English as L1, rank order*	Composite group	Cantonese speakers	Spanish speakers	Tagalog speakers	Ilokano speakers
Present progressive -ing 1	1	1	1	1	2
in, on 2	4	5	4	5	5
Plural 3	7	7	5	7	6.5
Past - irregular 4	3	4	3	3	6.5
Possessive 5	5	3	6	4	4
Articles 6	6	6	8	6	3
Past - regular 7	9	10	9	10	9
Third person - regular 8	8	8	7	8	8
Third person - irregular 9	10	9	10	9	10
Contractible copula 10	2	2	2	2	1
P-value**	.45	.38	.49	.39	.25

*Brown 1973

**Spearman rho correlation coefficient

First versus second language

Results of the analyses that examined the relationship between first-language (L1) and second-language (L2) rank orders revealed that the order in which the morphemes were learned when English was L1 did not correlate significantly with their order when English was L2 (see Table). This held true for all four home

language groups and for the composite group. Similarly, correlations were not consistently significant on the basis of either the children's age, as indicated by grade level, or their English proficiency.

Studies of first-language acquisition have shown that the order in which English morphemes are acquired by monolingual English-speaking children is largely invariant, and that the point of emergence and control of these morphemes is highly predictable on the basis of the length of the child's utterances, which correlates positively with chronological age. Brown (1973) has proposed that the primary determinants of first language acquisition are semantic complexity and grammatical complexity, and that other factors, such as frequency of occurrence in the language, play only minor roles. He has been able to demonstrate that the L1 order of acquisition of 14 morphemes can be predicted on the basis of semantic and grammatical orderings.

Dulay and Burt (1974a), using Brown's orderings, attempted to predict the L2 order of acquisition of 11 English morphemes by their combined Spanish and Chinese sample. The predictions were not borne out. For second-language learners, who are older and further along in their cognitive development and who have already acquired highly complex structures in their native languages, factors other than linguistic complexity apparently become major determinants of the L2 order. Results of this present study suggest that this is so.

Similarly, the stage the children were in in the acquisition process did not appear to influence the ordering of the morphemes. Children at the lowest proficiency level showed some use of all of the morphemes except 3rd person irregular. Use of two of the morphemes (present progressive-ing and copula) had reached criterion for mastery. Children at each succeeding proficiency level showed steady progress in correct use of the morphemes, with children at the highest proficiency level demonstrating mastery or near mastery of all but four of the morphemes.

These findings suggest that possibly the acquisition of English morphemes by second-language learners occurs, not in a linear fashion as it does in first-language learning, but rather in a global manner. That is to say, learners gain some control of all or most of the morphemes in the early stages of learning and gradually refine their use of them as they gain in knowledge of how the language works.

Relative similarity of the home language structure to that of English appeared to have affected only minimally the order in which the children in the study acquired English morphemes. Rank orders obtained for the children who speak non-Indo-European languages (Cantonese, Tagalog, and Ilokano) showed lower correspondence with the L1 sequence than did those for the native Spanish speakers, for whom moderate but nonsignificant relationships between L1 and L2 were evident.

Commonality among L2 rank orders

That there is commonality in the L2 acquisition of English morphemes is evident in the high degree of correspondence that exists among the L2 rank orders of the 10 morphemes for children in this study who speak highly dissimilar mother tongues. All L2 rank orders obtained on the basis of home language background correlated significantly. The value of the correlation coefficients ranged from .71 to .99.

A number of second-language acquisition studies (Bailey and others, 1974, Dulay and Burt 1974, Fathman 1975, Larsen-Freeman 1975, Rosansky 1976, Anderson, in press) have suggested that there is a common order in which L2 learners in an English-speaking environment acquire English morphemes. This is an important finding in and of itself, but if a common order does exist, then an understanding of the factors which influence this order may well have implications for second-language teaching.

Post-hoc analyses of the data from the present study (Mace-Matluck 1978) suggest that:-

- 1) frequency of occurrence of the morphemes in adult native speaker speech is a major determinant in the L2 rank order;
- 2) syntactic category and morpheme type can account for the systematicity in the degree of accuracy with which L2 learners use English morphemes;
- 3) native language influence may explain the variance that occurs across L2 rank orders in the rate, and thus in the order, in which certain of the morphemes reach mastery criterion.

Tentative conclusions

Results of this investigation to date lead to the following tentative conclusions:-

- 1) Young second-language learners acquire English grammatical morphemes in a different order than do children learning English as their first language. Factors that are major determinants in the L1 order of acquisition appear not to be the primary factors that determine the L2 sequence.
- 2) In an English-speaking environment, the order of acquisition of English structures by second-language learners will reflect a high degree of commonality. The structure of the native language affects the order of acquisition, but possibly to a lesser degree than do other factors."

While admitting that the acquisition of these functors may, as Dulay and Burt suggested, be especially reflexive of differences in cognitive abilities between first and second language learners McLaughlin (1978) to some extent criticises the methodology employed in the following way:-

"Dulay and Burt lumped together children with different amounts of exposure to the second language and measured acquisition sequence by examining the degree to which the functors were correctly supplied in speech samples. Their study was not a longitudinal study, as were most of the studies discussed earlier. They did not, strictly speaking, measure acquisition sequence but rather accuracy of use, since they measured the percent of times a subject correctly supplied a morpheme in an obligatory context. Longitudinal studies of individual children yield more direct information about the developmental sequence in the acquisition of functors and are more directly comparable to Brown's (1973) original (longitudinal) data."

In order to answer such criticisms it appears necessary to extend previous studies in two important directions. In the first place longitudinal studies will enable us to assess whether in fact there is any similarity between order of acquisition of functors and order of accuracy in their use such as has, in fact, been established by Dulay and Burt and Mace-Matluck. The second direction should be towards the collection of naturalistic data. Until this is done with larger numbers of children than the earlier studies with individual subjects we are largely confined to studying whether the acquisition of functors is in the same order as that in earlier first language acquisition studies. Only naturalistic data will allow for the emergence (should they exist) of altogether different morphemes which may be completely missing from first language acquisition lists, at least in early stages of acquisition. It is perfectly conceivable, for example, that second language learners gain an early ability to use such structures as the imperative which of course does not appear at all in the functors under examination. This therefore points to the essential limitation of elicited language in which only

certain functors are being looked for.

Naturally it is not difficult to understand the problems facing those who might wish to work with naturalistic language. The volume of language to be processed if the sample is to be of adequate size and the possible complexity of structures to be analysed makes this a daunting task. Nevertheless until at least a start has been made in this direction we can have little confidence in our understanding of the process of second language acquisition.

Research programs developed in Queensland under the Van Leer Foundation Project and the Mount Gravatt Language Research Project (Hart, Walker and Gray, 1977) have yielded practicable approaches for the gathering and analysis of speech data from large samples of children of various age groups. On the basis of this research it has been possible to study patterns of language development in first language acquisition and to develop appropriate literacy materials for children at various levels of primary education.

It quickly became obvious to interested members of staff at the Mount Lawley College of Advanced Education Language Centre that the Mt Gravatt procedures would provide a most useful basis on which to approach the gathering and analysis of naturalistic data from those learning English as a second language.

One of the most attractive features of these procedures was the possibility of progressing well beyond the study of individual morphemes to the question of how the words are used and how they combine together. This is well summarised by Hart (1973) as follows:-

"Researchers who have worked with frequency counts (Dolch 1942, Fry 1960) tell us that structure or function words combine with other words more frequently than adjectives do and adjectives combine more frequently than nouns.

This function of combinability is one of the fundamental characteristics of any item in a language, for it is the capacity of an item to combine with other items that largely determines its frequency and probability of occurrence. Words that occur over and over again are those that combine most readily with other words. These frequently occurring structure words are the words which signal the structure of the language. Words like *a* and *the*, *can* and *do* and *in* and *on* can combine with thousands of others. These words in their proper places make up the framework or skeleton upon which the rest of the language is built. Lefevre (1964) gives a list of common types of structure words with a few illustrative examples -

1. Noun Markers - *a, the, some, any, three, this, my, few ...*
2. Verb Markers - *am, are, is, was, have, had, has ...*
3. Phrase Markers - *up, down, in, out, above, below ...*
4. Clause Markers - *if, because, that, how, when, why ...*
5. Question Markers - *who, why, how, when, what, where ...*

Combinability is not limited to structure words however. Verbs like *put*, *take*, *give* and *get* can combine with a relatively large number of different nouns, adverbs and prepositions. A limited number of adjectives like *big* and *little* can combine with a large number of nouns. Pronouns also occur with very high frequency, especially first person *I*.

In studying children's language, it is possible to extend the idea of a frequency count for single words to a frequency count for sequences of words. Suppose a frequently used word were represented by algebraical *A*, then it is possible to find the most frequent word that follows *A*. Suppose this were represented by algebraical *B*. In a like manner it is possible to find the most frequent word following *A + B* - call this *C*. Eventually we end up with a sequence which represents a very frequently used "unit of language", e.g. *I'm going to, I want to.*"

A further obvious advantage of using the procedures devised in the earlier Van Leer and Mt. Gravatt studies was the possibility of carrying out direct comparisons between the language of L2 learners and that of L1 children. In the latter case language data has been collected from successive yearly cohorts of children from ages $2\frac{1}{2}$ to $10\frac{1}{2}$. Since basic procedures were similar to those adopted in the present study and outlined in the following chapter they will not be expanded on at this point. Table II outlines the L1 and Aboriginal L2 subjects or second dialect material held at the Mt. Gravatt Research and Learning Centre (Hart and Walker, 1976). (see next page) ...

TABLE II

LANGUAGE DATA HELD AT MT. GRAVATT RESEARCH AND LEARNING CENTRE

LANGUAGE SAMPLE	Form in which data is held					Data supplied by:*
	Tape Speech only	Tape Context Desc.	Folio Trans- cribed	Phase 1 Analysis (Lexico- grammatical)	Phase 2 Analysis (Semantic- Lexico- grammatical)	
Oral Language (Brisbane Children)						Dept. Education (Q) Van Leer Project Mt. Gravatt Research and Learning Centre
2½ yrs.	X		X	X		
3½ yrs.	X		X	X		
4½ yrs.	X		X	X		
5½ yrs.	X		X	X		
6½ yrs.	X		X	X	X	
8½ yrs.	X	X	X	X	X	
10½ yrs.	X	X	X	X		
Written Language (Brisbane Children)						
8½ yrs.			X	X		
10½ yrs.			X	X		
Close Contact Aboriginal Children (Petrie Tce.)						
5½ yrs.	X	X	X	X	X	
Medium Contact Aboriginal Children (Cherbourg)						
3½ yrs.			X	X		Dept. Education (Q) Van Leer Project
4½ yrs.			X	X		

CHAPTER 2

OUTLINE OF PROCEDURES

1. INITIAL OBJECTIVES

This project was undertaken to identify and tabulate the meaningful oral language units of Aboriginal and migrant children from homes in which normal communication is in a language other than English. Initially it was hoped to include the language of Aboriginal children from homes where a dialect other than standard English is used but difficulties with the collection of the required data and the number of research assistants involved plus the knowledge that closely related work with Aboriginal children is being undertaken in Queensland led to a withdrawal from this particular group.

The intention is that such data, made available to other researchers, curriculum planners, materials developers and teachers, should further research, curriculum, and materials development for second language learners. It is seen, in the first instance, as the provision of basic research data allowing for later studies at phonological, semantic, as well as more sophisticated syntactic and functional levels of analysis.

and discourse

2. PILOT PROJECT

Plainly such a large-scale project involving a commitment to several years of work by a team of researchers required preliminary investigation and trial. In the period April 1977 to January 1978 after permission had been obtained from the Western Australian Education Department, three to four hours of the language of each of four migrant children and one Aboriginal child were recorded from different parts of a normal day. The children were between 5½ and 6½ years of age and, with exception of the Aboriginal child, were from non English speaking homes.

The language was recorded, transcribed and coded for computer analysis by 10 third year Diploma of Teaching students working in pairs under the supervision of the research team. The language of the migrant students was computer analysed separately from that of the Aboriginal child.

The purpose of this feasibility study was to establish the suitability of the procedures for the proposed research, particularly with regard to gaining the co-operation of the subjects and their parents, using Diploma of Teaching students as research assistants, establishing the usefulness of the recording equipment, demonstrating the applicability of the computer concordance program to second language data, ascertaining the accuracy of estimated costs and providing research team members with experience in the procedures to be adopted. (For a fuller description of the pilot study see McGregor, 1978).

In addition, while no firm conclusions could be drawn from the language analysis in view of the smallness of the sample it was hoped that some pointers could be obtained as to possible differences between the sample and first language results obtained in previous studies.

The results of testing equipment and procedures, giving experience to the research team, using Diploma of Teaching students as research assistants and using the computer program to analyse the language data obtained all, broadly speaking, proved highly satisfactory.

Analysis of the language data seemed to suggest the possibility that while the use of single words by second language children might not be too different from their use by first language children two and three word sequences differed significantly not only from first language use but also within the sample between the migrant children and the Aboriginal child. It also seemed possible that the migrant children might have a much smaller range of two or three word sequences available for use than is the case with first language children. These could, of course, in view of the small sample, only be regarded as highly tentative findings requiring investigation in the full study.

3. SELECTION OF CHILDREN FOR THE STUDY

The initial aim was to obtain the language of Aboriginal and Migrant children in the upper five year old age range, that is those normally to be found in the first year of schooling in Western Australian schools.

There were several compelling reasons for starting with this particular cohort. They may be summarised as follows:-

- a) In selecting a starting point for the study it was desirable to enter at a level that would enable some realistic assessment to be made of future desirable directions. Should it become obvious that the language of 5½ year old L2 learners varies significantly from L1 learners it would be desirable to extend the study in future years to discover whether there are continuing differences in later years or whether there is a convergence of the samples. Again should it be clear that children at this level (or some of them) are already well advanced in L2 acquisition it might be desirable to extend the study to earlier years, i.e. collect data from the 2½, 3½ and 4½ age groups.
- b) Practicability of administrative arrangements favoured a start at school age. Contact with children and parents would be much more easy through education authorities with whom they were directly acquainted than through direct approaches to the homes. Such contact would facilitate later permission to work in subsequent years with younger children in the homes should this prove desirable.
- c) An influx during the research year of Vietnamese refugees gave an opportunity for the study of children gaining their first contact with English in the first year of schooling, thus ensuring that at least one section of the sample would be at the beginners' stage in the L2 learning process.

Thus children were selected from the 5½ year level from three populations:-

- (1) Migrant children, speakers of Macedonian as mother tongue; (8 subjects completed).
- (2) Migrant children, speakers of Vietnamese as mother tongue; (10 subjects completed).
- (3) Aboriginal children, speakers of Pitjantjatjara (in the main) or another Aboriginal language as mother tongue; (8 subjects completed).

This selection was based on the desirability of having samples from widely differing first language backgrounds allowing for comparison between these groups as well as comparisons of the whole sample with L1 data. Practically speaking it was also possible to deal with these groups in particular geographic areas as, for example, the Vietnamese children on arrival in Western Australia were grouped in one reception centre and for the most part attended one school. Similarly the Aboriginal children from one language background were, of course, to be found (in the main) in one tribal area.

No initial screening of the children was carried out other than ascertaining that English was not used in the homes from which they came and that they were of the required age level and language background.

4. RECORDING OF LANGUAGE

The research team consisted of - the Chief Investigator, two Co-Investigators and six senior staff members (originally seven but one had to withdraw through illness). These came from the English, Intercultural Studies, and Education and Psychology departments of the College. These had the help of forty-three second and final year and postgraduate students of Mount Lawley College who acted as research assistants in collection and preparation of the language data. Details of the full research team are provided on page iv of this Report.

The aims of the project and requirements for the recording of the language sample were explained at a meeting held at each school with the parents, the children's teachers and members of the research team involved, usually with the help of interpreters. The research team then spent some time in the child's home and at school testing equipment, establishing rapport with the child and family and becoming familiar with the child's daily routine.

Children's language was recorded by means of a radio microphone placed in the pocket of a T-shirt or specially made vest worn by the child whose language was being recorded. Since in all cases language was also being collected from at least one other member of the class in which each subject was found this enabled the microphones to be carried by two or more class members for a few days before and during the collection of data, so ensuring that the subjects and other class members could become familiar with the equipment being used and go about their normal routine quite naturally. The radio microphone enabled the child to move about freely and unhindered language sampling was thus made easier. Typically these preparatory/practice recording sessions in the classroom were spread over two or three weeks.

The signal from the radio microphone was received by a portable radio receiver which fed into a mixer-preamplifier and finally into a two-track tape recorder.

The second track of the tape was, simultaneously, having fed onto it a commentary by a second team member who was observing the movements and interaction of the child. This was recorded either through a second radio microphone or, more commonly, by direct line to the tape recorder. (Technical details of the equipment used are given in Appendix 1.) The responsibility of the other research assistant present was to monitor the incoming child and commentator language through earphones attached to the recorder, checking levels and taking notes of as much speech as possible. Notes were taken of any occurrences that could aid later transcription or of any interference that could not be explained from the tapes themselves.

It was considered that details contained in the team member's commentary on the pragmatics of the situation would be of great value in the transcription and later analysis of the data. Instructions on the type of commentary required as issued to the research assistants are contained in Appendix 2 and for obvious purposes of comparison of data, are identical with the 'pragmatics' instructions issued in the Mount Gravatt studies.

In the final collection of three hours' language from each child for transcription and coding each child was assigned to one of three periods during the day by means of a random, rotating design. The first period was the time before a child went to school in the morning and typically included at least one hour of the school morning. Here the child interacted with parents, siblings, classmates and teachers. The second period was considered the latter part of the school day and included the lunch period. In this time slot the child inter-

acted with teachers and classmates as well as wider school groups during the lunch and playtime intervals. The third period was after school when the child interacted with siblings, friends and parents. Since the language of 26 children was recorded in the study this gave a total of approximately 7 morning, 9 day and 10 evening periods altogether.

In the collection of language from the children every attempt was made to sample natural interactions. While situations in which the child read from a book or blackboard, recited poems or sang songs he had not made up were not excluded as they represented valid practice by the children, every effort was made to obtain recordings of normal interactions by the two research assistants present. 'Normal' in this case is taken to mean the avoidance of language elicited by any member of the research team. It is to be noted that, in the case of the migrant subjects particularly, the circumstances of their lives outside of school hours led to an almost entire absence of spoken English. In such cases team members occasionally arranged situations to facilitate the collection of language. Typically this would take the form of having one of their own children play with the subject to ascertain what English the subject could, in fact, use. No direction of any sort was given to the children beyond, for example, the provision of toys to play with or books to look at.

5. TRANSCRIPTION AND CODING OF DATA

As soon as was practicable after the child's language was recorded the tapes were transcribed by the same research assistants who had recorded them. This transcription was carried out on a standard form in which the language recorded (both the child's and those with whom the child was interacting) was transcribed on every alternate line, the intervening lines being used for the transcription, in a different colour of ink or type, of the pragmatic commentary on the second track of the recorded tape, together with any relevant written comments made by the researchers at the time of recording. (A typical page of transcription is reproduced, without, of course, the different colour of ink, in Appendix 3. The narrow left and right hand columns of the transcription sheets were used, respectively, for notes on the macro-situation (e.g. time, place of recording and participants) and cross referencing to the coded version of the transcription.)

The conventions for the coding of the data into a text acceptable to the computer were based on those used in the Mount Gravatt Project but were extensively revised to suit the needs of this project. Care was, however, taken that the coding should lead to the provision of data in the results that would be precisely parallel to the Mount Gravatt results, all revisions being to provide additional data deemed to be of particular interest in the study of second language development, such as verbs and nouns and mistakes in these, omissions, etc. Instructions issued to research assistants regarding coding are attached in Appendix 4. It is to be noted that original intentions were to mark the coding for finer computer analysis by indicating precise tenses of verbs, cases of nouns, etc. The complexity of the resultant coding scheme and compassion for the research assistants dictated a modification of these plans leaving fine analysis to be carried out from, for example, the listing of verb mistakes provided by the program.

6. COMPUTER PROCESSING OF LANGUAGE

(a) Capacity of Program

For the project a new computer program was written, an activity dictated by the use of a different computer from that used in the Mount Gravatt studies, the College computer being a PDP 11/34 running RSTS/E, and the desire to include functions not included in the Mount Gravatt results. It was however based upon the Mount Gravatt program.

This in turn was based on one written by Dr J E Grimes at the University of Oklahoma for the analysis of Aboriginal languages. This program was subsequently adapted for use in New Guinea by Professor Max Mackay at the University of Papua and New Guinea. In its turn this early program was used by Dr Margaret Sharpe for her Ph D thesis and later by the Queensland Department of Education's Van Leer Project to analyse children's speech. This was then revised for use in the Mount Gravatt study.

The new Mount Lawley program has a greatly increased capacity, eg. the 1979 project saw the handling of 9193 lines of data or 99395 morphemes as opposed to 5,000 lines or 43,709 morphemes in the Mt. Gravatt Program. The three major modifications are:-

- (1) the listing of all correct and incorrect nouns and verbs in separate listings at the end of print out;
- (2) the marking and listing of all other forms of errors, including omissions, within the text and at the end of the print out respectively and;
- (3) the printing out at the end of data of the frequency of one to three word sequences by alphabetic, raw frequency and index order.

This program when supplied with the language samples collected from the subjects produced a total of 1723 pages of computer print-out for analysis from an input of 36,872 valid (i.e. subject spoken) morphemes of language.

(b) The computer print out

The output of the computer is in three parts, the procedure being replicated for the Macedonian, Vietnamese and Aboriginal samples separately and for the complete sample together. The parts are:-

- (1) The exact text submitted for analysis. At the end of each line of the text is a code number for the child and a line number given to the text, e.g.

```

-----750?  #ARE WE  #ALLOWED TO 1500063
-----
TO  #USE THIS. ?  THESE ----- 1500064
-----

```

The first three digits at the end of the line (150) identify the block of data as belonging to one particular child from one particular home language group. The last four digits identify the line of that text as printed. This text contains the language of both the subjects and all others with whom they interacted. The purpose of this complete text is to provide a wider context for particular utterances than is supplied in the single line of concordance listing (see (2) below). It is recognised that while sufficient context is usually supplied in the concordance this broader setting may not infrequently be desirable. On each occasion a new speaker starts speaking within the text the utterance is prefixed by a number which identifies the speaker of that sentence as belonging to a particular grouping, eg. a Vietnamese child (boy), the father or mother, a male or female teacher, etc. A full explanation of the allocation of speaker numbers is included in the Coding Instructions set out in Appendix 4.

- (2) Section two is basically a concordance which lists each word spoken by the subjects in the text in alphabetic order and shows the context in which the words occur. This concordance is preceded by a

frequency table for the language of the subjects showing the number and percentage of the characters used by them (alphanumeric only) e.g.

CHARACTER	FREQUENCY	FREQUENCY
#	11271	
*	284	
@	584	
A	11110	6.83
B	2202	1.35
C	2804	1.72
D	4562	2.80
E	16130	10.09

This is followed by statistics relevant to the text. These are listed below:-

- a) Total number of lines in input file;
- b) Total number of words in input file;
- c) Number of valid morphemes extracted
(Morphemes are separated from words in this program only in cases where this will simplify later analysis. Thus, no attempt has been made to mark all morphemes in the text. The computer counts as morphemes all groups of letters separated by either a space or other designated morpheme marks in the input data);
- d) Number of correct verbs in text;
- e) Number of correct nouns in text;
- f) Number of incorrect verbs;
- g) Number of incorrect nouns;
- h) Total number of speakers selected;
- i) K-factor used to calculate Index (See Chapter 3.1 (a)).

These statistics are followed by the alphabetic listing in a concordance of words and morphemes spoken by the subjects only, justified towards the right from the middle of the page, e.g.

The first morpheme to be listed will be "A". Within the group of sentences beginning with "A" the morphemes following "A" are also alphabetised, thus "A BETTER" will be listed before "A BIG". The sequence "A BETTER" is then ordered according to the alphabetic characteristics of the third word, so that "A BETTER ONE" is listed before "A BETTER THING". The listing continues in this way for a sequence of five morphemes.

Context for sequences is included for 49 spaces before and continued for 50 spaces after the space down the centre of the page, e.g.

```

..... 102 AND I#LL      #GO TO SCHOOL. 132 #PRETEND ...1020200 51
.....150 YES.  BECAUSE I #WANT TO    #GO TO THE+TOILET .....1500159 62
.....154 ALL THE  PEOPLE #NEED TO    #GO TO+BED/++PEOPLES .....1540256 31

```

The numbers at the end of the line allow the sequence to be referred back to the complete text for a particular child. This text is given in the first part of the print out (see 5(b)(1) above). The first seven numbers are the text and line number within the complete text. The last two numbers give the position in that line for the first morpheme of the

sequence, e.g. the sequence "GO TO SCHOOL" is referred back to the complete text by looking in the 51st space along line 200 of text 102. (See 5 (b) (1) above).

This section also includes a morpheme count for:-

- a) all single morphemes;
- b) all two morpheme sequences provided both morphemes occur five times;
- c) all three, four and five morpheme sequences provided the first two morphemes occur five times and the third morpheme twice. (A four morpheme sequence is dependent on the occurrence of a three morpheme sequence. Similarly for five morpheme sequences).

The morpheme count specifies:-

- a) the number of morphemes in a sequence;
- b) the number of speakers using that particular sequence;
- c) their speaker numbers and the number of times the sequence is spoken by any one speaker, e.g.

after many listings of 'DONT' one concordance comes to the following sequences:-

(154) ...WHO #CARES I	#DONT #WANT TO BE CAUGHT
(150)YOU	#DONT #WANT QUERY .
.....102	#DONT #WANT ONE
.....101 I	#DONT #WANT THEM @EM.....
.....101 I	#DONT #WANT THEM @EM.....
.....102 I	#DONT #WANT THIS QUERY ...
.....101 I	#DONT #WANT TO
.....101 I	#DONT #WANT TO
.....150 I	#DONT #WANT TO #PLAY CHASEY
.....101 I	#DONT #WANT TO #SEE THAT +TRICK
.....151 NO I	#DONT #WANT TO #TALK
.....151 ? I	#DONT #WANT TO #TALK #CANT YOU #SEE

3 MORPHEME SEQUENCE (DONT WANT TO)	SPOKEN 6 TIMES BY 3 SPEAKERS
	SPEAKER 101 - FREQUENCY 3
	SPEAKER 150 - FREQUENCY 1
	SPEAKER 151 - FREQUENCY 2

2 MORPHEME SEQUENCE (DONT WANT)	SPOKEN 12 TIMES BY 5 SPEAKERS
	SPEAKER 101 - FREQUENCY 5
	SPEAKER 102 - FREQUENCY 2
	SPEAKER 150 - FREQUENCY 2
	SPEAKER 151 - FREQUENCY 2
	SPEAKER 154 - FREQUENCY 1

(The concordance continues and, after other instances of 'DONT WORRY' and 'DONT YOU' lists the following):-

SINGLE MORPHEME (DONT)	SPOKEN 152 TIMES BY 8 SPEAKERS
	PERCENTAGE FREQUENCY = 0.90
	SPEAKER 100 - FREQUENCY 9
	SPEAKER 101 - FREQUENCY 27
	SPEAKER 102 - FREQUENCY 36
	SPEAKER 103 - FREQUENCY 17
	SPEAKER 150 - FREQUENCY 15
	SPEAKER 151 - FREQUENCY 11
	SPEAKER 153 - FREQUENCY 35
	SPEAKER 154 - FREQUENCY 2

"DONT WANT TO" is a three morpheme sequence that is spoken by three different speakers (101, 150, 151). The two morpheme sequence "DONT WANT" is spoken by five different speakers (101, 102, 150, 151, 154), while "DONT" as a single morpheme is spoken by eight different speakers in all (100, 101, 102, 103, 150, 151, 153 and 154).

For the one morpheme count it will be seen that additional information is given, i.e. Percentage frequency of occurrence, which is calculated thus:-

$$\frac{\text{Particular morpheme frequency}}{\text{Total morpheme frequency for the whole text}} \times \frac{100}{1}$$

As may be seen from the above example in this concordance section the utterances of each speaker are listed alphabetically under the speaker's code number.

Details of data preparation and a sample of the computer print out are included in Appendices 4, 5 and 6.

- (3) Section 3 of the print out contains the first basic analysis of the data gathered in the form of one morpheme, two morphemes and three morpheme sequences listed in alphabetic, raw frequency and index value orders.

It should also be noted that prior to these counts this section contains alphabetic listings of all specially marked morphemes such as nouns, verbs, questions, plurals, possessives and omissions.

BACKGROUND DATA ON SUBJECTS

Since it is clear that any thorough-going analysis of the data gathered would require to take into account many items of sociological data regarding the subjects, research assistants devised a 15 point interview schedule for the most important information needed. (See Appendix 7 for this schedule).

It is stressed, of course, that this information was not obtained in any single interview, indeed that, for the most part, it was not gathered in interviews at all but that the data was compiled gradually and naturally in visits to the home by the researchers and through a growing acquaintance - in many cases leading to lasting friendships - with the parents and subjects.

LIMITATIONS OF STUDY

No study of this scope dealing with naturalistic language and necessarily encompassing so many uncontrollable variables can be without important limitations. Listing these will enable appropriate weight to be given both to the data and to any analysis of and conclusions from them. No attempt has been made to place these in any order of importance:-

- a) The occasional 'setting up' of a situation, particularly with the Vietnamese sample, has already been mentioned. Taking along a child for the Vietnamese subject to play with could, strictly interpreted, be taken as altering the situation and preventing the recording of natural language. However it is difficult in this case to give great weight to this possible objection as it is unlikely that much language would be immediately learned from the visiting child while the advantage of having English spoken appeared to outweigh heavily any possible interference with planned procedures.
- b) It was not possible to ensure completely equal periods of recording from each period of the day (see section 3: Recording of Language, above.)
- c) It was not possible to obtain equal numbers of children in each of the three language samples, withdrawal of researchers or subjects making the

initial start with equal numbers quickly impracticable to maintain.

- d) There are more doubts over the language obtained from the Aboriginal sample than from either of the other samples. This sample was obtained during a one-week trip to Wiluna by a team of researchers who camped there in the grounds of the primary school. In spite of most whole hearted cooperation from the school authorities the time was scarcely sufficient to establish the kind of familiarity with procedures or rapport with subjects and families that were achieved with the other samples. For example in few cases was it possible to obtain the background information from the families of the children themselves; school and other records had to supply as much information as possible (For a sample of the schedule of background information used with Aboriginal subjects see Appendix 8). It was not always possible, for example, to be certain of the tribal language background of the subjects or of their precise ages.
- e) While care has been taken by members of the research team to achieve accurate transcriptions of the recorded language there will always be differences of opinion on distorted or doubtful sections of the tapes. Indeed interference proved quite a serious problem in certain locations, another almost inevitable corollary of attempting to collect naturalistic language with maximum freedom of movement for the subjects.
- f) An obvious and largely unanswerable objection is that language collected from any one subject on a particular day of the year may well not be typical of the subject's language performance even if complete acceptance of equipment, researchers and recording has been obtained. To this only the efforts of the researchers to ensure that recordings were taken in what had been discovered to be normal situations for the subjects can be opposed.
- g) As far as amount of language collected for analysis is concerned any possible doubt can probably be discounted. Tests carried out at Mount Gravatt have revealed that, in fact, virtually the same results would have been obtained from much smaller samples of the subjects' language.
- h) Obviously present analysis of data is also extremely limited being confined in this publication to frequency listings only. This is, of course, a limitation of time but in no way implies any permanent limitation on the uses to which the data may be put. It is hoped in the near future to publish studies on various aspects of the analysis of this present mass of raw data. Similarly the research team is most willing to make available wherever practicable, to researchers and other interested parties, for the cost of reproducing them, any portions of the recorded, transcribed, coded, programmed or background materials that could assist them in their own studies or programs.

While certain aspects of the procedure and analysis must be, and have been, tightened in the second year of the research program it is the belief of the research team that within the constraints placed by the desired collection of 'real' language in natural situations the data collected, while it certainly should not be subjected to fine statistical analysis, can sustain broad conclusions and provide valuable comparisons and contrasts with data obtained from elicited language in more controlled situations.

CHAPTER 3

LANGUAGE USAGE OF FIVE AND A HALF YEAR OLD MIGRANT AND ABORIGINAL CHILDREN FOR WHOM ENGLISH IS A SECOND LANGUAGE

This chapter contains a listing and brief discussion of the most frequently used words and sequences of words in the language of the 5½ year old subjects studied.

1. SINGLE WORD FREQUENCIES

(a) Calculation of Index

Each word is ordered by a special weighting of the actual word frequency. The weighted frequency is defined as a composite of the actual number of times a word or sequence is spoken and the number of speakers using that word, e.g. within the Macedonian sample

$$\begin{aligned}
 \text{word} &= \text{all} \\
 \text{actual frequency} &= 42 \\
 \text{no. of speakers} &= 7 \\
 \text{total no. of speakers} &= 8 \\
 \text{weighted frequency} &= \frac{42}{1} \times \frac{7}{8} \\
 &= 36.75
 \end{aligned}$$

The actual frequency of usage is multiplied by the ratio of the number of speakers using that word, to the total possible number of speakers.

To allow comparison with data obtained for word frequency in the language of children within the Van Leer and Mount Gravatt Projects as well as inter-group comparisons within the present study, the above weighted frequency was further adjusted to an Index which allowed for differing volumes of language collected for each age group. The amount of language (in morphemes) recorded from four and a half year old children in the Van Leer Project was taken as the base and a constant K was calculated by expressing the total number of morphemes collected in the present sample, or in subgroupings within the present sample, as a fraction or inverse fraction (depending on whether the amount was smaller or larger) of the total number of morphemes spoken by the 4½ year old sample. The weighted frequency was then multiplied by K to give the final Index, e.g.

$$\begin{aligned}
 \text{word} &= \text{all} \\
 \text{weighted frequency} &= 36.75 \\
 \text{K for 5½ yr olds} &= .764832 \\
 \therefore \text{Frequency index} & \\
 \text{for "all"} &= 36.75 \times .764832 \\
 &= 28.10
 \end{aligned}$$

(b) Most Frequent Single Words

Table II shows the 25 words most frequently used in the language of 5½ year old children in the total sample and subgroups of the present sample compared with the most frequently used words in the language of 5½ year old Mount Gravatt samples of first language speakers of the same age, when both are listed by index value.

TABLE III

A COMPARISON OF THE TWENTY SINGLE WORDS MOST FREQUENTLY USED BY THE MT. LAWLEY AGE 5½
MIGRANT AND ABORIGINAL SAMPLE, AND THE MT. GRAVATT (FIRST LANGUAGE) AGE 5½ SAMPLE

ORDER	TOTAL MIGRANT AND ABORIGINAL		ABORIGINAL		MACEDONIAN		VIETNAMESE		MT. GRAVATT (FIRST LANGUAGE)	
	WORD	INDEX	WORD	INDEX	WORD	INDEX	WORD	INDEX	WORD	INDEX
1	I	489	I	480	I	593	the	624	I	449
2	the	347	you	396	you	448	one	433	it	310
3	one	325	a	334	's/is	308	yes	409	you	293
4	you	321	one	332	it	244	is/'s	375	is/'s	243
5	is/'s	304	what	328	the	239	two	366	the	214
6	a	247	is/'s	271	one	222	I	362	to	189
7	no	211	that	258	a	204	no	299	a	184
8	yes	199	it	253	to	184	three	295	that	180
9	two	191	this	210	that	179	and	289	and	170
10	and	187	no	174	this	172	a	258	yes	140

11	it	180	he	172	no	167	here	220	no	123
12	here	159	here	169	and	160	four	200	one(s)	110
13	this	154	the	169	what	151	five	165	on	101
14	why	147	there	158	two	141	come	149	what	98
15	three	138	got	156	me	122	you	145	oh	96
16	that	125	oh	143	yes	122	see	135	in	95
17	look	118	to	129	don't	116	can	128	have	94
18	to	102	look	128	'm	115	little	126	got	93
19	four	99	up	103	look	112	six	124	this	90
20	oh	92	in	98	here	112	look	117	my	86

2. LANGUAGE SEQUENCES

(a) Selection of Sequences

Lists of the most frequently used sequences of two and three morphemes were compiled. To exclude unusual sequences the computer was programmed to list only those sequences in which the first two morphemes both occurred five times. To be considered for listing as a three morpheme sequence, the two morpheme sequences then had to be spoken by five subjects and the third morpheme had to be spoken by at least two children:-

e.g. ON THE BACK
 ON THE BEACH
 ON THE BOOK
 ON THE BOOK
 ON THE BOOK
 ON THE CAP
 ON THE CASE
 ON THE SHIP
 ON THE TREE

'ON THE' would be accepted as a two word sequence if it is spoken by at least two speakers because both words occur together at least five times. However 'ON THE BOOK' would be the only acceptable three word sequence provided it too is used by two different speakers.

Because of these stringent criteria it is highly likely that all the two and three word sequences extracted are well integrated into the oral language system of the subjects.

b) Calculation of Indices

Indices for two morpheme sequences were obtained by multiplying the index for the total occurrence of the first morpheme by the number of times the second morpheme follows the first. In the case of three morpheme sequences this figure was again multiplied by the number of times the third morpheme followed the second.

The indices were first weighted for the ratio of actual speakers to possible speakers and adjusted for comparison with the Mount Gravatt study in a similar manner to one word frequencies.

This method of calculation of the index was selected by the Mount Gravatt research team to give extra power to sequences which make up meaningful units of language. Basic to these meaningful units with first language speakers are structure words which act as signalling devices in language. As in L1 the most frequently used single words are usually structure words, the use of the total index for the first word tends to order the units so that those beginning with structure words have extra weight, but the index is also influenced by the frequency of the words following it.

If, for example, we take two sequences both of which as a whole occur eleven times, 'I'M A' and 'ON THE GRASS'. We are interested in the usage of 'I'M' and 'ON THE' as much as the sequences 'I'M A' and 'ON THE GRASS' themselves. An index based on the occurrence of each phrase as a discrete entity would rate both sequences equally. The index calculated for this complete sample gives far more weight to the more widely used 'I' and 'I'M'. Such an index thus allows generalisations beyond the particular phrase 'I'M A' to I -- or I'M. The units which come out at the top of the list are thus likely to be meaningful units or meaningful parts of larger units.

c) Most frequent two and three morpheme sequences.

Tables IV and V below give the twenty most frequently occurring two and three word morphemes used in the language of 5½ year old children in the total group and subgroups of the present sample compared with the most frequently used morpheme sequences used in the language of the 5½ year old Mount Gravatt sample of first language speakers of the same age when ordered by Index value.

d) Brief discussion

After considerable discussion of the pilot study results it was decided that in the present project the team had really no option other than to use an unscreened sample of children for the study. In the first place it seems desirable to gain a true picture of the possible kinds of range in second language development which could scarcely be examined if extremes were to be cut from the sample. Secondly it was, in any case, considered that even should there be a desire to exclude extremes there was as yet no reliable and valid test established for this population upon which to base such selection. Most practically important however was the fact that the total population of the sub groups scarcely permitted any 'selecting out'. While it might seem that the twin requirements of falling within a particular age range and coming from a home where a specific language is the means of communication are minimal, when applied, together with a reasonable approach to travel and time, there was in fact little choice in the sample to be studied.

A second observation would be that the tentative conclusions reached and problems discovered through the pilot study proved surprisingly prophetic in spite of the tiny sample used in that study (McGregor, 1978). This appears to further confirm the conclusions reached by Mount Gravatt researchers that reliable data can be obtained from smaller samples of language than those presently being gathered.

As far as the results themselves are concerned it has become clear that language items which were, in the pilot study, thought to be atypical because of the effect of one talkative subject are, in fact, in virtually the same places on the present lists. Some interesting possibilities thus begin to emerge:-

- (i) the basic observation on total data obtained is that while a larger total number of words was obtained for the present study than from the the corresponding Mount Gravatt cohort (99,395 v 87,418) the computer print out was, in its concordance section, considerably smaller (1723 v 2662 pages). This simply means that the number of valid morphemes extracted for processing was much smaller in the case of the present sample; the migrant and Aboriginal children, taken as a group, spoke far less English than the 5½ year old native speakers - scarcely a surprising result. The K-factors for the sub group are a further pointer to some of the details of this general observation (Aboriginal: 1.87442, Macedonian: .764832, Vietnamese: 1.00328). This indicates more than twice the valid morphemes for the Macedonian group when compared with the Aboriginal group. This result would probably be not so far removed from the Vietnamese contrast also when it is remembered that there were ten Vietnamese subjects v eight in each of the other sub-groups. The obvious question is therefore what the smaller samples of language data collected from the Aboriginal and Vietnamese groups indicate. Close examination of the lists should give some indications on this point.

TABLE IV

A COMPARISON OF THE TWENTY TWO-MORPHEME SEQUENCES MOST FREQUENTLY USED BY THE MT. LAWLEY AGE 5½ MIGRANT AND ABORIGINAL SAMPLE, AND THE MT. GRAVATT (FIRST LANGUAGE) AGE 5½ SAMPLE

ORDER	TOTAL MIGRANT AND ABORIGINAL		ABORIGINAL		MACEDONIAN		VIETNAMESE		MT GRAVATT (FIRST LANGUAGE)	
	SEQUENCE	INDEX	SEQUENCE	INDEX	SEQUENCE	INDEX	SEQUENCE	INDEX	SEQUENCE	INDEX
1	I'm	430	I'm	459	I'm	676	one-two	859	I'm	350
2	one two	289	what's	261	I'll	236	two three	512	I'll	210
3	I can	151	I'll	140	I know	182	three four	405	it's (it is)	157
4	I am	131	that one	130	I can	179	the little	219	I don't	144
5	two three	130	this one	94	I don't	163	I can	195	I got	126
6	I don't	93	's that	84	I am	151	four five	168	I know	101
7	three four	82	I want	68	I've	127	I am	163	I can	90
8	I'll	66	's this	65	one two	120	the door	118	I've	76
9	I know	61	I can't	56	it's	112	five six	98	you're	74
10	what's	59	it's	50	that's	110	I'm	93	got a	72

11	it's	58	I don't	45	you're	90	see the	86	that's(is)	65
12	I want	56	one two	44	I want	89	the water	83	I can't	61
13	I have	43	got a	44	I have	74	no no	72	I want	54
14	four five	41	a little	43	you know	67	I don't	71	I think	41
15	I've	38	's a	42	you can	48	the boy	70	I didn't	38
16	the little	34	that's	41	I was	46	and the	63	's a	36
17	that's	28	I can	41	you have	45	the ball	58	you know	31
18	this one	26	look at	38	look at	45	the dog	53	I was	30
19	no no	25	I know	37	you are	44	and a	40	I have	29
20	look at	24	I am	34	you say	43	six seven	37	I did	29

TABLE V

A COMPARISON OF THE TWENTY THREE-MORPHEME SEQUENCES MOST FREQUENTLY USED BY THE MT. LAWLEY AGE 5½ MIGRANT AND ABORIGINAL SAMPLE, AND THE MT GRAVATT (FIRST LANGUAGE) AGE 5½ SAMPLE

ORDER	TOTAL MIGRANT AND ABORIGINAL		ABORIGINAL		MACEDONIAN		VIETNAMESE		MT GRAVATT	
	SEQUENCE	INDEX	SEQUENCE	INDEX	SEQUENCE	INDEX	SEQUENCE	INDEX	SEQUENCE	INDEX
1	one two three	17830	what's this	7590	I'm not	11320	one two three	118270	I'm going	6126
2	two three four	6100	I'm going	4730	I'm going	6720	two three four	4313	I'm not	2960
3	three four five	2820	what's that	4410	I'm telling	3230	three four five	3270	I don't know	1903
4	I'm going	1450	I'm gonna	1610	one two three	3060	four five six	957	I'll get	531
5	I'm not	1280	I'm a	1500	I don't know	2070	five six seven	282	I'll have	509
6	four five six	1030	I'm after	1070	I've got	1870	I can see	211	I got a	456
7	I don't know	1000	I'm not	860	I'm making	1360	the little dog	167	you're not	454
8	what's this	780	I'll go	590	one two one	900	see the little	120	I've got	453
9	I can see	490	one two three	560	two three four	850	I don't know	103	I don't want	375
10	one two one	420	I'm up	430	I'm finished	650	six seven eight	72	I'm doing	366
11	I'm gonna	410	that's a	350	I don't want	620	what's this	71	I want to	333

12	I'm a	320	it's a	230	I want to	610	I can play	53	I'll go	309
13	five six seven	280	I don't know	220	I'm a	580	no no no	48	I'll put	274
14	what's that	270	I'll take	200	I'm getting	580	one two five	34	I'm a	260
15	I'm telling	250	got a balloon	180	it's not	510	one two four	34	I'm the	252
16	I've got	210	that one over	180	I'll get	470	the little cat	22	I got to	207
17	I want to	190	two three four	140	I can do	410	two three three	21	I got one	205
18	it's a	160	I'll do	130	I'll do	410	I am playing	20	I'll show	168
19	I'm getting	100	I want to	130	I'll tell	340	I am sitting	20	I'll tell	154
20	I'm making	100	I've got	90	that's a	290	the little boy	18	I'll be	154

(ii) [As regards the single word lists the most obvious observation is the close correspondence between the Mount Lawley and Mount Gravatt lists; not only the words but also the orderings of these words are very similar.] In the same way the total Mount Lawley sample index figures are very much on the same scale as for first language children. Closer examination of the sub-groups, however, begins to show up some interesting distinctions. In the Vietnamese sample differences from first language are far more obvious; only nine of the first twenty 'Vietnamese' words occur in the first twenty Mount Gravatt words, with thirteen 'Macedonian' and 'Aboriginal' words coinciding. Most obviously absent from the early Vietnamese word order are such structure words as 'in' and 'my', and to a lesser extent this is true of the other sub-groups also. On the other hand the presence of some words (such as numbers, particularly in the Vietnamese list) and 'what' in the other two sub-lists would suggest fairly straightforward hypotheses which could be further examined from the contexts of use; learners would of course, in asking questions of identification (what), while it is abundantly clear that in early stages of learning a language children put to heavy use the vocabulary that they are most sure of (i.e. numbers with the Vietnamese subjects). These tend to be used in all forms of play as well as in the classroom.

(iii) In two and three word sequences, however, the degree of similarity markedly decreases. Even in the two word sequences less than half the sequences are common to the two total groups, with the three-word sequences showing barely a quarter identical. Even allowing for the fact that about two or three of those used occur fairly quickly after the first twenty in the Mount Gravatt sample it is obvious that structures differ considerably in the two samples.

In the break-down the nature of this difference becomes even more obvious. If one takes the Aboriginal and Macedonian samples there are seven and nine structures identical to the first language group. However about four other Aboriginal and seven other Macedonian structures occur within the first thirty Mount Gravatt structures. On the other hand only one Vietnamese structure is identical and few of the others occur within the first forty or fifty first language structures. Notable are structures that ask for the identity of or identity something (what's, this, whats that, I can see, it's a, that's a) and of course, particularly with the Vietnamese, the ubiquitous numbers (though they are not entirely absent from either of the other groups).

(iv) While the nature of these differences will require close study the total absence of the shortened form of auxiliary verbs or of the copula from the Vietnamese first twenty list is too clear to be ignored. Similarly the influence of books (possibly readers) and of the classroom learning situation are fairly obvious in both the numbers and structures (see the little, the little dog, I can see, the little boy, etc.) With the Macedonian group, however, it would almost be true to say that, apart from numbers, all their structures are found in early stages of the first language list. While this is not true to quite the same degree of the Aboriginal group another half dozen structures at least are found in the first forty Mount Gravatt structures.

It is therefore worth examining the type of Index numbers attached to these structures to discover if other differences or similarities have emerged.

- (v) Figures for Index numbers present an interesting contrast, particularly within the total group. The overall pattern is for Index numbers to be much larger near the top of the lists than with the first language sample. This however soon alters, and by the end of the first twenty words (with the exception of the Macedonian group) the structures being used have much lower numbers than for the corresponding first language group. Once again the two sharply contrasting sub-groups are the Macedonian and Vietnamese sample. The former has a pattern somewhat 'stronger' than the first language sample with numbers remaining steadily larger right down the Index list. With the Vietnamese, however, the numbers drop very rapidly, even in the first twenty words. The Macedonian range of values (apart from 'I'm not') is very much the same as for first language values and this is broadly true of the Aboriginal range also. The Vietnamese range, however, after the 'one two three' phrase is not as extensive and, as has been noted, drops very quickly.

Plainly this is not the place in which the exact nature of these differences should be examined. There can however be little doubt in the case of a few points:-

- 1) The Macedonian sample is not very markedly different from first language speakers, suggesting that as far as their structures and vocabulary samples go, at least, they are very close indeed to first language competence, indeed practically indistinguishable from that.
- 2) While it may be true that in the early stages of learning English structures have a far smaller range of subjects which they are using than first language children, this is not true of the total sample where the range is, if anything, slightly larger. Thus the answer to one question raised in the pilot study may possibly be that migrant and Aboriginal children do NOT have a smaller range of two and three word sequences available for use except in the very early stage of learning English. This should become more obvious with passing cohorts. What does seem to be emerging is that there is a tendency to use a more limited number of these sequences much more frequently than is true of first language speakers; other sequences are known within the sample but not yet brought into common use. This appears to be in line with Mace-Matluck's (1979) suggestion that second language learning occurs not in a linear position as it does in first-language learning but in a global manner.
- 3) Thirdly the structures used by very early learners do differ markedly from first language competence. This is particularly true of the Vietnamese sample but also is evident in the Aboriginal language. However, it is perfectly possible that there could be many reasons for these differences. Further analysis should help to discover whether one set of learners is in an inter-language situation, another choosing to speak a different dialect of English, and whether there is any obvious first language interference. Previously noted possibilities of developing this study and extending it into the areas of phonology, interaction patterns, the social area of language, etc., remain.

It is hoped that some such studies will shortly be published beginning with a comparison with what is known of first language acquisition patterns, and with previous results of second language acquisition studies in elicited situations; another obviously desirable study would be an examination of the morphology related to the presumably wider semantic uses of those sequences used more heavily by second language learners.

REFERENCES

- Anderson, R: An implicational model for second language research, in press
- Baily, N. and others: Is there a natural sequence in adult second language learning? Language Learning, Vol. 24, December (1974).
- Brown, R: A First Language: The early stages, Harvard University Press, Cambridge, Mass. (1973).
- Cancino, H., Rosansky, E.J. and Schumann, J.H.: Testing hypotheses about second language acquisition: The copula and the negative in three subjects. Working Papers in Bilingualism, No. 3 (1974).
- Cancino, H., Rosansky, E.J., and Schumann, J.H: The acquisition of the English auxiliary by native Spanish speakers, TESOL Quarterly, Vol. 9, (1975).
- Dolch, F.W.: The basic sight word test, The Garrard Press, Champaign, Ill. (1942).
- Dulay, H.C., and Burt, M.K.: Should we teach children syntax? Language Learning, Vol. 23 (1973).
- Dulay, H.C. and Burt, M.K.: A new perspective on the creative construction processes in child second language acquisition, Language Learning, Vol. 24, (1974) (a).
- Dulay, H.C. and Burt, M.K.: Natural sequence in child second language acquisition, Language Learning, Vol. 24 (1974) (b).
- Ervin-Tripp, S: Structure and process in language acquisition, Monograph Series on Language and Linguistics, No. 23, (1970).
- Ervin-Tripp, S: Is second language learning like the first? TESOL Quarterly Vol. 8, (1974).
- Fatham, A.K.: Language Background, age and the order of acquisition of English structures. In M.K. Burt and H.C. Dulay (eds.), New Directions in second language learning, teaching and bilingual education. Teachers of English to Speakers of other Languages, Washington, D.C. (1975).
- Fry, E: Teaching a basic vocabulary, Elementary English, April, (1960).
- Hakuta, K: Learning to speak a second language. What exactly does the child learn? In D.P. Dato (ed.) Georgetown University Round Table on Languages and Linguistics, Georgetown University Press, Washington, D.C. (1975).
- Halliday, M.A.K: Learning to mean: Explorations in the development of language, Edward Arnold, London, (1972).
- Hart, N.W.M: Language research project; Research report No. 1, Mount Gravatt College of Advanced Education, Brisbane, (1973).
- Hart, N.M.W., and Walker, R.F.: Language research project: Research report No. 9: Mount Gravatt College of Advanced Education, Brisbane, (1976).

- Hart, N.W.M., Walker, R.F., and Gray, B., The Language of Children, Addison-Wesley Publishing Co., Reading, Mass. (1977).
- Larsen-Freeman, D.E.: The acquisition of grammatical morphemes by adult ESL students, TESOL Quarterly, Vol. 9, December (1975).
- Mace-Matluck, B.J.: The order of acquisition of certain oral English structures by native speaking children of Spanish, Cantonese, Tagalog, and Ilokano learning English as a second language between the ages of five and ten. Unpublished dissertation, The University of Texas at Austin, (1977).
- Mace-Matluck, B.J.: Order of acquisition: Same or different in first - and second-language acquisition? The Reading Teacher, Vol. 33, March (1979).
- McGregor, A.L.: The Development of English as a second Language in Aboriginal and migrant children: A pilot study, Mount Lawley College of Advanced Education, Perth, (1978).
- McLaughlin, B.: Second Language Acquisition in Childhood, Lawrence Erlbaum Associates, New Jersey (1978).
- Malmberg, B.: Drag ur en fy raarig finsk fluchas sprakliga utvectling, Nordisk Tidskrift for Vetenskap, Vol. 21 (1945).
- Matluck, J.H., and Mace-Matluck, B.J.: MAT-SEA-CAL Oral Proficiency Tests; English, Spanish, Cantonese, Mandarin, Tagalog, Ilokano. Field test edition in six volumes, Seattle Public Schools and Centre for Applied Linguistics, Arlington, Va (1974).
- Milon, J.P.: The development of negation in English by a second language learner, TESOL Quarterly, Vol. 8, (1974).
- Oller, J.W.: On syntax, semantics and pragmatics, Lingua (1972).
- Politzer, R.L.: Developmental sentence scoring as a method of measuring second language acquisition, Modern Language Journal, Vol. 58, (1974).
- Price, E.: Early bilingualism. In C.J. Dodson, E. Price, and L.T. Williams (Eds.), Toward Bilingualism, University of Wales Press, Cardiff, (1968).
- Ravem, R.: The development of Wh - questions in 1st and 2nd language learners, in J.C. Richards (ed.): Error analysis: Perspectives on second language acquisition, Longmans, London, (1974).
- Rosansky, E.J.: Methods and morphemes in second language acquisition research, Language Learning, Vol. 26, December, (1976).
- Wode, H.: Developmental principles in naturalistic L1 acquisition, Arbeitspapiere zum Spracherwerb, No. 16, Kiel University (1976).

APPENDIX 1

EQUIPMENT USED FOR RECORDING LANGUAGE

EQUIPMENT USED FOR RECORDING LANGUAGE

MICROPHONES:

AKAI ADM 40 (Commentator)
SONY ECM 16 (Subject)

TRANSMITTERS:

BEYER DYNAMIC "TRANSISTO PHONE"
TYPE TS 83

RECEIVERS:

BEYER Type NE 84

RECORDERS:

SONY TC 158 SD Stereo Cassette
A.W.A./CLARION
and MODEL:- MD 8080 K
STEREO DUAL CASSETTE DECK

MONITORING HEADPHONES:

SANSUI Type SS-10

APPENDIX 2

GUIDELINES FOR RECORDING PRAGMATIC INFORMATION

APPENDIX 2

GUIDELINES FOR RECORDING PRAGMATIC INFORMATION

What is said (or written) arises from the pragmatics of the utterance. Oller (1972) has pointed this out quite clearly and, with him, we are sure that language activity, at both syntactic and semantic levels, is inextricably interwoven with the context which gives rise to that activity and to which the language activity relates.

Halliday (1972) and others have made a very strong case that interpersonal purpose determines to a great degree the form of utterance.

Our conclusions are that utterance arises from what we could call meta-pragmatics (which includes interpersonal purpose) and that, at hierarchically lower levels of analysis, utterance cannot be interpreted (semantic) with reasonable confidence without mapping segments of utterance onto its situational and personal context.

Consequently, there are some features of the context which we want to record carefully, so that we can relate utterance to context. We think that we now know how to do this better than it has been done before and that we have both the technical and human resources to do this, and provide language data which will be uniquely useful to linguists everywhere.

We consider that there are three hierarchically different pragmatic aspects which are useful in interpreting utterance. They are:-

1. THE COMMUNICATION CONTEXT
2. INTERPERSONAL PURPOSE of a conversational sequence
3. THE FOCUS OF UTTERANCE

1. THE COMMUNICATION CONTEXT (of communication episodes)

This is the macro-pragmatic envelope within which the actions and utterances of all participants in any communication episode takes place. We want you to note three things about this context.

- a) The physical setting (Place)
- b) Other persons present (or alone)
- c) The activity situation (playing with x, watching an x cartoon on TV, getting ready for school, etc.)

2. INTERPERSONAL PURPOSE (of conversational sequence)

Each communication episode will be made up of a number of largely self-contained conversational sequences, bounded on either side by silence or by a change of topic or purpose. We are confident that you will be able to distinguish where one conversational sequence ends and another begins and classify each sequence in respect to its interpersonal purpose.

Six interpersonal purposes have been distinguished by Wells (and Slobin):-

a) Control

the control of the present or future behaviour of one or more the participants in the conversation.

Report this as "Control Interpersonal purpose: Jane is trying to get Mary to help her to x."

b) Expressive

the expression of feelings, emotions, attitudes and evaluation. Report this as "Annoyance at Mary about x or jubilant over winning the game of ..." or "shouting angrily" or "whining" or "describing which is best" or "Peter doesn't think much of this".

c) Social

the establishment of social relations. Report this as "Jane is trying to join in playing with x" or "Jane is trying to attract the attention of x" or "greeting class mates on arrival at school".

d) Representational

the giving and requesting of information. Report this as "Jane is telling/asking Mary about x".

e) Tutorial

deliberate teaching how to do something. Report this as "Jane is showing Mary how to x".

(NOTE: Don't try to force situations into this categorization if you do not think they fit. Before you attempt to describe the pragmatic situation you should make yourself very familiar with this classification system. Once you know this, the aim is to describe what is happening in terms of the interpersonal purpose of the utterance.)

3. FOCUS OF UTTERANCE (for each utterance if different from the preceding one)

a) Who is the recipient of the utterance?

Report this as "Jane is speaking to her teacher".

b) What the speaker is focussing on?

Report this as "She is looking at/pointing to/holding up x".

c) Activity - report changes here in activity to which utterance is related e.g. "Pouring water into a tin. The tin fell over. Mother is wiping up the water", but note only those activities to which her utterance directly relates.

GENERAL

1. RECORDING CONTEXT OF UTTERANCE

- a) First, state the COMMUNICATION CONTEXT and refer to this again only if it changes, i.e. the child moves elsewhere, someone else comes in or the activity situation changes - be sure to refer to

this as change in context.

- b) As soon as possible, try to give INTERPERSONAL PURPOSE of the conversational sequence. Meanwhile, direct your attention to the FOCUS OF UTTERANCE, being sure to state specifically the INTERPERSONAL PURPOSE as soon as it is clear to you. After this, go on with FOCUS OF UTTERANCE until a change in interpersonal purpose has become clear.
- c) In the FOCUS, two things are very important -
 - i) To whom is the child speaking;
 - ii) What objects or events are influencing her utterance.

You'll find sometimes, that the child is virtually talking to herself, or she will direct her utterance to others but they ignore or do not hear it. These things are important.

2. STYLE OR RECORDING

Try to be brief and to the point. If you use consistent terminology this will help transcription and keep the volume of commentary to a minimum. The combining of both aspects of focus in one short sentence, for example, should be possible, e.g. "She's showing her teacher her painting. The teacher walked away. She's showing it to the next door child."

Be sure that the child does not hear you making comments. The objective is to avoid any interference with the normal, on-going activity.

APPENDIX 3

EXAMPLE OF PAGE OF
TRANSCRIPTION

Participants
Time

Subject / Other Children
12 noon - 12:50 pm.

Child ELIKA

Line No

(Children trying to open drinks, talking and play at lunch time	(E) Oh don't ... Oh no! I'm going to Mrs Johnson. Wo Wo. The last resort, but she decides against it.	4/15	(RED TYPE)
	(Mac) Betty...Aah... (Ch) O.K. Here, give us it..... She asks her friend to open her juice carton.	4/16	(RED TYPE)
	(Ch) Careful, cause it's going to tip out. (E) Thank you.....Jennifer Friend is successful. Subject has gone	4/17	(RED TYPE)
	what you doing?....Hey, no that. That's Mrs. Johnson over to the climbing equipment. A child pulls the aerial hanging behind subject.	4/18	(RED TYPE) 34
	(Ch) That fell down. (E) Doesn't marrer. doesn't marrer..... You like it Subject explains that aerial is all right.	4/19	(RED TYPE)
	in the monkey bar?...Jennifer. Jennifer.. You - You like it? - You stop in Jennifer is climbing on the bars.	4/20	(RED TYPE)
	the parneh (?)....(Ch) Robert said if any girl's got a ring over here, he's going to marry - marry the girl if she's got a ring. (E) Ah, Alec.	4/21	
	Group is gathered around speaker.	4/22	(RED TYPE)

APPENDIX 4

INSTRUCTIONS FOR CODING

CODING OF TEXT FOR COMPUTER ANALYSIS

Alphabetic	A - Z (Capitals only)
Numeric	0 - 9
Space	
Special Characters	. (Period)

Special Characters (cont'd)

- (Hyphen)
/ (Slash)

+
*
@
: (Colon)
?

The use of any other characters in the coded text may cause errors in the computer processing.

- Use the following coding conventions to prevent confusion between numbers and letters:

Letter O is coded as 0, number zero as 0
Letter I is coded as 1, number one as 1
Letter Z is coded as 2, number two as 2

Identification of Speakers

Each speaker recorded in the text is to be identified by a unique three digit identification number. This number is assigned according to the characteristics of the speaker, as follows:

Experimental Subjects

<u>Number range</u>	<u>Speaker Classification</u>
100 - 199	Greek languages
200 - 299	Vietnamese languages
300 - 399	Aboriginal (tribal dialects)
400 - 499	Aboriginal (english dialect)
500 - 599	Other children

Within the range assigned to each classification above, the lower 50 numbers will signify a male subject (e.g. 100 - 149), while the upper 50 will signify a female subject (e.g. 150 - 199).

As well as the experimental subjects, the speech of other persons may be present in the text, and will be identified as follows:

<u>Number range</u>	<u>Speaker Classification</u>
700 - 799	Teachers
800 - 899	Parents
900 - 999	Other adults

The speaker's identification number must appear in the coded text, preceding his speech. The number must be preceded, and followed, by a space. All text after the number will be regarded as having been spoken by that speaker, until a new number is encountered in the text. For example:

150 THERE ARE THE BOYS. 155 I CAN SEE THEM .

4. Sentence Structure

The sentence is an important unit in the computer processing of the text. A coded sentence may contain the corrected text, as well as the original versions of the corrected words.

A sentence may begin anywhere on a coded line, and must be ended by a period. The period must be preceded and followed by a space (see the coding examples in the following sections).

5. Identification of Parts of Speech

Two parts of speech, nouns and verbs, will be identified in the text through the use of special identification characters as follows:

Nouns - Precede the word with a plus sign (+) to identify it as a noun.

For example, code 'BOY' as +BOY, etc.

(Do not use + for proper names)

Verbs - Precede the word with a 'hash' character (#) to identify it as a verb.

For example, code 'RUN' as #RUN, etc.

Contracted verb forms - Where a verb is contracted, using an apostrophe, replace the apostrophe with a 'hash' to identify the contracted verb. For example, code 'HE'S' as HE#S, etc.

6. Corrections

6.1 General

Corrections to speech will involve the substitution of corrected words in the text. Where the word has been identified as a noun or a verb, the applicable code (+ or #) is used twice to identify an incorrect part of speech (++ or ##). The original, incorrect word is placed at the end of the sentence, preceded by a special character to indicate that it is the original version. (The special characters to be used are given below.) Any identification codes used on the correct version of the word are also used on the original version to identify the error.

6.2 Grammatical Corrections

Corrections are applied as described above. The original, incorrect version of each corrected word is placed at the end of the sentence, and is immediately preceded by the slash character (/) to indicate a grammatical mistake.

Examples: "Them is bad boys." coded as:
HE #IS A BAD ++BOY /THEM /++BOYS .

"He cutted the string." coded as:
HE ##CUT THE +STRING /##CUTTED .

"We's going" coded as:
WE##RE GOING /##S .

6.3. Pronunciation Corrections

Corrections are applied as described in section 6.1. The original version of each corrected word is immediately preceded by the "at" character ("@") to indicate a pronunciation mistake. No correction codes are used.

For example, the text "Yeah I know you lika beansa". may be coded as:-

YEAH I #KNOW YOU #LIKE +BEAN-SP @LIKA @BEANSA

Below is a list of commonly used abbreviations in children's speech which would not be regarded as mispronunciations:-

Word	Abbreviations
and	an, a
around	round, roun
banana	nana
going to	gonna
him	im
he	e
no	nup, nuh
of	a
round	roun
them	em
to	ta
verb + "ing"	verb + "in" (comin, lookin)
want to	wanna
yes	yep, yup, yeah
you	ya
your	ya

Incorrect pronunciation must be distinguished from what is called incorrect grammar (i.e. forms restricted to child language). Examples of words which would not be marked as incorrect pronunciation are when the omitted sound or sounds is an inflection or if one word is substituted for another, e.g.:

"cat" for "cats"
 "open" for "opened"
 "me" for "my"

7. Expansion of Contracted Forms

The verbs "be", "have", "do", "will", "would", and "did" will be expanded from their contracted forms as follows:-

Contracted Form	Verb	Expanded Form
Noun's	is	NOUN#S
Noun's	had (got)	NOUN#SHA
Noun'd	had	NOUN#DHA
Noun'd	would	NOUN#DWO

<u>Contracted Form</u>	<u>Verb</u>	<u>Expanded Form</u>
Noun're	are	NOUN#RE
Noun've	have	NOUN#VE
Noun'm	am	NOUN#M
Noun'll	will	NOUN#LL
Noun's	was	NOUN#SWA
How'd	did	HOW#DDI
How'd	would	HOW#DWO
What's	does	WHAT#SDO

Negation:

can't	CANT
won't	WONT
don't	DONT

etc.

Note: The use of the 'hash' to replace the apostrophe is restricted to contracted VERB forms. In all other cases the apostrophe is simply omitted.

8. Expansion of Plural Forms

Regular plural forms are coded as follows:

NOUN-SP

Some collective nouns are not coded in this way, for example:

Scissors	SCISSORS (Not SCISSOR-SP)
Trousers	TROUSERS
Pants	PANTS
Glasses	GLASSES

etc.

9. Expansion of Possessive Forms

Possessive forms are coded as follows:

NOUN-SS

Forms which are possessive and plural are coded as:

NOUN-SPSS

10. Questions

Questions in the text, whether marked by sentence structure or by speaker intonation only, are identified by the insertion of a question mark ("?",) at the beginning of the sentence. The question mark must be preceded and followed by a space.

For example, the text "Why are you doing it." would be coded as:

? WHY #ARE YOU #DOING IT .

11. Uncertain Transcriptions

Where the transcription of the speaker's text is uncertain, it is to be identified as such so that the computer program may take the uncertainty into account. The coding procedures are as follows:

- a) Where a word or part of a word is uncertain, but a reasonable guess is made, the letter Q is added as the last letter of the word. For example:

"I'm doing it" would be coded as I#M #DOINGQ IT .

- b) Where a word or phrase cannot be identified the word "QUERY" may be substituted for it. For example, the text "He is" would be coded as:

HE #IS QUERY .

- c) If enough is heard to know the part of speech of an uncertain word, the name of the part may be substituted for the word, with the letter Q added as the first letter, i.e. QNOUN, QNAME etc. For example, the text "Hey (proper name).....what." would be coded as:

HEY QNAME WHAT .

12. Counting and Numerals

Numbers are written in word form when said by the subject, for example:

ONE TWO THREE etc.

Do not use digits as these will be confused with the speaker identification numbers in the text.

Conjoined numbers such as "Twenty-one", "Thirty-seven", etc are coded as:

TWENTYONE THIRTYSEVEN etc.

A number such as "One hundred and twenty seven" is coded as ONE HUNDRED AND TWENTYSEVEN

When the subject is counting, the words are written as a sequence with no commas or periods between:

ONE TWO THREE FOUR etc.

If the counting sequence extends past ten, it can be shortened to a range and coded as:

ONE-ONE HUNDRED AND TEN

3. Words spelt out by subject, or when quoting.

When the subject spells out a word, naming individual letters, these are marked by placing a colon (":") in front of each letter. For example, the text "C....A....T spells cat" would be coded as:

:C:A:T SPELLS +CAT .

Note: The colon and the letter are separate characters and are not joined together in the same space on the form. In the above example the entry ":C:A:T" would occupy six spaces.

4. Statements as a Group

Where the text is a statement made by the subject as part of a group, an asterisk ("*") is placed before the sentence to indicate this. The asterisk must be preceded and followed by a space. For example, the group statement "We are boys" would be coded as :

* WE ARE +BOY-SP .

(Where the subject is included in the group use the subject's number before the asterisk; otherwise another number representing the appropriate group of children.) An asterisk should also be used for singing.

5. Some miscellaneous words

The following coding conventions are to be used:

Okay	-	OKAY
Yuk	-	YUK
O'clock	-	OCLOCK
Let's	-	LETS

(In general, apostrophies are left out. In some circumstances they will be replaced by a hash (#) - see section 7.)

6. Missing Words

Put a bracket round any words omitted by the subject, e.g. "He is bad boy" coded as

HE #IS (A) BAD +BOY .

Such omitted words do not need to be listed as corrections at the end of the sentence.

Another example: "I done it" coded as:

I(#VE) #DONE IT .

Put a square bracket round any words added unnecessarily by the subject, e.g. "I not dont want it" coded as I [NOT] #DONT #WANT IT

(N.B. Be very careful to distinguish square from round brackets)

APPENDIX 5

MOUNT LAWLEY COLLEGE
SECOND LANGUAGE CONCORDANCE
PROGRAM:
PROGRAM DESCRIPTION

1. GENERAL

The 'MORPH' system consists of a set of programs for the analysis of speech, using the morpheme as the basic unit. Morphemes are extracted from text, sorted alphabetically, then printed in context. The frequencies of single morphemes and morpheme sequences are interspersed in the morpheme list. The system then prints frequency listings for one, two and three morpheme sequences. The system was designed by the Division of Computing Services of the Mount Lawley College of Advanced Education at the request of the English, Speech & Drama Department. The concept is based on a program written in Fortran for a large computer and it was not possible to simply adapt this to a mini computer. Consequently an entirely new set of programs written in BASIC was designed to provide the language concordance from a PDP11 based computer system where avoiding interference to other users rather than the total amount of run time was an important factor.

2. DATA INPUT

Data is input to the system from a terminal, using an input editing and formatting program which is part of the MORPH suite of programs. Text is entered in columns 1 to 72, as it is read from the coding sheets. (see appendix 4). Line numbers in sequence are appended automatically by the input program in columns 73-79.

When entering the text, the typist need only enter the principal speaker number once. The line numbers appended by the system consist of this number (3 digits) followed by a line sequence number (4 digits). The appended line number is used only as a label and is not used in computation.

The input data is stored as an ASCII text file, with a separate file being generated for each principal speaker. Each file is run through a preliminary editing program which reports any errors in text coding formats, these errors are then corrected using a standard text editor. Error free files for speakers of the same classification are then concatenated to form the input data file for an analysis run. Group classification is usually identified by some part of the speaker number.

3. PROGRAM CONTROL INPUT

All programs are designed to be run from a terminal, with control information being entered as required. For full analysis runs, the system is run in batch mode, with these commands coming from a control file.

4. OUTPUT

The output consists of the following:-

- a) An echo listing of the input data;
- b) A summary section containing details on the numbers of lines, words, and morphemes, etc., in the input file;
- c) A concordance listing of morphemes in alphabetical order, with details of frequencies and speakers for morphemes and morpheme sequences;
- d) Listings of correct and incorrect nouns and verbs in context;
- e) Frequency listings of one, two and three morpheme sequences, in alphabetic, frequency and index value sequences.

5. PROGRAM DESCRIPTIONS

5.1. General

The MORPH system consists of several programs written in

extended BASIC-PLUS, running on a PDP 11/34A minicomputer under the RSTS/E operating system. The computer system has 128K words of memory and 200Mb of on-line disk storage. As well as the programs written for the MORPH suite, the system also uses the utility sort program.

5.2. MORPH1 - Text input program

This is a text input program which accepts data from a keyboard and enters it into a data file on disk, after appending a line number to each line.

5.3. MORPH2 - Preliminary text check

This program runs through the data file and verifies that the coding is in accordance with the text coding rules. Any errors are reported with line and character position to facilitate editing with a standard utility text editor.

5.4. MORPH3 - Morpheme extraction

This program accepts up to 50 speaker identification numbers, and then extracts all morphemes spoken by these speakers. Any single morpheme or sequence of morphemes is limited to 50 characters in length. Certain classifications of morphemes are marked so as to be sorted to the end of the morpheme file, rather than in the machine's normal sequence.

5.5. Utility sort

The morpheme file is sorted alphabetically within the broad sequence of:-

- valid morphemes beginning with alpha, # or +
- valid morphemes beginning with -, :, ? or *
- invalid morphemes beginning with / or @
- missing text
- inserted text
- invalid sequences beginning with \$

5.6. MORPH5 - Morpheme Analysis

The analysis of morpheme sequences is done in this program, using the sorted morpheme file. This program outputs the text list, summary information and concordance listing to a listing file, at the same time constructing a file of morpheme sequence frequency information.

5.7. Utility sort

The morpheme sequence frequency information is sorted into three output files, based on alphabetic, frequency and index value sequences.

5. MORPH6

This program formats and lists the frequency information in the above files, appending it to the main listing produced by MORPH5.

5.9. MORPH9

This program is used where a number of data files are to be concatenated before processing. The program accepts file names and then creates a new file containing data from the named files.

6. BATCH RUN

The following is a typical control file used to run the MORPH system:-

```
$JOB/NAME=VIETNM/NOLIMIT/PRIORITY:-16
$!
$!CONCATENATE THE DATA FILES FOR THIS RUN
$MESSAGE CREATING THE DATA FILE FOR THIS RUN
$RUN MORPH9
VIETNM
NOELWR
JOHNS
DANIEL
DOLMAN
SINGH
JOHNS2

$!
$!EXTRACT THE MORPHEMES FOR DESIRED SPEAKERS
$MESSAGE STARTING MORPHEME EXTRACTION
$!
$RUN MORPH3
VIETNM
222
225
271
272
274
275

$!
$!SORT THE MORPHEME FILE INTO ALPHABETIC ORDER
$MESSAGE STARTING MORPHEME SORT
$!
$RUN "$SORT"
VIETNM.MPS/FO:S:67=VIETNM.MPH/FO:S:67/KE:1.50
$EOD
$!
$!DELETE THE UNSORTED MORPHEME FILE
$DELETE VIETNM.MPH
$!
$!RUN THE MORPHEME SEQUENCE EXTRACTION
$MESSAGE STARTING MORPHEME SEQUENCE ANALYSIS
$!
$RUN MORPH5
VIETNM
$!DELETE THE FILES NO LONGER NEEDED
$DELETE VIETNM.DAT VIETNM.LIN VIETNM.MPS VIETNM.WRK
$!
```


5.9. MORPH9

This program is used where a number of data files are to be concatenated before processing. The program accepts file names and then creates a new file containing data from the named files.

6. BATCH RUN

The following is a typical control file used to run the MORPH system:-

```
$JOB/NAME=VIETNM/NOLIMIT/PRIORITY:-16
$!
$!CONCATENATE THE DATA FILES FOR THIS RUN
$MESSAGE CREATING THE DATA FILE FOR THIS RUN
$RUN MORPH9
VIETNM
NOELWR
JOHNS
DANIEL
DOLMAN
SINGH
JOHNS2

$!
$!EXTRACT THE MORPHEMES FOR DESIRED SPEAKERS
$MESSAGE STARTING MORPHEME EXTRACTION
$!
$RUN MORPH3
VIETNM
222
225
271
272
274
275

$!
$!SORT THE MORPHEME FILE INTO ALPHABETIC ORDER
$MESSAGE STARTING MORPHEME SORT
$!
$RUN "$SORT"
VIETNM.MPS/FO:S:67=VIETNM.MPH/FO:S:67/KE:1.50
$EOD

$!
$!DELETE THE UNSORTED MORPHEME FILE
$DELETE VIETNM.MPH
$!
$!RUN THE MORPHEME SEQUENCE EXTRACTION
$MESSAGE STARTING MORPHEME SEQUENCE ANALYSIS
$!
$RUN MORPH5
VIETNM
$!DELETE THE FILES NO LONGER NEEDED
$DELETE VIETNM.DAT VIETNM.LIN VIETNM.MPS VIETNM.WRK
$!
```

```

$!SORT THE FREQUENCY FILE TO THE VARIOUS REQUIRES SEQUENCES
$MESSAGE STARTING MORPHEME FREQUENCY FILE SORT
$!
$RUN "$SORT"
VIETNM.MFA/FO:S:72=VIETNM.MPF/FO:S:72/KE:C1.1:C22.50
VIETNM.MFF/FO:S:72=VIETNM.MPF/FO:S:72/KE:C1.1:C02.6:C22.50
VIETNM.MFI/FO:S:72=VIETNM.MPF/FO:S:72/KE:C1.1:C014.8:C22.50
$EOD
$!
$!PRINT THE SEQUENCE LISTINGS
$MESSAGE STARTING SEQUENCE LISTINGS
$!
$RUN MORPH6
VIETNM
$!DELETE THE REMAINDER OF THE RUNNING FILES
$DELETE VIETNM.MPF VIETNM.MFA VIETNM.MFF VIETNM.MFI
$!
$!PRINT THE LOG FILE
$PRINT VIETNM.LOG
$!
$!END OF MORPHEME RUN
$!
$EOJ

```

7. BATCH OUTPUT

The following log file is the result of a run with the above control file. As the job is run at a low priority on a multi-user machine, the run times are lengthy.

```

-----$JOB/NAME=VIETNM/NOLIMIT/PRIORITY:-16
:22:54 PM HELLO

:22:55 PM RSTS V06C-04 Mt Lawley CAE Job 16 KB1 12-Jun-80 03:22 PM
:22:56 PM #66/6
:22:57 PM Password:
:22:57 PM 1 other user is logged in under this account

:22:57 PM Ready

-----$!
-----$!CONCATENATE THE DATA FILES FOR THIS RUN
-----$MESSAGE CREATING THE DATA FILE FOR THIS RUN
-----$RUN MORPH9
:22:58 PM RUN MORPH9

:23:05 PM OUTPUT FILE
:23:05 PM VIETNM
:23:06 PM INPUT FILE(S)
:23:06 PM ? NOELWR
:23:30 PM ? JOHNS
:23:33 PM ? DANIEL
:24:16 PM ? DOLMAN
:24:48 PM ? SINGH
:25:37 PM ? JOHNS2
:25:52 PM ?

:25:53 PM Ready

```

```

-----$!
-----$!EXTRACT THE MORPHEMES FOR DESIRED SPEAKERS
-----$MESSAGE STARTING MORPHEME EXTRACTION
-----$!
-----$RUN MORPH3
03:25:53 PM RUN MORPH3
03:25:55 PM MORPH3 - Morpheme analysis data extraction
03:25:55 PM Input file ? VIETNM
03:25:56 PM Subject ID ? 222
03:25:56 PM Subject ID ? 225
03:25:57 PM Subject ID ? 271
03:25:58 PM Subject ID ? 272
03:25:59 PM Subject ID ? 274
03:26:00 PM Subject ID ? 275
03:26:01 PM Subject ID ?

04:48:46 PM No errors encountered in 2716 lines of text.
04:48:46 PM 29493 words processed in text
04:48:46 PM 17344 lines written to morpheme file
04:48:46 PM 15729 valid morphemes written to morpheme file

04:48:47 PM Ready

-----$!
-----$!SORT THE MORPHEME FILE INTO ALPHABETIC ORDER
-----$MESSAGE STARTING MORPHEME SORT
-----$!
-----$RUN "$SORT"
04:48:47 PM RUN $SORT

04:48:50 PM SRT>VIETNM.MPS/FO:S:67=VIETNM.MPH/FO:S:67/KE:1.50

04:55:37 PM SRT -- M:ELAPSED REAL TIME: 00:06:41
SRT -- M:TOTAL RECORDS SORTED: 17344
-----$EOD
04:55:37 PM SRT>^Z

04:55:38 PM Ready

-----$!
-----$!DELETE THE UNSORTED MORPHEME FILE
-----$DELETE VIETNM.MPH
04:55:38 PM RUN $PIP
04:55:39 PM *VIETNM.MPH/DE
04:55:40 PM *^Z

04:55:41 PM Ready

-----$!
-----$!RUN THE MORPHEME SEQUENCE EXTRACTION
-----$MESSAGE STARTING MORPHEME SEQUENCE ANALYSIS
-----$!
-----$RUN MORPH5
04:55:42 PM RUN MORPH5
04:56:15 PM MORPH5 - Morpheme sequence analysis
04:56:15 PM Input File? VIETNM

```

12:24 AM Ready

```
-----$!DELETE THE FILES NO LONGER NEEDED
-----$DELETE VIETNM.DAT VIETNM.LIN VIETNM.MPS VIETNM.WRK
12:24 AM RUN $PIP
12:26 AM *VIETNM.DAT,VIETNM.LIN,VIETNM.MPS,VIETNM.WRK/DE
12:29 AM *^Z
```

12:30 AM Ready

```
-----$!
-----$!SORT THE FREQUENCY FILE TO THE VARIOUS REQUIRES SEQUENCES
-----$MESSAGE STARTING MORPHEME FREQUENCY FILE SORT
-----$!
-----$RUN "$SORT"
12:30 AM RUN $SORT
```

12:33 AM SRT>VIETNM.MFA/FO:S:72=VIETNM.MPF/FO:S:72/KE:C1.1:C22.50

```
12:56 AM SRT -- M:ELAPSED REAL TIME: 00:00:21
SRT -- M:TOTAL RECORDS SORTED: 861
```

12:56 AM SRT>VIETNM.MFF/FO:S:72=VIETNM.MPF/FO:S:72/KE:C1.1:C02.6:C22.50

```
12:21 AM SRT -- M:ELAPSED REAL TIME: 00:00:23
SRT -- M:TOTAL RECORDS SORTED: 861
```

12:21 AM SRT>VIETNM.MFI/FO:S:72=VIETNM.MPF/FO:S:72/KE:C1.1:C014.8:C22.50

```
12:47 AM SRT -- M:ELAPSED REAL TIME: 00:00:24
SRT -- M:TOTAL RECORDS SORTED: 861
```

```
-----$EOD
12:47 AM SRT>^Z
```

12:48 AM Ready

```
-----$!
-----$!PRINT THE SEQUENCE LISTINGS
-----$MESSAGE STARTING SEQUENCE LISTINGS
-----$!
-----$RUN MORPH6
12:48 AM RUN MORPH6
12:54 AM MORPH6 - Sequence Frequency Listing Program
12:54 AM Input File ? VIETNM
```

06:51:46 AM Ready

```
-----$!DELETE THE REMAINDER OF THE RUNNING FILES
-----$DELETE VIETNM.MPF VIETNM.MFA VIETNM.MFF VIETNM.MFI
06:51:46 AM RUN $PIF
06:51:47 AM *VIETNM.MPF,VIETNM.MFA,VIETNM.MFF,VIETNM.MFI/DE
06:51:50 AM *^Z
```

06:51:51 AM Ready

```
-----$!
-----$!PRINT THE LOG FILE
-----$PRINT VIETNM.LOG
06:51:51 AM RUN $QUE
06:51:53 AM QUE V06C-03 RSTS V06C-04 Mt Lawley CAE
06:51:53 AM *Q VIETNM=VIETNM.LOG
06:51:55 AM *^Z
```

06:51:56 AM Ready

```
-----$!
-----$!END OF MORPHEME RUN
-----$!
-----$EOJ
06:51:56 AM BYE
06:51:58 AM Confirm: YES
06:52:00 AM Saved all disk files; 8840 blocks in use
06:52:01 AM Job 16 User 66,6 logged off KB1 at 13-Jun-80 06:52 AM
06:52:02 AM System RSTS V06C-04 Mt Lawley CAE
06:52:02 AM Run time was 12 hours, 30 minutes, 53.1 seconds
06:52:02 AM Elapsed time was 15 hours, 29 minutes
06:52:02 AM Good morning
```

APPENDIX 6

SAMPLE PAGE FROM A LANGUAGE
CONCORDANCE PRINT OUT FOR
FIVE AND A HALF YEAR OLD
CHILDREN (MACEDONIAN SAMPLE)

CONCORDANCE LISTING

#DONT #BREAK IT YOU +STUPID ? WHERE#S BEITY
 LIONS 103 YOU #HAD YOUR +SHOT 501 QUERY 103
 560 QUERY YOUR MUM #DOES THIS 151 UM AH QUERY
 T (ONE) (ON) HER /AFTER /AFTER /AFTER @GOIN 100
 554 QUERY WHEN YOU #WERE #TALKING 150 I #KNOW
 O OH PLEASE I #WANT TO #SHOW YOU A +TRICK 101
 LL MY +MUM 502 YOU #KNOW WHAT THAT #MEANS 103
 100 OH THE +LADY 511 +COME ON TONY @CMON 100
 502 #DID 501 YES #LOOK HOW CLOSE YOU #ARE 103
 E QUERY 554 QUERY WHEN YOU #WERE #TALKING 150
 #DIDNT YOU @YA I #M NOT #PLAYING WITH YOU 101
 WHEN YOU #WERE #TALKING 150 I #KNOW I #KNOW
 #M #TELLING ON YOU 103 OH THAT#S +BUGS +BUNNY
 3 MY +DAD-@S LAUGH 502 YOUR +DAD-@S HAPPY 103
 ONT #KNOW ENGLISH YOU #DONT #KNOW ENGLISH 150
 EE FOUR FOUR +INCH-SP MATE 502 OH QUERY 103
 LL ON YOU 103 I##VE #BEEN #DRAWING ON A QUERY
 TH DINAH @WIF * SOMEONE (*IS) (IN) THE +KITCHEN
 -SP AND PINK +SHOE +LACE-SP POLKA DOT +VEST AND
 SIDE #DO YOU #WANT 150 SIDE ONE NOT SIDE TWO

*** 3 MORPHEME SEQUENCE (I KNOW I)

BE AT +NIGHT +TIME 191 OH BOB #SIT DOWN 101
 Y YOUR MUM #DOES THIS 151 UM AH QUERY I #KNOW
 H @WIF * SOMEONE (*IS) (IN) THE +KITCHEN I KNOW
 132 ? YOU READY 102 NO @NUH YOU#RE #COMING
 U #DONT EVEN #CARE 190 QUERY THE +WORD-SP 101
 #KNOW THE +STORY ON THIS 190 ? WHAT /HMM 101
 OU +LEAVE THESE YOU CAN #PRESS THIS YES @YEAH
 #DOESNT #KNOW WHAT DUMB #MEANS 103 I #AM QUERY
 SNT #KNOW WHAT A HUNDRED #ADD A HUNDRED #IS 103
 ? LEEANNE (#WILL) YOU #HELP ME I#LL #HELP YOU
 SAW YOU #COPING IT 501 THAT#S NOTHING 103 AND

*** 3 MORPHEME SEQUENCE (I KNOW WHAT)

P #SPOILT IT AND MY +BROTHER#S #BREAKING IT 102
 HUCKED 132 I #CANT #FIND THAT FOR +SCHOOL 102
 YEAH 501 THIS CUNT #DOESNT #KNOW ANYTHING 103
 S I #GOTTA #GO LIKE THIS 102 NO NOT THAT +WAY

*** 3 MORPHEME SEQUENCE (I KNOW YOU)

*** 2 MORPHEME SEQUENCE (I KNOW)

I #KNOW 584 ? WHAT #IS THIS 153 THAT#S (A) ++G	1530108	58	I
I #KNOW 501 THIS #IS WHAT YOU #CALL AN EXPERT +T	1030294	5	I
I #KNOW I #KNOW SINK 560 ? WHO 151 MY MUM #T	1510246	60	I
I #KNOW 505 NOW QUERY #GOT ONE 505 NOW THAT#S	1000033	13	I
I #KNOW I #KNOW 553 AH THIS (#WAS) ##GIVEN (TO)	1500078	51	I
I #KNOW QUERY I #WANT TO #TAKE THE +KING-SP OUT	1010612	35	I
I #KNOW 502 ? WHAT 103 I #DONT #HAVE TO #TELL	1030140	5	I
I #KNOW ? #DID YOU #CALL ME 515 THAT +WIRE #BR	1000119	39	I
I #KNOW 502 YES BECAUSE I #CANT #DO IT FROM THER	1030311	47	I
I #KNOW I #KNOW I #KNOW 553 AH THIS (#WAS) #	1500078	41	I
I #KNOW 501 ? THEN WHY #DID YOU #HAVE TO #GET TH	1030013	29	I
I #KNOW 553 AH THIS (#WAS) ##GIVEN (TO) (ME) (BE	1500078	61	I
I #KNOW YOU#RE #COPING +BUGS +BUNNY I #KNOW /##	1030445	30	I
I #KNOW 501 ? WHERE ##ARE YOU #GOING /##ARE @YA	1030227	38	I
I #KNOW ENGLISH BY QUERY ON THE +BUM OH YOU	1500041	5	I
I #KNOW FOUR +INCH-SP 502 YOU #HIT IT @YA NOT	1030259	41	I
I #KNOW HOW TO #DRAW A QUERY 501 JUST #DO IT ON	1030404	29	I
I #KNOW I #KNOW SOMEONE (#IS) (IN) THE +KITCHEN#WIT	1000002	14	I
I #KNOW I #MEAN WITH TAN +SHOE-SP AND PINK +SHOE +	1510348	43	I
I #KNOW I #KNOW 926 ? ARE YOU ALLOWED TO PUT IT	1500177	37	I

SPOKEN 3 TIMES BY 3 SPEAKERS

SPEAKER 100 - FREQUENCY 1

SPEAKER 150 - FREQUENCY 1

SPEAKER 151 - FREQUENCY 1

I #KNOW LOTS OF TRICKS I #WANT TO #SHOW A #TRICK	1010608	5	I
I #KNOW SINK 560 ? WHO 151 MY MUM #TALKED TO H	1510247	1	I
I #KNOW SOMEONE (#IS) (IN) THE +KITCHEN WITH DINAH	1000002	21	I
I #KNOW THAT 132 ? WHAT #ARE YOU #DOING 102 #C	1020499	58	I
I #KNOW THE +STORY ON THIS 190 ? WHAT /HMM 101	1010346	5	I
I #KNOW THE +STORY ON THIS 100 #GO +OUTSIDE 10	1010346	56	I
I #KNOW THIS ONE IT ##GOES LIKE THIS AND YOU #TU	1500195	1	I
I #KNOW TOO 501 ? WHAT 103 STUPID 502 YES @Y	1030125	1	I
I #KNOW TWO MILLION +DOLLAR-SP 501 TWO MILLION	1030211	47	I
I #KNOW WHAT #GOES THAT#S GOOD NO NO THAT#	1530179	29	I
I #KNOW WHAT IT IS 501 IT#S ONE OF THE EASTER +E	1030437	9	I

SPOKEN 2 TIMES BY 2 SPEAKERS

SPEAKER 103 - FREQUENCY 1

SPEAKER 153 - FREQUENCY 1

I #KNOW WHEN I #WAS HEY +STUPID 889 EH #DONT YOU	1020460	49	I
I #KNOW WHERE THE FAT +ONE #IS 132 ? WHERE 102	1020229	24	I
I #KNOW YOU #DONT #KNOW @DONNO 502 ONE THOUSAND	1030142	44	I
I #KNOW YOU#RE #LOOKING #DONT #LOOK #DONT #LOO	1020570	24	I

SPOKEN 2 TIMES BY 2 SPEAKERS

SPEAKER 102 - FREQUENCY 1

SPEAKER 103 - FREQUENCY 1

SPOKEN 40 TIMES BY 8 SPEAKERS

APPENDIX 7

INTERVIEW SCHEDULE FOR
BACKGROUND INFORMATION
(MIGRANT SAMPLE)

MOUNT LAWLEY COLLEGE OF ADVANCED EDUCATION

DEPARTMENT OF ENGLISH SPEECH & DRAMA

LANGUAGE RESEARCH PROJECT

BACKGROUND INFORMATION REQUIRED

1. Name and address of subject.
2. Sex, exact age and first language of the child (dialect, if applicable).
3. Length of residence in Australia:-
 - a) for child
 - b) for any other members of family, if different.
4. Is the family living in a very close knit community (e.g. hostel or other close grouping), an integrated community, or completely isolated as regards their own ethnic community.
5. a) Does the family have access to T.V.? Is the subject spending much time watching programmes?
 b) Are there English books/magazines/papers available in the home?
6. Employment of parents. If both employed then who cares for the children during period both parents are at work and what is the language situation? To what degree do parents need to speak or/and understand English in their employment?
7. Details of siblings - sexes, ages, occupations.
8. Standard of English of parents and siblings. Have the parents attended English classes of any sort either before or since coming to Australia? Did the child learn any English before coming to Australia?

 (For standard of English of parents and siblings it is suggested that a 4-point scale be applied:-
 1. Presystematic: random words or phrases known, but no systematic knowledge of generalisations, structures.
 2. Systematic: Some knowledge of structures and generalisations learned.
 3. Fluent-one: Good grasp of the structures of English but limited vocabulary. Still at concrete level.
 4. Fluent-two: Freer use of language and approaching first language competence.)
9. Standard of education of parents in native country?
10. Were there other relatives in Australia before the parents came?
11. Has the child attended nursery, kindergarten or day-care centre before starting to attend school? Which language medium?
12. Is the child attending classes to learn any other language?

13. The attitude of the parents towards integrating with Australian society or remaining largely within their own family or language group? Is the family fairly mobile or static within the one district? Answers should be impressionistic. Some pointers towards parents' desire (or otherwise) for integration would be gained from participation by parents or children in ethnic or integrated clubs/activities, ethnic language learning, attendance at English classes, etc., and the interrelationship of these. Some impression could also be gained of the amount of travelling done by the family over the previous few months or during school holidays, their social habits, whether they have a 'host family', etc.
14. Religious influence/background.
15. Playmates of subject, migrant language speakers or English speakers?

ALASTAIR MCGREGOR

APPENDIX 8

INTERVIEW SCHEDULE FOR
BACKGROUND INFORMATION
(ABORIGINAL SAMPLE)

MOUNT LAWLEY COLLEGE OF ADVANCED EDUCATIONLANGUAGE DEVELOPMENT RESEARCH PROJECTFIELD PROJECT WILUNA 1979

Research Pair: _____

Subject's Name: _____ Code Number: _____

Subject's Age: _____ Date of Birth: _____

Home Language:(Tick Correct One)

Katutjara ☐Mantjiltjara ☐Ngaanjatjara ☐Pitjantjatjara ☐Pintupi ☐

Comment: _____

Place of birth of subject: _____

Subject: _____ has been to (be careful of leading questions):

Meekatharra ☐Jigalong ☐Strelley ☐Yandeyarra ☐Hedland ☐Cue ☐Leonora ☐Kalgoorlie ☐Cosmo ☐Warburton ☐Alice Springs ☐Cundeelee ☐Other: ☐

Comment: _____

EDUCATION

When did subject start school? _____

Has subject been to another school? _____

If yes, give details: _____

Has subject been to Pre-school? _____

If yes, give details: _____

SIBLINGS

Indicate age and sex of siblings by marking ages - brothers: circle number
sisters: cross number

1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17,
18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30.

Comment: _____

Subject's home and environment: _____

Comment: _____

Local economy and social conditions: (Please use the back of this page if
necessary)

FATHER

Age (approx.): _____

Linguistic Group: (Tick one)Katutjara ☐Mantjiltjara ☐Ngaanjatjara ☐Pitjantjatjara ☐Pintupi ☐

Home (tribal) country: _____

EDUCATION

Did father go to school as a child? _____

If so, where? _____

If so, for how many years? _____

EMPLOYMENT

Present employment: _____

Past employment: _____

Pensioner: _____

Unemployed: _____

Comment: _____

MOTHER

Age (approx.): _____

Linguistic Group: (Tick One)Katutjara ☐Mantjiltjara ☐Ngaanjatjara ☐Pitjantjatjara ☐Pintupi ☐

Home (tribal) country: _____

EDUCATION

Did mother go to school as a child? _____

If so, where? _____

If so, for how many years? _____

EMPLOYMENT

Present employment: _____

Past employment: _____

Pensioner: _____

Unemployed: _____

Comment: _____

APPENDIX 9

ONE, TWO AND THREE WORD
SEQUENCES OF THE TOTAL SAMPLE
OF UPPER FIVE YEAR OLD
ABORIGINAL AND MIGRANT CHILDREN,
IN ORDER OF INDEX VALUE

EXPLANATORY NOTES

1. Full lists are also available of single words and sequences in alphabetic and raw frequency order. It is regretted that space forbids their inclusion in this report.
2. In order to provide important information on such matters as number of possessives, plurals, questions, etc. used, relevant non-alphabetic symbols are listed along with other words and morphemes on the following pages.
3. For details on the calculation of an Index Number for single words and two or three word sequences see pages 17 and 20 respectively. Index numbers for three word sequences should be multiplied by 100 to make them comparable with the index numbers used on the other lists.
4. For details on the interpretation of all symbols used in the Print out see Appendix 4: Instructions for Coding.

SINGLE WORDS USED IN THE TOTAL SAMPLE OF ABORIGINAL
AND MIGRANT CHILDREN IN ORDER OF INDEX VALUE

SEQUENCE	FREQ	SPEAKERS	INDEX
I	1402	26	489.2
?	1197	25	401.6
THE	1036	25	347.6
ONE	933	26	325.6
YOU	957	25	321.1
A	767	24	247.0
NO	655	24	211.0
YES	617	24	198.7
TWO	619	23	191.1
AND	556	25	186.5
IT	558	24	179.7
S	549	23	169.5
-SP	502	24	161.7
HERE	456	26	159.1
THIS	478	24	154.0
WHAT	498	22	147.0
THREE	448	23	138.3
IS	434	23	134.0
THAT	426	22	125.8
LOOK	339	26	118.3
*	545	16	117.0
TO	331	23	102.2
FOUR	336	22	99.2
OH	287	24	92.4
ME	279	24	89.9
CAN	299	22	88.3
COME	280	22	82.7
SEE	256	23	79.0
ON	265	22	78.2
IN	240	24	77.3
FIVE	287	20	77.0
M	235	24	75.7
UP	254	22	75.0
GOT	226	24	72.8
LITTLE	221	24	71.2
HE	222	23	68.5
MY	233	21	65.7
DONT	222	22	65.5
SIX	225	19	57.4
THERE	203	19	51.8
NOT	166	23	51.2
KNOW	169	20	45.4
GO	153	21	43.1
ARE	152	21	42.8
HAVE	144	22	42.5
WHERE	126	23	38.9
AT	120	24	38.7
FOR	136	21	38.3
DO	148	18	35.8
SEVEN	149	17	34.0

SEQUENCE	FREQ	SPEAKERS	INDEX
WANT	118	21	33.3
YOUR	112	21	31.6
WE	122	19	31.1
GOING	103	21	29.0
AM	103	20	27.6
DOWN	110	18	26.6
PLAY	104	19	26.5
BOY	103	19	26.3
COLOUR	92	20	24.7
NOW	108	17	24.6
DOG	100	18	24.2
GET	114	15	22.9
THEY	90	19	22.9
PUT	99	17	22.6
LIKE	88	19	22.4
ALL	80	20	21.5
EIGHT	109	14	20.5
RE	85	17	19.4
WITH	75	19	19.1
LL	99	14	18.6
OKAY	103	13	18.0
RUN	98	13	17.1
GOOD	74	17	16.9
CAT	68	18	16.4
HOUSE	76	16	16.3
HELLO	79	15	15.9
BIG	59	20	15.8
HEY	64	18	15.5
MINE	76	15	15.3
OF	63	18	15.2
SAY	81	14	15.2
YELLOW	71	16	15.2
OUT	63	17	14.4
BLACK	59	18	14.3
VE	69	15	13.9
GIRL	64	16	13.7
HOW	73	14	13.7
TEN	69	14	13.0
NINE	79	12	12.7
--SS	58	16	12.5
CANT	61	15	12.3
SOME	54	17	12.3
DID	60	15	12.1
FINISHED	56	16	12.0
BE	52	17	11.9
MORE	53	16	11.4
MAN	59	14	11.1
TALK	68	12	11.0
AH	58	14	10.9
CAR	58	14	10.9

SEQUENCE	FREQ	SPEAKERS	INDEX
CAR	58	14	10.9
MAKE	60	13	10.5
TREE	56	14	10.5
WATER	59	13	10.3
WILL	59	13	10.3
HORSE	63	12	10.1
GIVE	61	12	9.8
AEROPLANE	45	16	9.7
RED	40	18	9.7
BALL	65	11	9.6
HIM	51	14	9.6
YEAH	52	13	9.1
AWAY	49	13	8.5
DUCK	48	13	8.4
PLAYING	38	16	8.2
TAKE	36	17	8.2
KITTEN	75	8	8.1
MRS	55	11	8.1
WAS	60	10	8.1
BACK	36	16	7.7
JUST	48	12	7.7
GREEN	43	13	7.5
OFF	43	13	7.5
SHE	51	11	7.5
MISS	50	11	7.4
SITTING	50	11	7.4
BOAT	49	11	7.2
VERY	54	10	7.2
TOO	40	13	7.0
GONNA	51	10	6.8
THANK	39	13	6.8
WHO	42	12	6.8
EAT	45	11	6.6
PLEASE	49	10	6.6
DOOR	40	12	6.4
LOOKING	34	14	6.4
RIGHT	36	13	6.3
AN	33	14	6.2
ELEPHANT	41	11	6.1
TEACHER	35	13	6.1
THEM	37	12	6.0
RABB T	44	10	5.9
HIS	33	13	5.8
BUT	38	11	5.6
OVER	35	12	5.6
WAY	30	14	5.6
WHEN	46	9	5.6
EATING	45	9	5.4
GROUND	45	9	5.4
FISH	35	11	5.2

SEQUENCE	FREQ	SPEAKERS	INDEX
PIG	39	10	5.2
TELL	30	13	5.2
TIME	30	13	5.2
WELL	32	12	5.2
BOOK	31	12	5.0
SAID	37	10	5.0
FLOWER	33	11	4.9
BLUE	30	12	4.8
HIT	40	9	4.8
INSIDE	36	10	4.8
BECAUSE	50	7	4.7
WENT	32	11	4.7
ORANGE	31	11	4.6
YOURS	34	10	4.6
DOING	26	13	4.5
JUMP	42	8	4.5
NOTHING	28	12	4.5
HER	30	11	4.4
HOME	25	13	4.4
HAD	29	11	4.3
THESE	29	11	4.3
QUICK	38	8	4.1
BROWN	30	10	4.0
DIDNT	27	11	4.0
NAME	27	11	4.0
SO	33	9	4.0
THING	27	11	4.0
WANNA	37	8	4.0
IF	32	9	3.9
LET	32	9	3.9
COMING	28	10	3.8
TURN	28	10	3.8
WHY	35	8	3.8
CATCH	31	9	3.7
FIRST	34	8	3.7
MANY	31	9	3.7
WHITE	25	11	3.7
AGAIN	30	9	3.6
DRAW	38	7	3.6
ANOTHER	24	11	3.5
HUH	43	6	3.5
LEAVE	26	10	3.5
MOUSE	33	8	3.5
OUR	24	11	3.5
THEN	29	9	3.5
EGG	36	7	3.4
GOTTA	35	7	3.3
UM	41	6	3.3
OPEN	30	8	3.2
AFTER	23	10	3.1

SEQUENCE	FREQ	SPEAKERS	INDEX
AFTER	23	10	3.1
FAT	33	7	3.1
HAS	21	11	3.1
MORNING	33	7	3.1
NEVER	29	8	3.1
SUN	26	9	3.1
RUNNING	20	11	3.0
APPLE	31	7	2.9
BIRD	36	6	2.9
JANET	43	5	2.9
JOHN	43	5	2.9
THOSE	24	9	2.9
HA	23	9	2.8
SHUT	23	9	2.8
BEAUTIFUL	34	6	2.7
CHAIR	20	10	2.7
HAND	20	10	2.7
MUM	25	8	2.7
EH	28	7	2.6
LION	24	8	2.6
KANGAROO	27	7	2.5
AROUND	16	11	2.4
HAVENT	20	9	2.4
STANDING	22	8	2.4
UH	36	5	2.4
BABY	34	5	2.3
SNAKE	25	7	2.3
ANY	18	9	2.2
CDN	23	7	2.2
QNAME	18	9	2.2
WATCH	18	9	2.2
ISNT	17	9	2.1
LEG	17	9	2.1
PAPER	16	10	2.1
STOP	20	8	2.1
TOILET	16	10	2.1
YUK	26	6	2.1
CRAYON	19	8	2.0
HURRY	30	5	2.0
QUICKLY	19	8	2.0
SCHOOL	21	7	2.0
HAT	23	6	1.9
MUCH	20	7	1.9
GOES	19	7	1.8
OR	19	7	1.8
OUTSIDE	19	7	1.8
THINK	15	9	1.8
FINISH	16	8	1.7
FRIEND	18	7	1.7
HEAD	14	9	1.7

SEQUENCE	FREQ	SPEAKERS	INDEX
MOTHER	21	6	1.7
NAUGHTY	18	7	1.7
OTHER	18	7	1.7
POTATO	62	2	1.7
SAT	32	4	1.7
WHICH	21	6	1.7
WINDOW	21	6	1.7
ABOUT	14	8	1.5
NICE	19	6	1.5
ONLY	14	8	1.5
TAIL	16	7	1.5
BAG	13	8	1.4
BEAR	13	8	1.4
BEEN	15	7	1.4
BROTHER	27	4	1.4
COULD	12	9	1.4
FLYING	15	7	1.4
GONE	12	9	1.4
HOLD	15	7	1.4
KING	15	7	1.4
LETS	21	5	1.4
MAKING	17	6	1.4
MONKEY	17	6	1.4
PINK	26	4	1.4
SAW	18	6	1.4
STAR	17	6	1.4
BEE	16	6	1.3
BIT	14	7	1.3
CAME	14	7	1.3
CUT	16	6	1.3
DOES	12	8	1.3
ELEVEN	16	6	1.3
FROM	16	6	1.3
GETTING	11	9	1.3
LONG	12	8	1.3
SHOE	24	4	1.3
SIDE	14	7	1.3
WERE	16	6	1.3
BALLOON	22	4	1.2
EXCUSE	15	6	1.2
FUNNY	13	7	1.2
HAIR	11	8	1.2
SHOULD	13	7	1.2
TALKING	13	7	1.2
WALKING	13	7	1.2
BETTER	14	6	1.1
BICYCLE	16	5	1.1
BIN	12	7	1.1
CARE	17	5	1.1
DAY	20	4	1.1

SEQUENCE	FREQ	SPEAKERS	INDEX
DAY	20	4	1.1
FULL	12	7	1.1
GUN	14	6	1.1
KEEP	12	7	1.1
KICK	21	4	1.1
KOALA	21	4	1.1
ON	14	6	1.1
RIDE	17	5	1.1
SHOP	16	5	1.1
SORRY	12	7	1.1
TWELVE	12	7	1.1
WAIT	16	5	1.1
BOX	11	7	1.0
CAUGHT	15	5	1.0
DONE	11	7	1.0
FOURTEEN	13	6	1.0
GAVE	11	7	1.0
HEAR	11	7	1.0
LAST	12	6	1.0
MADE	13	6	1.0
NEFD	13	6	1.0
PUSH	13	6	1.0
SOMETHING	13	6	1.0
STILL	13	6	1.0
STORY	12	6	1.0
THROW	12	6	1.0
TOOK	11	7	1.0
WALL	15	5	1.0
BETTY	22	3	0.9
DIGGER	34	2	0.9
DOO	22	3	0.9
DUMPTY	22	3	0.9
FOX	14	5	0.9
FUCK	16	4	0.9
GRASS	16	4	0.9
HAMMER	32	2	0.9
HUMPTY	23	3	0.9
JENNIFER	11	6	0.9
NOSE	11	6	0.9
OOH	11	6	0.9
REALLY	13	5	0.9
SAYS	16	4	0.9
SECOND	13	5	0.9
TOUCH	11	6	0.9
US	11	6	0.9
YET	8	8	0.9
BAA	15	4	0.8
CARROT	12	5	0.8
CHICKEN	12	5	0.8
CLOCK	21	3	0.8

SEQUENCE	FREQ	SPEAKERS	INDEX
COLD	9	7	0.8
ER	15	4	0.8
EVERYBODY	12	5	0.8
FLOOR	14	4	0.8
HARD	10	6	0.8
HOT	10	6	0.8
HULLO	30	2	0.8
NUMBER	10	6	0.8
OCLOCK	14	4	0.8
OLD	12	5	0.8
PENCIL	14	4	0.8
SHOW	9	7	0.8
SISTER	15	4	0.8
SIT	10	6	0.8
STAY	9	7	0.8
TOP	9	7	0.8
BED	9	6	0.7
BREAK	11	5	0.7
BROKEN	10	5	0.7
BUS	10	5	0.7
CHOCOLATE	11	5	0.7
COLOURING	11	5	0.7
DUMB	11	5	0.7
FALL	11	5	0.7
FATHER	13	4	0.7
FIRE	11	5	0.7
GOODBYE	11	5	0.7
GUESS	13	4	0.7
HELP	18	3	0.7
KID	17	3	0.7
LADY	11	5	0.7
MATE	13	4	0.7
MM	9	6	0.7
MOVE	9	6	0.7
PARK	9	6	0.7
READY	10	5	0.7
SH	11	5	0.7
SWING	10	5	0.7
TELLING	11	5	0.7
TRY	7	7	0.7
USE	9	6	0.7
WATT	26	2	0.7
ALWAYS	14	3	0.6
BELL	15	3	0.6
BY	8	6	0.6
CARD	15	3	0.6
CLASS	9	5	0.6
DAD	11	4	0.6
DOESNT	8	6	0.6
DOLL	8	6	0.6

DOLL	8	6	0.6
EAR	7	6	0.6
EYE	11	4	0.6
FIX	7	6	0.6
FRDG	12	4	0.6
GLUE	15	3	0.6
HICKORY	14	3	0.6
LA	24	2	0.6
LADDER	11	4	0.6
LOTS	7	6	0.6
MAMMA	24	2	0.6
OHH	16	3	0.6
SIR	12	4	0.6
STICK	7	6	0.6
THAN	9	5	0.6
THANKYOU	11	4	0.6
TWINKLE	22	2	0.6
WOOL	9	5	0.6
ALLOWED	8	5	0.5
ANYTHING	6	6	0.5
CAKE	8	5	0.5
CALL	7	5	0.5
DICKORY	13	3	0.5
DRINK	8	5	0.5
EASTER	12	3	0.5
EVEN	7	5	0.5
FALLING	7	5	0.5
FIND	8	5	0.5
FLY	9	4	0.5
HO	9	4	0.5
HUNG	20	2	0.5
HUNGRY	8	5	0.5
LOVE	13	3	0.5
LUNCH	10	4	0.5
MEN	10	4	0.5
MIGHT	6	6	0.5
MILK	9	4	0.5
MONEY	10	4	0.5
NEXT	8	5	0.5
PAM	20	2	0.5
PAT	20	2	0.5
PEN	8	5	0.5
PICK	9	4	0.5
PUTTING	8	5	0.5
RAT	17	2	0.5
SAME	10	4	0.5
SCISSORS	10	4	0.5
STUPID	8	5	0.5
TABLE	9	4	0.5
THIRTY	7	5	0.5

TODAY	9	4	0.5
TOGETHER	6	6	0.5
TONY	13	3	0.5
TRUCK	9	4	0.5
TWENTY	10	4	0.5
WASH	9	4	0.5
WEIR	17	2	0.5
WOULD	8	5	0.5
WRONG	7	5	0.5
ANN	10	3	0.4
BANANA	9	3	0.4
BRENDON	7	4	0.4
BUTTERFLY	7	4	0.4
DAVID	9	3	0.4
DESK	9	3	0.4
DILLIS	15	2	0.4
DONKEY	8	4	0.4
DORA	14	2	0.4
DRAWING	10	3	0.4
EASY	8	4	0.4
EMU	9	3	0.4
GOAT	7	4	0.4
GRADE	10	3	0.4
HAPPENED	8	4	0.4
HIGH	7	4	0.4
HONG	10	3	0.4
JUMPED	7	4	0.4
LINE	10	3	0.4
MAMA	10	3	0.4
MEAN	6	5	0.4
OK	16	2	0.4
PEOPLE	7	4	0.4
READ	7	4	0.4
READING	8	4	0.4
ROUND	8	4	0.4
SAM	15	2	0.4
SHA	9	3	0.4
SHANE	14	2	0.4
SHEEP	6	5	0.4
SHIRT	7	4	0.4
SHOT	9	3	0.4
SIXTEEN	7	4	0.4
SLEEPING	7	4	0.4
SWAN	7	4	0.4
TA	9	3	0.4
THIRD	9	3	0.4
TIGER	7	4	0.4
TRAIN	6	5	0.4
TURTLE	7	4	0.4
TYRE	8	4	0.4

SEQUENCE	FREQ	SPEAKERS	INDEX
TYRE	8	4	0.4
WALKIE	10	3	0.4
WORK	6	5	0.4
BAND	7	3	0.3
BAR	5	4	0.3
BISCUIT	7	3	0.3
COMES	6	4	0.3
COUSIN	5	4	0.3
COWBOY	7	3	0.3
DICK	12	2	0.3
DIRTY	6	4	0.3
DOCK	8	3	0.3
DRAGAN	10	2	0.3
DRAGON	7	4	0.3
DRESS	5	4	0.3
ENGLISH	7	3	0.3
FAST	7	3	0.3
FOOT	7	3	0.3
FUCKING	7	3	0.3
GAME	5	4	0.3
GREAT	8	3	0.3
GREY	5	5	0.3
HAVING	6	4	0.3
HELICOPTER	6	4	0.3
HOLDING	6	4	0.3
ICECREAM	6	4	0.3
INTO	5	4	0.3
LAMB	12	2	0.3
LE	13	2	0.3
LISTEN	5	4	0.3
LIVE	7	3	0.3
LIZARD	11	2	0.3
LOST	8	3	0.3
MARY	7	3	0.3
MAT	12	2	0.3
MATTER	5	4	0.3
NEARLY	5	4	0.3
NIACONI	13	2	0.3
OUCH	7	3	0.3
PICTURE	5	4	0.3
PUPPY	10	2	0.3
PUSSYCAT	8	3	0.3
RAN	13	2	0.3
RIDING	5	4	0.3
ROBERT	10	2	0.3
ROOM	12	2	0.3
SAND	8	3	0.3
SHAI	11	2	0.3
SHELL	7	3	0.3
SHUTTING	6	4	0.3

SING	6	4	0.3
SOMEONE	7	3	0.3
SONG	6	4	0.3
SPEAK	6	4	0.3
START	6	4	0.3
STOOD	7	3	0.3
SWIMMING	6	4	0.3
T	8	3	0.3
THIRTEEN	7	3	0.3
THUMBKIN	8	3	0.3
TOMORROW	5	4	0.3
TONGUE	8	3	0.3
TRACTOR	5	4	0.3
UNCLE	10	2	0.3
UNDER	6	4	0.3
YA	8	3	0.3
ZIP	10	2	0.3
ALRIGHT	5	3	0.2
ARMSTRONG	8	2	0.2
B	7	2	0.2
BA	6	2	0.2
BASH	6	2	0.2
BEHIND	5	3	0.2
BOBBY	8	2	0.2
BOING	8	2	0.2
BREAD	5	3	0.2
BRETT	9	2	0.2
BUCKET	6	2	0.2
BUM	5	3	0.2
BYE	6	3	0.2
CATCHING	6	3	0.2
CHANGE	9	2	0.2
CHARLIE	5	3	0.2
CHASEY	8	2	0.2
CHURCH	5	3	0.2
CRACK	5	3	0.2
DESERT	5	3	0.2
EIGHTEEN	6	2	0.2
FARM	5	3	0.2
FENCE	5	3	0.2
FIFTEEN	6	3	0.2
FINGER	8	2	0.2
FOOTBALL	7	2	0.2
GIMME	5	3	0.2
HAH	5	3	0.2
HEN	6	3	0.2
HILL	6	2	0.2
HUNDRED	6	2	0.2
JUMPING	5	3	0.2
KISS	9	2	0.2

SEQUENCE	FREQ	SPEAKERS	INDEX
KISS	9	2	0.2
KONG	7	2	0.2
LEANNE	6	2	0.2
LIM	6	2	0.2
LING	6	2	0.2
MR	6	3	0.2
NA	5	3	0.2
NIGHT	5	3	0.2
NINETEEN	6	3	0.2
ONCE	7	2	0.2
OWN	5	3	0.2
PETER	6	2	0.2
PLAYED	6	3	0.2
PURPLE	7	2	0.2
QUEEN	5	2	0.2
QUIET	7	2	0.2
QVERB	5	3	0.2
ROW	5	3	0.2
SCRIBBLED	6	2	0.2
SEVENTEEN	6	3	0.2
SLEEP	5	3	0.2
STAMP	7	2	0.2
STOPPED	6	2	0.2
SUPPOSED	6	2	0.2
TALLMAN	9	2	0.2
TEAPOT	6	2	0.2
TEDDYBEAR	6	2	0.2
TEETH	6	2	0.2
TIN	6	2	0.2
TOY	5	3	0.2
TRYING	6	3	0.2
TUESDAY	7	2	0.2
TURNIP	6	3	0.2
NET	6	2	0.2
WOMAN	6	2	0.2
BASKET	5	2	0.1
BUST	5	2	0.1
CALLING	5	2	0.1
DEIDRE	5	3	0.1
ELVIS	5	2	0.1
FEATHER	5	2	0.1
FRIDAY	5	2	0.1
GILLIGAN	5	2	0.1
KAY	5	2	0.1
KIM	5	2	0.1
LUN	5	2	0.1
MARK	5	2	0.1
NIP	5	2	0.1
O	5	2	0.1
PAP,	5	2	0.1
PART	5	2	0.1
SPIDER	5	2	0.1
TIE	5	2	0.1
WEEK	5	2	0.1

TWO WORD SEQUENCES OF THE TOTAL SAMPLE OF ABORIGINAL AND
MIGRANT CHILDREN IN ORDER OF INDEX VALUE

SEQUENCE	FREQ	SPEAKERS	INDEX
? WHAT	380	21	430.1
I M	232	24	365.6
ONE TWO	368	18	289.4
I CAN	121	19	150.9
I AM	100	20	131.3
TWO THREE	281	18	129.7
? WHERE	95	19	97.3
I DONT	79	18	93.4
THREE FOUR	222	20	82.4
I LL	77	13	65.7
I KNOW	66	14	60.7
WHAT S	175	17	58.7
IT S	114	21	57.7
I WANT	50	17	55.8
I HAVE	48	15	47.3
FOUR FIVE	190	16	40.5
I VE	45	13	38.4
? HOW	54	13	37.8
THE LITTLE	61	12	34.1
? CAN	40	15	32.3
THAT S	104	16	28.1
? YOU	35	14	26.4
THIS ONE	80	16	26.4
NO NO	68	13	25.0
LOOK AT	78	19	23.5
S THIS	72	14	22.9
YOU CAN	34	15	22.0
I LIKE	22	15	21.7
YOU RE	40	12	20.7
FIVE SIX	132	15	20.5
S A	54	16	19.6
IN THE	97	19	19.1
THE DOOR	34	12	19.0
THIS IS	58	15	18.0
A LITTLE	41	12	16.3
I CANT	23	10	15.1
SEE THE	109	13	15.0
I GOT	25	9	14.8
THAT ONE	63	13	13.8
ON THE	87	15	13.7
I SEE	29	7	13.3
YOU KNOW	34	9	13.2
? DO	23	10	12.4
YOU HAVE	23	12	11.9
AND A	39	12	11.7
S THAT	49	10	11.1
THE DOG	20	11	10.3
NO I	29	12	9.9
? WHY	25	7	9.4
? IS	21	8	9.1

SEQUENCE	FREQ	SPEAKERS	INDEX
AND I	30	12	9.0
COME ON	60	13	8.7
YOU ARE	22	9	8.5
? WHO	22	7	8.3
AND THE	37	9	8.3
DONT KNOW	59	16	8.3
SIX SEVEN	89	12	8.2
YOU WANT	23	8	7.9
THE BALL	28	6	7.8
YOU SAY	30	6	7.8
ONE MORE	16	11	7.7
? I	14	10	7.5
THE WATER	23	7	7.5
I WAS	28	4	7.4
I WILL	18	6	7.1
THE BOY	19	8	7.1
THE GIRL	19	8	7.1
I DIDNT	13	8	6.8
ONE FOR	25	6	6.6
YOU GOT	17	9	6.6
? ARE	13	9	6.3
IS IT	35	10	6.3
THE TREE	17	8	6.3
S NOT	24	11	6.0
-SP AND	30	9	5.9
A BIG	16	11	5.8
IS A	27	12	5.8
ONE HERE	22	6	5.8
ARE YOU	60	16	5.5
HE S	40	15	5.5
I HAVENT	12	7	5.5
TWO ONE	36	6	5.5
AND SEE	31	7	5.4
YOU GO	14	9	5.4
CAN SEE	34	13	5.2
I DID	11	7	5.1
WHAT IS	32	8	5.1
AND ONE	20	10	5.0
OH NO	40	10	5.0
THE CAT	18	6	5.0
? SEE	13	7	4.9
? WHICH	18	5	4.9
I WANNA	15	5	4.9
YOU DO	14	8	4.8
GOT A	37	13	4.7
I JUST	9	8	4.7
I NEVER	18	4	4.7
YES I	17	10	4.5
YOU DONT	15	7	4.5
HERE I	34	6	4.4

SEQUENCE	FREQ	SPEAKERS	INDEX
HERE I	34	6	4.4
S THE	24	8	4.4
I GO	11	6	4.3
YOU SEE	11	9	4.3
? DID	11	7	4.2
A MAN	21	6	4.2
COME HERE	29	13	4.2
* YES	33	8	4.1
* ONE	36	7	4.0
A GIRL	17	7	3.9
CAN I	30	11	3.9
I SAW	15	4	3.9
IT IS	18	9	3.9
SEVEN EIGHT	70	12	3.8
ONE ONE	14	6	3.7
WANT TO	55	15	3.7
? THIS	11	6	3.6
YOU VE	13	6	3.4
YOU DID	11	7	3.3
YOU GOING	11	7	3.3
YOU PUT	11	7	3.3
A DOG	12	8	3.2
WHAT DO	23	7	3.2
CAN YOU	24	11	3.1
DO YOU	50	13	3.1
GOING TO	61	13	3.1
IS THE	17	10	3.1
IT IN	13	10	3.1
THE BIN	11	6	3.1
THE OTHER	11	6	3.1
M GOING	37	8	3.0
NO YOU	15	7	3.0
THE FAT	13	5	3.0
HERE HERE	17	8	2.9
THE CHAIR	9	7	2.9
THE QNOUN	15	4	2.8
THE TEACHER	12	5	2.8
YES YES	15	7	2.8
YOU LIKE	11	6	2.8
M NOT	29	9	2.7
A COW	11	7	2.6
A DUCK	13	6	2.6
COME AND	39	6	2.6
YOU GET	10	6	2.6
AND DOWN	25	4	2.5
LOOK UP	26	6	2.5
UP THERE	28	9	2.5
* NO	25	6	2.4
? A	9	5	2.4
I DO	6	6	2.4

SEQUENCE	FREQ	SPEAKERS	INDEX
THE HOUSE	13	4	2.4
THE KITTEN	13	4	2.4
THIS WAY	13	9	2.4
? HAVE	7	6	2.3
A PIG	14	5	2.3
I SAID	7	5	2.3
NO IT	10	8	2.3
TO THE	15	11	2.3
? OH	8	5	2.2
A HOUSE	13	5	2.2
HERE S	13	8	2.2
I EAT	11	3	2.2
IN HERE	19	11	2.2
THE MAN	8	6	2.2
-SP ARE	12	8	2.1
LITTLE DOG	31	7	2.1
S MINE	13	7	2.1
S MY	13	7	2.1
S WHAT	13	7	2.1
THE BOAT	11	4	2.1
THE CAR	9	5	2.1
THE MOUSE	11	4	2.1
ONE OF	9	5	2.0
TWO TWO	13	6	2.0
A BOY	14	4	1.9
IT ON	11	7	1.9
THE HORSE	8	5	1.9
TO PLAY	17	8	1.9
YOU CANT	11	4	1.9
YOU DIDNT	9	5	1.9
A CAT	11	5	1.8
GOT TWO	23	8	1.8
I NEED	7	4	1.8
THE CLOCK	13	3	1.8
THE GRASS	13	3	1.8
UP AND	25	7	1.8
WHERE IS	35	10	1.8
WHERE S	27	13	1.8
YES A	11	6	1.8
YOU THIS	14	3	1.8
? HUH	8	4	1.7
AT THE	22	15	1.7
IS THIS	16	6	1.7
THE KING	9	4	1.7
THE THE	12	3	1.7
TWO KITTEN	17	4	1.7
YOU COME	8	5	1.7
? AND	6	5	1.6
? THAT	6	5	1.6
A GOOD	8	6	1.6

SEQUENCE	FREQ	SPEAKERS	INDEX
A GOOD	8	6	1.6
I HAD	6	4	1.6
THE TWO	7	5	1.6
YOU LL	9	4	1.6
YOU WANNA	12	3	1.6
YOU YOU	6	6	1.6
* I	16	6	1.5
? EH	9	3	1.5
A HORSE	11	4	1.5
EIGHT NINE	46	12	1.5
ONE THREE	7	5	1.5
\$ YOUR	8	8	1.5
THE BOX	8	4	1.5
THE SAME	8	4	1.5
UP THE	19	8	1.5
WHAT COLOUR	11	7	1.5
YOU DOING	7	5	1.5
IS MY	13	6	1.4
IT TO	10	6	1.4
SEE YOU	13	10	1.4
THE AEROPLANE	6	5	1.4
THE BIG	6	5	1.4
THE FLOOR	10	3	1.4
YOU SAID	8	4	1.4
? BUT	6	4	1.3
? ISNT	6	4	1.3
? IT	6	4	1.3
? WILL	6	4	1.3
A CAR	10	4	1.3
FOR ME	26	10	1.3
THE TIME	7	4	1.3
WHAT ARE	11	6	1.3
YES SIR	12	4	1.3
A BIRD	9	4	1.2
DO IT	23	11	1.2
HAVE TO	31	7	1.2
I PLAY	6	3	1.2
IT UP	10	5	1.2
LOOK LOOK	13	6	1.2
TO DO	14	6	1.2
? NO	5	4	1.1
? OKAY	5	4	1.1
? WANT	7	3	1.1
A BALLOON	11	3	1.1
A RABBIT	8	4	1.1
DONT WANT	15	8	1.1
GOT ONE	13	9	1.1
IS THAT	9	7	1.1
THE FLOWER	6	4	1.1
THE LAST	6	4	1.1

SEQUENCE	FREQ	SPEAKERS	INDEX
THE TOILET	6	4	1. 1
THIS HERE	9	6	1. 1
TO ME	13	6	1. 1
VE GOT	47	13	1. 1
? MISS	9	2	1. 0
A YELLOW	6	5	1. 0
AND TWO	8	5	1. 0
AND YOU	8	5	1. 0
GOT IT	13	8	1. 0
GOT TO	13	8	1. 0
HAVE A	17	10	1. 0
I COME	5	3	1. 0
I MADE	5	3	1. 0
LOOK HERE	11	6	1. 0
ON A	13	7	1. 0
THE ELEPHANT	7	3	1. 0
THE FIRST	7	3	1. 0
THIS FOR	8	6	1. 0
THREE THREE	9	6	1. 0
TO GET	10	7	1. 0
WHAT YOU	7	7	1. 0
YOU ANY	6	4	1. 0
YOU THINK	6	4	1. 0
A TREE	7	4	0. 9
A WINDOW	9	3	0. 9
AND PLAY	9	4	0. 9
AND THAT	6	6	0. 9
AND THIS	9	4	0. 9
CAN PLAY	15	5	0. 9
IS HERE	12	4	0. 9
M GONNA	21	4	0. 9
ME TOO	11	7	0. 9
ONE AND	5	4	0. 9
THERE S	16	8	0. 9
THIS MORNING	14	3	0. 9
THREE ONE	16	3	0. 9
TO GO	11	6	0. 9
YOU PLAYING	5	4	0. 9
* COME	16	3	0. 8
? DONT	7	2	0. 8
? SHALL	7	2	0. 8
? YEAH	5	3	0. 8
A BALL	6	4	0. 8
A STORY	6	4	0. 8
A WALL	8	3	0. 8
AND ME	6	5	0. 8
FOR YOU	21	7	0. 8
HAVE U	15	9	0. 8
IT OFF	7	5	0. 8
PUT IT	29	9	0. 8

SEQUENCE	FREQ	SPEAKERS	INDEX
PUT IT	29	9	0.8
TO BE	14	4	0.8
TO YOU	10	6	0.8
* AND	15	3	0.7
A GREEN	5	4	0.7
A LOOK	5	4	0.7
A RED	5	4	0.7
AND MY	7	4	0.7
CAN DO	11	5	0.7
CAN GO	11	5	0.7
FIVE ONE	17	4	0.7
GO TO	14	9	0.7
HERE YOU	7	5	0.7
IN A	9	8	0.7
IS FLYING	8	5	0.7
IS SITTING	10	4	0.7
IS STANDING	10	4	0.7
IT FOR	7	4	0.7
KNOW WHAT	12	9	0.7
LIKE THIS	25	9	0.7
LITTLE BOY	18	4	0.7
M A	11	6	0.7
NO THAT	5	5	0.7
OH OH	10	6	0.7
ON IT	9	7	0.7
ONE FOUR	5	3	0.7
ONE NOW	5	3	0.7
S GOING	8	4	0.7
S THERE	6	5	0.7
THE CARD	5	3	0.7
THE WAY	5	3	0.7
THEY RE	24	9	0.7
TWO FOUR	7	4	0.7
YOU GOTTA	8	2	0.7
YOU MAKE	8	2	0.7
* ?	9	4	0.6
* SEE	9	4	0.6
* THREE	8	5	0.6
? PLEASE	6	2	0.6
ALL THE	22	9	0.6
AND IT	5	5	0.6
DO THAT	15	8	0.6
IS IN	8	4	0.6
IT WENT	8	3	0.6
LIKE THAT	19	11	0.6
LOOK OUT	7	5	0.6
ME A	10	5	0.6
NO HE	5	4	0.6
OH LOOK	10	5	0.6
S HERE	5	5	0.6

SEQUENCE	FREQ	SPEAKERS	INDEX
THE BABY	6	2	0.6
THE HILL	6	2	0.6
THE KANGAROO	6	2	0.6
THE RED	6	2	0.6
WHERE ARE	14	8	0.6
YES YOU	6	4	0.6
YOU BECAUSE	5	3	0.6
? UH	5	2	0.5
A BABY	5	3	0.5
A GREAT	7	2	0.5
A NICE	5	3	0.5
A QNOUN	5	3	0.5
AND HE	5	4	0.5
AND JUMP	9	2	0.5
AND PINK	10	2	0.5
AND YELLOW	5	4	0.5
CAN RUN	10	4	0.5
COME IN	8	6	0.5
FOR THE	17	6	0.5
GIVE ME	35	11	0.5
GOT THE	7	7	0.5
HE IS	11	5	0.5
HERE WE	6	4	0.5
IN THERE	8	6	0.5
IT HERE	5	4	0.5
IT OUT	5	4	0.5
IT WAS	10	2	0.5
M TELLING	10	5	0.5
ME SOME	10	4	0.5
MY HOUSE	14	4	0.5
MY MUM	9	6	0.5
NINE TEN	29	10	0.5
NO ONE	6	3	0.5
OH DONT	7	6	0.5
ON YOU	8	6	0.5
S ALL	5	4	0.5
THANK YOU	39	13	0.5
THAT WE	6	5	0.5
THE BIRD	5	2	0.5
THE BUCKET	5	2	0.5
THE DUCK	5	2	0.5
TO MAKE	8	5	0.5
UP HERE	9	6	0.5
WE RE	16	8	0.5
WHAT HAPPENED	6	4	0.5
YOU TODAY	6	2	0.5
YOU W E	6	2	0.5
* A	7	4	0.4
* HUMPTY	12	2	0.4
* THE	6	4	0.4

SEQUENCE	FREQ	SPEAKERS	INDEX
PUT IT	29	9	0.8
TO BE	14	4	0.8
TO YOU	10	6	0.8
* AND	15	3	0.7
A GREEN	5	4	0.7
A LOOK	5	4	0.7
A RED	5	4	0.7
AND MY	7	4	0.7
CAN DO	11	5	0.7
CAN GO	11	5	0.7
FIVE ONE	17	4	0.7
GO TO	14	9	0.7
HERE YOU	7	5	0.7
IN A	9	8	0.7
IS FLYING	8	5	0.7
IS SITTING	10	4	0.7
IS STANDING	10	4	0.7
IT FOR	7	4	0.7
KNOW WHAT	12	9	0.7
LIKE THIS	25	9	0.7
LITTLE BOY	18	4	0.7
M A	11	6	0.7
NO THAT	5	5	0.7
OH OH	10	6	0.7
ON IT	9	7	0.7
ONE FOUR	5	3	0.7
ONE NOW	5	3	0.7
S GOING	8	4	0.7
S THERE	6	5	0.7
THE CARD	5	3	0.7
THE WAY	5	3	0.7
THEY RE	24	9	0.7
TWO FOUR	7	4	0.7
YOU GOTTA	8	2	0.7
YOU MAKE	8	2	0.7
* ?	9	4	0.6
* SEE	9	4	0.6
* THREE	8	5	0.6
? PLEASE	6	2	0.6
ALL THE	22	9	0.6
AND IT	5	5	0.6
DO THAT	15	8	0.6
IS IN	8	4	0.6
IT WENT	8	3	0.6
LIKE THAT	19	11	0.6
LOOK OUT	7	5	0.6
ME A	10	5	0.6
NO HE	5	4	0.6
OH LOOK	10	5	0.6
S HERE	5	5	0.6

SEQUENCE	FREQ	SPEAKERS	INDEX
THE BABY	6	2	0.6
THE HILL	6	2	0.6
THE KANGAROO	6	2	0.6
THE RED	6	2	0.6
WHERE ARE	14	8	0.6
YES YOU	6	4	0.6
YOU BECAUSE	5	3	0.6
? UH	5	2	0.5
A BABY	5	3	0.5
A GREAT	7	2	0.5
A NICE	5	3	0.5
A QNOUN	5	3	0.5
AND HE	5	4	0.5
AND JUMP	9	2	0.5
AND PINK	10	2	0.5
AND YELLOW	5	4	0.5
CAN RUN	10	4	0.5
COME IN	8	6	0.5
FOR THE	17	6	0.5
GIVE ME	35	11	0.5
GOT THE	7	7	0.5
HE IS	11	5	0.5
HERE WE	6	4	0.5
IN THERE	8	6	0.5
IT HERE	5	4	0.5
IT OUT	5	4	0.5
IT WAS	10	2	0.5
M TELLING	10	5	0.5
ME SOME	10	4	0.5
MY HOUSE	14	4	0.5
MY MUM	9	6	0.5
NINE TEN	29	10	0.5
NO ONE	6	3	0.5
OH DONT	7	6	0.5
ON YOU	8	6	0.5
S ALL	5	4	0.5
THANK YOU	39	13	0.5
THAT WE	6	5	0.5
THE BIRD	5	2	0.5
THE BUCKET	5	2	0.5
THE DUCK	5	2	0.5
TO MAKE	8	5	0.5
UP HERE	9	6	0.5
WE RE	16	8	0.5
WHAT HAPPENED	6	4	0.5
YOU TODAY	6	2	0.5
YOU W E	6	2	0.5
* A	7	4	0.4
* HUMPTY	12	2	0.4
* THE	6	4	0.4

SEQUENCE	FREQ	SPEAKERS	INDEX
* THE	6	4	0.4
* TWO	9	3	0.4
-SP IN	5	4	0.4
A COWBOY	6	2	0.4
AND AFTER	8	2	0.4
AND GET	5	3	0.4
CAN COME	8	4	0.4
COME COME	8	4	0.4
DONT YOU	12	4	0.4
FOR A	10	8	0.4
IS EATING	6	4	0.4
IS RUNNING	5	4	0.4
IS THUMBKIN	8	3	0.4
IS YOUR	5	5	0.4
LITTLE BIT	9	5	0.4
ME ME	10	3	0.4
NO NOT	5	3	0.4
OH I	7	5	0.4
OH YES	6	5	0.4
ON THERE	7	5	0.4
ONE KITTEN	5	2	0.4
ONE RABBIT	5	2	0.4
S BEAUTIFUL	6	3	0.4
SEE MY	9	4	0.4
SIX SIX	10	5	0.4
THAT THERE	11	2	0.4
THREE BAG	6	4	0.4
TO TALK	7	4	0.4
TO TELL	6	5	0.4
WHAT A	5	4	0.4
* DORA	8	2	0.3
* GOOD	7	3	0.3
* LOOK	8	2	0.3
-SP FOUR	5	3	0.3
-SP ON	5	3	0.3
A LINE	5	2	0.3
A PINK	5	2	0.3
A TEACHER	5	2	0.3
AT MINE	8	7	0.3
AT THAT	11	6	0.3
AT THIS	20	3	0.3
BOY -SP	12	8	0.3
CAN HAVE	7	4	0.3
CAN JUMP	9	3	0.3
DO THIS	12	5	0.3
DOWN HERE	12	8	0.3
FOUR ONE	12	2	0.3
GO HOME	9	6	0.3
GO ON	11	5	0.3
GOT ANY	7	4	0.3

SEQUENCE	FREQ	SPEAKERS	INDEX
HAVE IT	8	7	0.3
HAVE THIS	10	5	0.3
HE GOT	10	3	0.3
HE SAID	11	3	0.3
HERE IS	5	3	0.3
HOW ARE	24	6	0.3
IS ON	6	3	0.3
IS TALLMAN	9	2	0.3
IS VERY	5	3	0.3
KNOW I	7	6	0.3
LITTLE KITTEN	10	3	0.3
LITTLE ONE	7	4	0.3
LOOK AND	7	3	0.3
LOOK DOWN	6	3	0.3
ME SEE	9	3	0.3
MY DAD	10	3	0.3
NOT A	7	7	0.3
OH ONE	7	3	0.3
OH WHAT	5	5	0.3
OH YOU	5	5	0.3
ON MY	6	5	0.3
ON THIS	9	3	0.3
PLAY WITH	12	7	0.3
S EASY	5	3	0.3
S GOT	5	3	0.3
S INSIDE	5	3	0.3
SITTING ON	29	9	0.3
THAT LITTLE	5	3	0.3
THAT THAT	6	3	0.3
THAT THING	5	4	0.3
THAT UP	5	4	0.3
THEY ARE	12	7	0.3
THIS ON	5	3	0.3
THREE BOAT	8	2	0.3
TO DRAW	12	2	0.3
TO EAT	10	2	0.3
TO HAVE	8	3	0.3
WANT A	11	7	0.3
WHAT DID	5	3	0.3
* HERE	5	3	0.2
* IN	5	2	0.2
* JOHN	6	2	0.2
* ON	5	3	0.2
* WATER	5	2	0.2
ALL RIGHT	17	4	0.2
CAN MAKE	7	3	0.2
CAN P'DE	7	3	0.2
COME LOOK	7	3	0.2
DID YOU	20	7	0.2
DOG RUN	13	5	0.2

SEQUENCE	FREQ	SPEAKERS	INDEX
	13	5	0.2
DOG RUN	7	4	0.2
DONT HAVE	7	3	0.2
DONT TOUCH	11	6	0.2
DOWN THERE	6	3	0.2
FIVE FIVE	10	7	0.2
GET OUT	7	4	0.2
GO AND	7	4	0.2
GO AWAY	8	4	0.2
GO DOWN	7	5	0.2
GO IN	10	4	0.2
GO UP	5	5	0.2
GOT AN	6	3	0.2
GOT THREE	9	3	0.2
HAVE SOME	6	4	0.2
HE CAN	7	3	0.2
IN MY	5	4	0.2
IN OUR	5	2	0.2
IS BIG	8	4	0.2
KNOW WHERE	13	2	0.2
LITTLE BOAT	6	3	0.2
LITTLE CAT	8	2	0.2
LITTLE GIRL	10	7	0.2
LL GO	7	2	0.2
LOOK JOHN	6	2	0.2
LOOK SEE	5	4	0.2
M GETTING	7	3	0.2
M MAKING	22	8	0.2
MAKE A	5	4	0.2
ME ONE	5	4	0.2
ME TO	9	2	0.2
MY ROOM	5	4	0.2
MY THING	12	2	0.2
MY TURN	7	4	0.2
NOT THAT	13	7	0.2
OF THE	5	4	0.2
OH THAT	27	3	0.2
RUN AWAY	5	4	0.2
SEE A	5	3	0.2
SEE TWO	12	2	0.2
SIX ONE	5	2	0.2
THAT FOR	6	4	0.2
THERE AND	5	2	0.2
THIS TIME	5	3	0.2
TO SCHOOL	5	4	0.2
UP TO	7	3	0.2
UP UP	11	5	0.2
WE ARE	6	2	0.2
WHAT IT	11	7	0.2
WITH THE	6	3	0.1
ALL OF			

SEQUENCE	FREQ	SPEAKERS	INDEX
AM HERE	8	2	0.1
AM PLAYING	5	4	0.1
ARE HERE	5	5	0.1
ARE THE	5	3	0.1
AT HIS	5	3	0.1
BIG ONE	9	4	0.1
BOY AND	5	4	0.1
BUT I	10	8	0.1
CAR -SP	11	4	0.1
COLOUR -SP	9	5	0.1
COLOUR COLOUR	6	4	0.1
COLOUR IN	7	4	0.1
COME BACK	6	2	0.1
COME SEE	5	2	0.1
COME UP	6	2	0.1
DOG -SP	7	5	0.1
DOG CAN	9	3	0.1
DONT CARE	5	3	0.1
DOWN ON	5	4	0.1
DOWN UP	16	2	0.1
GET A	7	6	0.1
GET DOWN	6	4	0.1
GET IT	7	4	0.1
GET ME	6	3	0.1
GET YOU	6	4	0.1
GIRL -SP	13	6	0.1
GOOD MORNING	15	2	0.1
GOT MY	5	3	0.1
HE HE	5	2	0.1
HE WAS	5	3	0.1
HELLO HELLO	9	4	0.1
HEY YOU	5	5	0.1
HOUSE -SP	8	4	0.1
HOUSE AND	7	4	0.1
HOW MUCH	15	5	0.1
KITTEN -SP	21	5	0.1
KNOW HOW	5	3	0.1
KNOW THAT	5	4	0.1
KNOW YOU	7	3	0.1
LEAVE IT	14	8	0.1
LIKE IT	6	4	0.1
LIKE THE	8	4	0.1
LL BE	6	4	0.1
LL DO	10	5	0.1
LL GET	9	5	0.1
LL TELL	7	4	0.1
LOOKING AT	17	9	0.1
M AFTER	5	2	0.1
M FINISHED	5	2	0.1
MAN AND	7	5	0.1

SEQUENCE	FREQ	SPEAKERS	INDEX
MAN AND	7	5	0.1
MRS WATT	26	2	0.1
MY CAR	6	2	0.1
MY EASTER	5	2	0.1
MY FATHER	5	2	0.1
MY HAND	5	3	0.1
MY MOTHER	5	3	0.1
MY QNOUN	5	3	0.1
NOT ALLOWED	5	3	0.1
NOT GOING	5	4	0.1
NOT GONNA	7	3	0.1
NOW I	6	4	0.1
NOW YOU	7	3	0.1
OKAY I	7	3	0.1
OUT OF	11	6	0.1
OVER THERE	18	8	0.1
PUT THE	7	3	0.1
PUT THEM	7	3	0.1
PUT THIS	8	4	0.1
PUT YOUR	5	4	0.1
RE A	6	4	0.1
RE NOT	10	5	0.1
RUN AND	8	4	0.1
SAY IT	8	5	0.1
SHE S	12	6	0.1
TALK TALK	12	3	0.1
TEN ELEVEN	10	5	0.1
VERY WELL	16	4	0.1
WANT IT	6	4	0.1
WANT ME	7	3	0.1
WE CAN	5	3	0.1
WE GO	7	3	0.1
WE HAVE	5	3	0.1
WE WENT	6	3	0.1
WILL YOU	11	7	0.1
WITH A	7	4	0.1
WITH ME	6	5	0.1
WITH THAT	5	4	0.1
WITH THIS	6	4	0.1
WITH YOU	7	6	0.1
YOUR HAND	5	4	0.1
AEROPLANE -SP	9	3	0.0
AN AEROPLANE	7	4	0.0
ANOTHER ONE	8	5	0.0
BAA BAA	9	4	0.0
BAG -SP	6	4	0.0
BALL -SP	7	3	0.0
BE BE	5	2	0.0
BE QUIET	5	2	0.0
BECAUSE I	10	5	0.0

SEQUENCE	FREQ	SPEAKERS	INDEX
BECAUSE WE	5	3	0.0
BECAUSE YOU	9	2	0.0
BEE -SP	8	4	0.0
BIRD -SP	13	3	0.0
BLACK AND	5	2	0.0
BLACK SHEEP	5	4	0.0
BOAT -SP	16	3	0.0
BOAT AND	5	2	0.0
BOING BOING	5	2	0.0
BOOK -SP	7	4	0.0
BOY IS	5	2	0.0
BREAK IT	7	4	0.0
CANT YOU	5	4	0.0
CAR IS	7	2	0.0
CARD -SP	12	3	0.0
CATCH ME	13	2	0.0
CATCH YOU	5	3	0.0
CAUGHT ME	6	4	0.0
COLOUR IT	5	3	0.0
CRAYON -SP	8	5	0.0
DICKORY DOCK	6	3	0.0
DID IT	5	5	0.0
DRAW A	8	4	0.0
DUMPTY HAD	7	2	0.0
DUMPTY SAT	9	2	0.0
EASTER EGG	7	2	0.0
EAT IT	10	2	0.0
EATING THE	15	4	0.0
EGG -SP	11	5	0.0
EIGHT TEN	5	2	0.0
ELEPHANT AND	8	6	0.0
ELEVEN TWELVE	8	4	0.0
EXCUSE ME	14	5	0.0
EYE -SP	7	3	0.0
FAT CAT	5	2	0.0
FEATHER -SP	5	2	0.0
FLOWER -SP	8	6	0.0
GET LOST	7	2	0.0
GET YOUR	5	2	0.0
GIVE IT	5	4	0.0
GONNA HAVE	6	3	0.0
GOOD ONE	5	3	0.0
GOTTA TALK	12	3	0.0
GREAT FALL	7	2	0.0
HA HA	9	5	0.0
HAD A	11	5	0.0
HAND -SP	12	7	0.0
HAVENI GOT	11	7	0.0
HICKORY DICKORY	12	3	0.0
HIT ME	5	3	0.0

SEQUENCE	FREQ	SPEAKERS	INDEX
HIT ME	5	3	0.0
HIT THE	5	2	0.0
HIT YOU	9	4	0.0
HOLD IT	5	3	0.0
HONG KONG	5	2	0.0
HORSE -SP	6	4	0.0
HOW COME	7	2	0.0
HOW MANY	6	4	0.0
HOW TO	5	3	0.0
HUMPTY DUMPTY	20	3	0.0
HURRY UP	16	4	0.0
IF I	5	4	0.0
IF YOU	15	6	0.0
JANET JOHN	10	3	0.0
JOHN COME	10	3	0.0
JUST LET	7	2	0.0
JUST WHAT	6	2	0.0
KICK IT	10	3	0.0
KID -SP	17	3	0.0
KING -SS	9	5	0.0
KITTEN JUMP	5	2	0.0
KITTEN KITTEN	8	3	0.0
LA LA	14	2	0.0
LAST ONE	8	4	0.0
LEG -SP	8	3	0.0
LET ME	15	6	0.0
LETS GO	5	2	0.0
LETS SEE	5	3	0.0
LL DRAW	5	2	0.0
LOOKING FOR	7	4	0.0
MADE A	5	2	0.0
MAKING A	5	3	0.0
MAMMA MAMMA	5	2	0.0
MAN IS	8	4	0.0
MANY COLOUR	6	3	0.0
MINE S	5	2	0.0
MISS ARMSTRONG	7	2	0.0
MISS WEIR	17	2	0.0
MUCH IS	7	2	0.0
NAME IS	6	4	0.0
NAUGHTY BOY	6	3	0.0
NINETEEN TWENTY	6	3	0.0
OF A	5	3	0.0
OF YOU	5	2	0.0
OPEN IT	5	2	0.0
OPEN THE	10	3	0.0
PAT DIGGER	9	2	0.0
PLAYING WITH	7	6	0.0
PUT ONE	5	3	0.0
QUICK QUICK	12	2	0.0

SEQUENCE	FREQ	SPEAKERS	INDEX
RABBIT -SP	5	2	0.0
RABBIT AND	5	2	0.0
RE COMING	5	2	0.0
RE GONNA	6	3	0.0
READING BOOK	5	2	0.0
RIDE A	6	2	0.0
SAID YOU	5	2	0.0
SAT ON	16	3	0.0
SAW A	7	3	0.0
SAY I	6	3	0.0
SAY THAT	6	3	0.0
SHALL I	9	2	0.0
SHELL -SP	7	3	0.0
SHOE -SP	11	2	0.0
SHOW YOU	5	4	0.0
SHUT THE	7	5	0.0
SHUT UP	13	4	0.0
SHUTTING THE	6	4	0.0
SIR THREE	6	4	0.0
SIR YES	6	4	0.0
SIT DOWN	7	4	0.0
SITTING IN	5	4	0.0
SOME MORE	6	3	0.0
STANDING IN	5	4	0.0
STOP IT	8	5	0.0
SUPPOSED TO	6	2	0.0
TAKE IT	6	6	0.0
TAKE THIS	7	6	0.0
TEACHER -SP	7	3	0.0
TELL ME	7	4	0.0
TELL YOU	11	6	0.0
THEM IN	9	4	0.0
THEN YOU	7	3	0.0
THING -SP	6	3	0.0
TREE IS	5	2	0.0
TREE TREE	5	3	0.0
TRYING TO	6	3	0.0
TWINKLE TWINKLE	8	2	0.0
UH UH	12	2	0.0
UM ?	8	2	0.0
VE FINISHED	5	4	0.0
VERY FAST	5	3	0.0
VERY GOOD	6	2	0.0
WANNA GO	7	3	0.0
WANNA PLAY	12	2	0.0
WASH MY	6	3	0.0
WATT I	8	2	0.0
WATT MRS	7	2	0.0
WELL I	11	5	0.0
WENT UP	7	3	0.0
WENT UP	7	3	0.0
WHEN I	10	4	0.0
WHICH ONE	13	5	0.0
WHO S	7	3	0.0
WHY DID	5	2	0.0
WITH MANY	5	3	0.0
YELLOW YELLOW	6	2	0.0
YOURS IS	8	3	0.0

THREE WORD SEQUENCES OF THE TOTAL SAMPLE OF ABORIGINAL AND MIGRANT
CHILDREN IN ORDER OF INDEX VALUE

SEQUENCE	FREQ	SPEAKERS	INDEX
ONE TWO THREE	270	17	178.3
? WHAT S	152	16	140.4
TWO THREE FOUR	206	17	61.0
THREE FOUR FIVE	170	15	28.2
I M GOING	37	8	14.5
I M NOT	29	9	12.8
FOUR FIVE SIX	127	15	10.3
I DONT KNOW	50	16	10.0
WHAT S THIS	71	14	7.8
? WHAT IS	19	7	7.7
? WHAT DO	19	5	5.5
I CAN SEE	24	10	4.9
? WHAT COLOUR	11	7	4.4
ONE TWO ONE	27	4	4.2
I M GONNA	21	4	4.1
? WHERE IS	33	8	3.4
? WHERE S	21	12	3.3
I M A	11	6	3.2
FIVE SIX SEVEN	84	12	2.8
WHAT S THAT	43	8	2.7
? WHAT ARE	9	5	2.6
I M TELLING	10	5	2.5
I VE GOT	34	12	2.1
I WANT TO	28	9	1.9
IT S A	21	10	1.6
? WHERE ARE	14	8	1.5
I M GETTING	5	4	1.0
I M MAKING	7	3	1.0
I M LOOKING	4	4	0.8
I M VERY	4	4	0.8
IT S NOT	15	7	0.8
SIX SEVEN EIGHT	62	11	0.8
? HOW ARE	24	6	0.7
? WHAT DID	4	3	0.7
I CAN DO	9	4	0.7
I DONT WANT	11	5	0.7
I M HERE	4	3	0.6
NO NO NO	28	6	0.6
THAT S A	19	9	0.6
? CAN I	15	8	0.5
? CAN YOU	18	7	0.5
? WHAT HAPPENED	3	3	0.5
I CAN PLAY	9	3	0.5
I LL GO	9	6	0.5
I M AFTER	5	2	0.5
I M FINISHED	5	2	0.5
THE LITTLE DOG	20	5	0.5
I AM PLAYING	5	4	0.4
I CAN MAKE	6	3	0.4
I LL DO	8	5	0.4

SEQUENCE	FREQ	SPEAKERS	INDEX
I M EATING	3	3	0.4
? DO YOU	20	10	0.3
? HOW MUCH	13	5	0.3
? WHAT RE	3	2	0.3
I AM GOING	4	4	0.3
I AM HERE	8	2	0.3
I CAN GO	5	3	0.3
I LL GET	8	4	0.3
I M PLAYING	3	2	0.3
I M PUTTING	3	2	0.3
SEE THE LITTLE	31	5	0.3
SEVEN EIGHT NINE	45	11	0.3
? WHAT ABOUT	2	2	0.2
I AM SITTING	4	3	0.2
I CAN RIDE	5	2	0.2
I DONT CARE	5	3	0.2
I DONT HAVE	4	3	0.2
I HAVE A	6	5	0.2
I HAVE SOME	9	3	0.2
I KNOW I	5	5	0.2
I LL TELL	7	4	0.2
I M CAUGHT	2	2	0.2
I M DOING	2	2	0.2
I M HUNGRY	2	2	0.2
I M NEARLY	2	2	0.2
I M THE	2	2	0.2
I M UP	2	2	0.2
I WANT A	6	5	0.2
IN THE BIN	10	6	0.2
LOOK AT THAT	11	6	0.2
LOOK AT THIS	20	3	0.2
ONE TWO AND	2	2	0.2
ONE TWO FIVE	2	2	0.2
ONE TWO FOUR	3	2	0.2
THAT S WHAT	8	5	0.2
THIS IS A	14	6	0.2
TWO THREE THREE	3	3	0.2
* ONE TWO	21	6	0.1
? ARE YOU	12	9	0.1
? HOW MANY	5	3	0.1
? IS IT	12	4	0.1
? YOU KNOW	7	4	0.1
? YOU WANNA	8	2	0.1
A LITTLE BIT	6	4	0.1
AND SEE THE	16	5	0.1
CAN I HAVE	20	10	0.1
EIGHT NINE TEN	27	10	0.1
I AM EATING	3	2	0.1
I AM HITTING	2	2	0.1
I AM MAKING	2	2	0.1

SEQUENCE	FREQ	SPEAKERS	INDEX
I AM MAKING	2	2	0.1
I AM NOT	2	2	0.1
I AM STANDING	2	2	0.1
I CAN CATCH	2	2	0.1
I CAN HEAR	3	2	0.1
I CAN RUN	2	2	0.1
I CAN SHUT	2	2	0.1
I CAN TAKE	2	2	0.1
I DONT LIKE	2	2	0.1
I DONT THINK	3	2	0.1
I DONT WANNA	2	2	0.1
I HAVE THIS	5	4	0.1
I HAVE TO	4	3	0.1
I KNOW WHAT	3	3	0.1
I KNOW YOU	6	3	0.1
I LIKE THE	6	3	0.1
I LL BE	3	2	0.1
I LL DRAW	5	2	0.1
I LL PUT	3	2	0.1
I LL TAKE	3	2	0.1
I SEE THE	15	3	0.1
I VE FINISHED	5	4	0.1
I WANT SOME	3	3	0.1
IN THE BOX	7	3	0.1
IN THE WATER	7	3	0.1
IT S AN	4	2	0.1
IT S GOING	4	2	0.1
IT S HERE	4	4	0.1
IT S MINE	4	4	0.1
IT S SUPPOSED	4	2	0.1
IT S THE	6	3	0.1
LOOK AT HIM	4	4	0.1
LOOK AT MINE	6	5	0.1
LOOK AT THE	5	5	0.1
M GOING TO	29	7	0.1
NO I M	10	6	0.1
SEE THE BOAT	10	4	0.1
SEE THE KITTEN	10	3	0.1
THAT S EASY	5	3	0.1
THAT S MINE	5	3	0.1
THAT S MY	7	4	0.1
THAT S WHY	4	4	0.1
THE LITTLE BOY	5	3	0.1
THIS IS MY	12	5	0.1
THIS ONE HERE	18	2	0.1
TWO ONE TWO	32	5	0.1
TWO THREE FIVE	2	2	0.1
WHAT DO YOU	21	6	0.1
WHAT IS IT	16	5	0.1
WHAT S THE	6	2	0.1

SEQUENCE	FREQ	SPEAKERS	INDEX
WHAT S WHAT	5	3	0.1
WHAT S YOUR	3	3	0.1
YOU HAVE TO	13	6	0.1
YOU RE A	6	4	0.1
* ? WHERE	4	3	0.0
* GOOD AFTERNOON	2	2	0.0
* HUMPTY DUMPTY	12	2	0.0
* I CAN	10	4	0.0
* IN THE	4	2	0.0
* ON THE	4	2	0.0
* SEE THE	5	2	0.0
* THREE FOUR	2	2	0.0
-SP AND ALL	3	2	0.0
-SP AND ONE	2	2	0.0
-SP AND THE	2	2	0.0
-SP ARE HERE	3	3	0.0
-SP ON IT	4	2	0.0
? AND WHAT	2	2	0.0
? CAN WE	2	2	0.0
? DID YOU	9	5	0.0
? DONT YOU	7	2	0.0
? HAVE YOU	7	6	0.0
? HOW COME	4	2	0.0
? I CAN	6	5	0.0
? I M	3	3	0.0
? IS THIS	5	2	0.0
? IT S	3	2	0.0
? MISS WEIR	5	2	0.0
? NO WHAT	2	2	0.0
? OH WHAT	4	4	0.0
? SEE THAT	3	3	0.0
? SHALL I	7	2	0.0
? THAT S	3	2	0.0
? THIS ONE	7	4	0.0
? WANT TO	5	2	0.0
? WHICH ONE	11	5	0.0
? WHO S	4	2	0.0
? WHY ARE	2	2	0.0
? WHY DID	4	2	0.0
? WILL YOU	6	4	0.0
? YOU CAN	4	3	0.0
A BIG ONE	5	4	0.0
A GREAT FALL	7	2	0.0
A MAN AND	4	3	0.0
A YELLOW ONE	2	2	0.0
ALL OF IT	2	2	0.0
ALL THE KING	5	2	0.0
ALL THE TIME	2	2	0.0
AND A WOMAN	2	2	0.0
AND AFTER I	3	2	0.0

SEQUENCE	FREQ	SPEAKERS	INDEX
AND AFTER I	3	2	0.0
AND DOWN UP	11	2	0.0
AND GET IT	2	2	0.0
AND I KNOW	2	2	0.0
AND I LI	3	3	0.0
AND I M	4	3	0.0
AND IT S	2	2	0.0
AND ONE FOR	7	4	0.0
AND ONE RABBIT	2	2	0.0
AND THAT ONE	2	2	0.0
AND THAT S	3	3	0.0
AND THE BOY	3	3	0.0
AND THE DOG	3	3	0.0
AND THE GIRL	2	2	0.0
AND THE LION	3	2	0.0
AND THE LITTLE	3	3	0.0
AND THE QNOUN	2	2	0.0
AND THE TWO	2	2	0.0
AND THIS IS	3	2	0.0
AND THIS ONE	4	3	0.0
ARE YOU DOING	4	4	0.0
ARE YOU DRAWING	2	2	0.0
ARE YOU GOING	4	3	0.0
ARE YOU PLAYING	3	3	0.0
ARE YOU THIS	13	2	0.0
ARE YOU TODAY	6	2	0.0
AT THE DOOR	2	2	0.0
BAA BAA BLACK	4	3	0.0
BAG -SP FULL	4	3	0.0
BECAUSE YOU SAY	3	2	0.0
BLACK SHEEP HAVE	4	3	0.0
BOOK -SP IN	2	2	0.0
BOY -SP AND	2	2	0.0
BOY AND A	2	2	0.0
BOY AND THE	2	2	0.0
BUT I M	5	5	0.0
CAN COME UP	5	2	0.0
CAN DO THAT	2	2	0.0
CAN DO THIS	5	2	0.0
CAN GO HOME	3	2	0.0
CAN GO UP	6	2	0.0
CAN HAVE IT	4	3	0.0
CAN HAVE THIS	2	2	0.0
CAN JUMP AND	2	2	0.0
CAN MAKE A	3	2	0.0
CAN PLAY WITH	4	2	0.0
CAN/ RIDE A	6	2	0.0
CAN RUN AND	4	2	0.0
CAN SEE THE	9	6	0.0
CAN SEE YOU	4	3	0.0

SEQUENCE	FREQ	SPEAKERS	INDEX
CAN YOU CAN	2	2	0.0
CAN YOU SEE	4	3	0.0
COME AND PLAY	6	2	0.0
COME AND SEE	19	4	0.0
COME IN HERE	2	2	0.0
COME LOOK AND	4	2	0.0
COME ON I	2	2	0.0
COME ON QUICKLY	2	2	0.0
COME SEE THE	5	2	0.0
DID YOU GO	2	2	0.0
DID YOU SAY	3	2	0.0
DID YOU SEE	2	2	0.0
DO YOU KNOW	7	5	0.0
DO YOU LIKE	3	3	0.0
DO YOU THINK	4	3	0.0
DO YOU WANT	13	6	0.0
DOG CAN COME	4	2	0.0
DOG CAN RUN	4	3	0.0
DOG RUN AND	3	2	0.0
DONT HAVE TO	7	4	0.0
DONT KNOW HOW	2	2	0.0
DONT KNOW WHAT	2	2	0.0
DONT KNOW WHERE	4	3	0.0
DONT WANT THIS	2	2	0.0
DONT WANT TO	7	4	0.0
DOWN ON THE	3	2	0.0
DOWN UP AND	11	2	0.0
DUMPTY HAD A	7	2	0.0
DUMPTY SAT ON	8	2	0.0
EATING THE BREAD	2	2	0.0
EATING THE FLOWER	3	2	0.0
EIGHT NINE NINE	3	2	0.0
ELEPHANT AND THE	3	2	0.0
ELEVEN TWELVE THIRTEEN	4	2	0.0
FIVE ONE THREE	2	2	0.0
FIVE ONE TWO	14	3	0.0
FOR A LITTLE	2	2	0.0
FOR A MINUTE	2	2	0.0
FOR THE DAME	2	2	0.0
FOR THE LITTLE	2	2	0.0
FOR THE MASTER	4	2	0.0
FOUR FIVE EIGHT	2	2	0.0
FOUR ONE TWO	12	2	0.0
GET A BALLOON	2	2	0.0
GET OUT OF	4	3	0.0
GIRL --SP ARE	2	2	0.0
GIVE IT TO	3	3	0.0
GIVE ME A	3	2	0.0
GIVE ME IT	2	2	0.0
GIVE ME ONE	5	4	0.0

SEQUENCE	FREQ	SPEAKERS	INDEX
GIVE ME ONE	5	4	0.0
GIVE ME SOME	9	3	0.0
GIVE ME THAT	3	3	0.0
GO DOWN HERE	2	2	0.0
GO DOWN THERE	5	2	0.0
GO IN THE	2	2	0.0
GO ON THIS	3	2	0.0
GO TO SCHOOL	3	2	0.0
GO TO THE	3	3	0.0
GOING TO GET	6	3	0.0
GOING TO GIVE	2	2	0.0
GOING TO MAKE	2	2	0.0
GOING TO PLAY	4	3	0.0
GOING TO SCHOOL	2	2	0.0
GOT A BALLOON	9	2	0.0
GOT A WHITE	2	2	0.0
GOT A YELLOW	3	2	0.0
GOT IT IN	3	3	0.0
GOT ONE OF	2	2	0.0
GOT THE SAME	2	2	0.0
GOT TO RUN	2	2	0.0
HA HA HA	4	2	0.0
HAD A GREAT	7	2	0.0
HAND -SP UP	2	2	0.0
HAVE A CAR	2	2	0.0
HAVE A LOOK	3	2	0.0
HAVE IT AT	2	2	0.0
HAVE TO BE	2	2	0.0
HAVE TO GO	3	2	0.0
HAVE TO SAY	2	2	0.0
HAVE TO TALK	3	2	0.0
HAVE TO TELL	2	2	0.0
HAVE YOU ANY	6	4	0.0
HAVE YOU GOT	5	2	0.0
HAVENT GOT ANY	6	3	0.0
HE IS IN	2	2	0.0
HE S A	2	2	0.0
HE S EATING	3	2	0.0
HE S HUNGRY	2	2	0.0
HE S NOT	2	2	0.0
HERE HERE HERE	3	2	0.0
HERE I AM	24	2	0.0
HERE I COME	2	2	0.0
HERE I GO	4	2	0.0
HERE IS THE	2	2	0.0
HERE S A	5	3	0.0
HERE S THE	2	3	0.0
HERE WE ARE	4	2	0.0
HICKORY DICKORY DOCK	5	2	0.0
HORSE -SP AND	3	2	0.0

SEQUENCE	FREQ	SPEAKERS	INDEX
HOUSE AND A	2	2	0.0
HOW ARE YOU	24	6	0.0
HOW MUCH IS	7	2	0.0
HUMPTY DUMPTY HAD	7	2	0.0
HUMPTY DUMPTY SAT	7	2	0.0
I CANT DO	2	2	0.0
I DID IT	2	2	0.0
I DIDNT HEAR	3	2	0.0
I EAT THE	2	2	0.0
I GO TO	2	2	0.0
I GOT A	3	3	0.0
I GOT ONE	3	2	0.0
I GOT TWO	2	2	0.0
I HAD TWO	2	2	0.0
I HAVENT GOT	5	3	0.0
I KNOW HOW	2	2	0.0
I LIKE IT	2	2	0.0
I LIKE THAT	4	4	0.0
I LL BUY	2	2	0.0
I LL MAKE	2	2	0.0
I MADE A	3	2	0.0
I SEE TWO	4	2	0.0
I WANNA GO	6	3	0.0
I WANT THAT	2	2	0.0
I WAS PLAYING	2	2	0.0
IN OUR CLASS	3	2	0.0
IN THE AIR	2	2	0.0
IN THE BACK	2	2	0.0
IN THE BAG	3	3	0.0
IN THE BUCKET	2	2	0.0
IN THE CHAIR	3	2	0.0
IN THE CREEK	3	2	0.0
IN THE DOOR	4	3	0.0
IN THE FRONT	2	2	0.0
IN THE PARK	2	2	0.0
IN THE QNOUN	3	2	0.0
IN THE SAND	2	2	0.0
IN THE SKY	3	3	0.0
IN THE TREE	2	2	0.0
IS EATING THE	3	2	0.0
IS IN THE	5	4	0.0
IS MY HOUSE	7	3	0.0
IS ON THE	5	3	0.0
IS RUNNING VERY	2	2	0.0
IS SITTING IN	2	2	0.0
IS SITTING ON	3	3	0.0
IS STANDING IN	2	2	0.0
IT FOR ME	4	2	0.0
IT IS A	3	3	0.0
IT ON MY	2	2	0.0

SEQUENCE	FREQ	SPEAKERS	INDEX
IT ON MY	2	2	0.0
IT S CATCHING	2	2	0.0
IT S COLD	2	2	0.0
IT S COMING	2	2	0.0
IT TO ME	4	4	0.0
IT TO YOU	2	2	0.0
IT UP THERE	4	2	0.0
JANET JOHN COME	7	2	0.0
JOHN COME LOOK	6	2	0.0
KING -SS HORSE	3	2	0.0
KING -SS PARK	3	2	0.0
KITTEN KITTEN KITTEN	4	2	0.0
KNOW HOW TO	3	2	0.0
KNOW I KNOW	3	3	0.0
KNOW WHAT THIS	2	2	0.0
KNOW WHERE IT	2	2	0.0
LA LA LA	7	2	0.0
LET ME SEE	9	3	0.0
LIKE THAT ?	2	2	0.0
LIKE THIS AND	3	2	0.0
LIKE THIS ONE	2	2	0.0
LITTLE BOY AND	2	2	0.0
LITTLE DOG AND	2	2	0.0
LITTLE DOG CAN	6	2	0.0
LITTLE DOG RUN	9	4	0.0
LL BE HERE	2	2	0.0
LL DO IT	6	3	0.0
LL DO THAT	3	2	0.0
LL GET YOU	4	3	0.0
LL TELL YOU	5	3	0.0
LOOK AND SEE	6	3	0.0
LOOK AT IT	2	2	0.0
LOOK AT ME	2	2	0.0
LOOK AT ONE	2	2	0.0
LOOK AT YOUR	3	3	0.0
LOOK DOWN HERE	3	2	0.0
LOOK JOHN SEE	2	2	0.0
LOOK LOOK LOOK	3	2	0.0
LOOK SEE THE	5	2	0.0
LOOK UP THERE	2	2	0.0
LOOK UP UP	2	2	0.0
LOOKING AT THE	7	4	0.0
LOOKING AT YOU	2	2	0.0
M A GOOD	2	2	0.0
M GONNA TO	3	2	0.0
M NOT GOING	3	2	0.0
M NOT GONNA	5	2	0.0
M TELLING THE	2	2	0.0
M TELLING YOU	2	2	0.0
M TELLING YOUR	2	2	0.0

SEQUENCE	FREQ	SPEAKERS	INDEX
MAN AND A	4	2	0.0
MANY COLOUR -SP	6	3	0.0
ME ME ME	5	2	0.0
MRS WATT I	8	2	0.0
MRS WATT MRS	7	2	0.0
MY DAD -SS	2	2	0.0
MY THING -SP	3	2	0.0
NINE TEN ELEVEN	9	5	0.0
NINE TEN TEN	2	2	0.0
NO I AM	2	2	0.0
NO I DONT	5	4	0.0
NO I LL	2	2	0.0
NO IT ISNT	2	2	0.0
NO IT S	6	6	0.0
NO NOT THAT	3	2	0.0
NO THAT S	3	3	0.0
NOT ALLOWED TO	2	2	0.0
NOT GOING TO	3	2	0.0
NOT THAT ONE	4	2	0.0
NOW I M	5	4	0.0
OF THE HOUSE	3	2	0.0
OH I M	2	2	0.0
OH LOOK AT	3	3	0.0
OH NO NO	4	2	0.0
OH THAT S	3	3	0.0
OH WHAT S	3	3	0.0
ON A WALL	7	2	0.0
ON MY HEAD	2	2	0.0
ON THE CHAIR	4	4	0.0
ON THE FLOOR	8	2	0.0
ON THE GRASS	11	2	0.0
ON THE HEAD	3	2	0.0
ONE AND TWO	2	2	0.0
ONE FOR THE	14	4	0.0
ONE FOR YOU	2	2	0.0
ONE OF THESE	2	2	0.0
ONE OF THOSE	4	3	0.0
ONE ONE ONE	4	3	0.0
ONE ONE TWO	5	4	0.0
ONE THREE FOUR	3	3	0.0
OPEN THE DOOR	10	3	0.0
OUT OF THE	7	4	0.0
PAT DIGGER SAM	3	2	0.0
PLAY WITH ME	2	2	0.0
PLAY WITH THIS	2	2	0.0
PLAYING WITH THE	2	2	0.0
PUT IT BACK	2	2	0.0
PUT DOWN	4	3	0.0
PUT IT IN	5	4	0.0
PUT IT ON	7	6	0.0

SEQUENCE	FREQ	SPEAKERS	INDEX
PUT IT ON	7	6	0.0
PUT IT THIS	2	2	0.0
PUT THEM IN	6	3	0.0
PUT THIS ON	2	2	0.0
PUT YOUR HAND	4	3	0.0
RE GONNA HAVE	2	2	0.0
RIDE A HORSE	6	2	0.0
S A BIG	3	3	0.0
S A DOG	4	3	0.0
S A GOAT	2	2	0.0
S A PIG	3	3	0.0
S GOING TO	5	2	0.0
S NOT A	3	3	0.0
S NOT YOUR	2	2	0.0
S THIS FOR	3	2	0.0
S WHAT I	2	2	0.0
S WHAT S	5	3	0.0
S YOUR NAME	3	3	0.0
SAT ON A	7	2	0.0
SAT ON THE	9	2	0.0
SAW A LITTLE	6	2	0.0
SEE THE AEROPLANE	2	2	0.0
SEE THE BOY	2	2	0.0
SEE THE BUTTERFLY	2	2	0.0
SEE THE CAT	8	2	0.0
SEE THE DOG	5	3	0.0
SEE THE DUCK	3	2	0.0
SEE THE ELEPHANT	4	2	0.0
SEE THE TWO	3	2	0.0
SEE TWO KITTEN	4	2	0.0
SEE YOU TOMORROW	2	2	0.0
SHE S GOT	2	2	0.0
SHE S WATCHING	2	2	0.0
SHOW YOU THE	2	2	0.0
SHUT THE DOOR	5	4	0.0
SHUT THE WINDOW	2	2	0.0
SHUTTING THE DOOR	6	4	0.0
SIR THREE BAG	6	4	0.0
SIR YES SIR	6	4	0.0
SIT DOWN HERE	2	2	0.0
SITTING IN THE	4	3	0.0
SITTING ON THE	22	4	0.0
SIX ONE TWO	12	2	0.0
SIX SIX SIX	3	2	0.0
STANDING IN A	2	2	0.0
SUPPOSED TO BE	5	2	0.0
TAKE IT HOME	2	2	0.0
TAKE IT OFF	2	2	0.0
TAKE THIS OFF	3	3	0.0
TEN ELEVEN TWELVE	8	4	0.0

SEQUENCE	FREQ	SPEAKERS	INDEX
THAT ONE OVER	3	2	0.0
THAT ONE TOO	2	2	0.0
THAT S ALL	2	2	0.0
THAT S BEAUTIFUL	5	2	0.0
THAT S NOT	4	3	0.0
THAT S RIGHT	2	2	0.0
THAT S THE	3	2	0.0
THAT S TWO	2	2	0.0
THAT THAT S	3	2	0.0
THE AEROPLANE IS	2	2	0.0
THE BOY -SP	2	2	0.0
THE BOY IS	3	2	0.0
THE CAR -SP	2	2	0.0
THE CARD -SP	4	3	0.0
THE CAT AND	3	2	0.0
THE DOG CAN	3	2	0.0
THE DOOR IS	3	3	0.0
THE DOOR PLEASE	3	2	0.0
THE DUCK RUN	2	2	0.0
THE ELEPHANT AND	3	2	0.0
THE FAT CAT	4	2	0.0
THE GIRL -SP	7	3	0.0
THE GIRL IS	2	2	0.0
THE HORSE IS	2	2	0.0
THE KING -SS	6	3	0.0
THE KITTEN JANET	2	2	0.0
THE KITTEN JUMP	3	2	0.0
THE LAST ONE	6	4	0.0
THE LITTLE APPLE	2	2	0.0
THE LITTLE CAT	5	2	0.0
THE MAN IS	5	4	0.0
THE OTHER ONE	2	2	0.0
THE THE THE	5	2	0.0
THE TWO KITTEN	5	3	0.0
THEM IN THE	2	2	0.0
THERE S A	4	4	0.0
THEY ARE NOT	3	2	0.0
THEY ARE PLAYING	3	3	0.0
THEY RE COMING	3	2	0.0
THIS FOR ME	4	3	0.0
THIS IS MINE	3	2	0.0
THIS IS THE	3	3	0.0
THIS WAY ?	2	2	0.0
THREE BAG -SP	6	4	0.0
THREE BOAT -SP	8	2	0.0
THREE FOUR SIX	2	2	0.0
THREE ONE TWO	14	2	0.0
THREE HREE THREE	2	2	0.0
TO DO IT	3	2	0.0
TO DO THAT	2	2	0.0

SEQUENCE	FREQ	SPEAKERS	INDEX
TO DO THAT	2	2	0.0
TO DO THIS	2	2	0.0
TO DRAW A	3	2	0.0
TO GO TO	2	2	0.0
TO MAKE A	3	2	0.0
TO PLAY WITH	5	3	0.0
TO TALK IN	2	2	0.0
TO TELL YOU	3	3	0.0
TO THE TOILFT	5	3	0.0
TREE TREE TREE	2	2	0.0
TWINKLE TWINKLF LITTLE	2	2	0.0
TWO FOUR FIVE	4	3	0.0
TWO KITTEN -SP	16	4	0.0
TWO TWO FOUR	3	2	0.0
TWO TWO TWO	5	2	0.0
UH UH UH	4	2	0.0
UM ? WHAT	8	2	0.0
UP AND DOWN	22	4	0.0
UP THE CLOCK	10	3	0.0
UP THE TREE	4	3	0.0
UP THERE AND	3	2	0.0
UP THERE IN	2	2	0.0
UP THERE YOU	2	2	0.0
UP UP UP	3	2	0.0
VE GOT A	12	7	0.0
VE GOT MY	3	2	0.0
VE GOT ONE	2	2	0.0
VE GOT TO	4	4	0.0
VE GOT TWO	8	2	0.0
WANT ME TO	3	2	0.0
WANT TO BE	2	2	0.0
WANT TO GET	2	2	0.0
WANT TO GO	4	3	0.0
WANT TO HAVE	2	2	0.0
WANT TO HIT	3	2	0.0
WANT TO PLAY	8	5	0.0
WANT TO SEF	2	2	0.0
WASH MY HANDS	2	2	0.0
WATT MRS WATT	7	2	0.0
WE ARE GOING	2	2	0.0
WE RE NOT	6	3	0.0
WENT UP THE	5	2	0.0
WHAT ARE YOU	10	6	0.0
WHAT COLOUR IS	3	2	0.0
WHAT DID YOU	3	2	0.0
WHAT IS THAT	5	4	0.0
WHAT IS THIS	4	2	0.0
WHAT S HE	3	2	0.0
WHAT S THERE	3	2	0.0
WHAT S WRONG	3	2	0.0

SEQUENCE	FREQ	SPEAKERS	INDEX
WHEN I WAS	5	2	0.0
WHERE ARE ALL	3	2	0.0
WHERE ARE THEY	2	2	0.0
WHERE ARE YOU	5	4	0.0
WHERE IS IT	3	2	0.0
WHERE IS RINGMAN	2	2	0.0
WHERE IS TALLMAN	9	2	0.0
WHERE IS THUMBKIN	8	3	0.0
WHERE IS YOUR	2	2	0.0
WHERE S MINE	2	2	0.0
WHERE S MY	6	4	0.0
WHERE S THE	4	4	0.0
WHERE S YOUR	3	3	0.0
WHO S THAT	4	3	0.0
WHY DID YOU	5	2	0.0
WILL YOU DO	2	2	0.0
WILL YOU GET	3	2	0.0
WITH A GUN	2	2	0.0
WITH MANY COLOUR	2	2	0.0
WITH THE BALL	2	2	0.0
YELLOW YELLOW YELLOW	2	2	0.0
YES I AM	4	2	0.0
YES I CAN	2	2	0.0
YES I LIKE	2	2	0.0
YES SIR THREE	6	4	0.0
YES SIR YES	6	4	0.0
YES YES YES	2	2	0.0
YES YOU CAN	2	2	0.0
YOU ANY WOOL	3	3	0.0
YOU ARE A	2	2	0.0
YOU BECAUSE YOU	3	2	0.0
YOU CAN HAVE	4	3	0.0
YOU CAN KEEP	2	2	0.0
YOU CAN PLAY	5	3	0.0
YOU CAN SEE	3	3	0.0
YOU CAN YOU	2	2	0.0
YOU COME AND	2	2	0.0
YOU COME HERE	3	2	0.0
YOU DO IT	5	5	0.0
YOU DO THAT	5	4	0.0
YOU DONT KNOW	4	2	0.0
YOU DONT WANT	2	2	0.0
YOU GET A	2	2	0.0
YOU GO AND	2	2	0.0
YOU GO IN	2	2	0.0
YOU GOING TO	4	2	0.0
YOU GOT A	2	2	0.0
YOU GOT IT	2	2	0.0
YOU HAVE A	3	3	0.0
YOU HAVE GOT	2	2	0.0

SEQUENCE	FREQ	SPEAKERS	INDEX
YOU HAVE GOT	2	2	0.0
YOU KNOW THAT	2	2	0.0
YOU KNOW UM	2	2	0.0
YOU KNOW WHAT	7	4	0.0
YOU KNOW WHERE	3	3	0.0
YOU LIKE IT	3	2	0.0
YOU LIKE THE	2	2	0.0
YOU LL BE	2	2	0.0
YOU LL SEE	2	2	0.0
YOU PUT IT	3	3	0.0
YOU PUT THIS	3	2	0.0
YOU RE GOING	3	2	0.0
YOU RE GONNA	4	3	0.0
YOU RE LOOKING	2	2	0.0
YOU RE NOT	3	3	0.0
YOU SAY I	2	2	0.0
YOU SAY THAT	5	3	0.0
YOU SAY WHAT	2	2	0.0
YOU SEE THE	2	2	0.0
YOU THIS MORNING	13	2	0.0
YOU VE GOT	11	5	0.0
YOU VE GOTTA	2	2	0.0
YOU WANNA PLAY	9	2	0.0
YOU WANT A	3	3	0.0
YOU WANT ME	6	2	0.0
YOU WANT TO	7	5	0.0
YOUR HAND -SP	3	3	0.0
YOURS IS YUK	2	2	0.0

APPENDIX 10

ONE, TWO AND THREE WORD
SEQUENCES OF THE ABORIGINAL
SAMPLE OF UPPER FIVE YEAR OLD
CHILDREN IN ORDER OF INDEX
VALUE

EXPLANATORY NOTES

1. Full lists are also available of single words and sequences in alphabetic and raw frequency order. It is regretted that space forbids their inclusion in this report.
2. In order to provide important information on such matters as number of possessives, plurals, questions, etc. used, relevant non-alphabetic symbols are listed along with other words and morphemes on the following pages.
3. For details on the calculation of an Index Number for single words and two or three word sequences see pages 17 and 20 respectively. Index numbers for three word sequences should be multiplied by 100 to make them comparable with the index numbers used on the other lists.
4. For details on the interpretation of all symbols used in the Print out see Appendix 4: Instructions for Coding.

SINGLE WORDS USED IN THE ABORIGINAL SAMPLE OF UPPER FIVE YEAR OLD
CHILDREN IN ORDER OF INDEX VALUE

SEQUENCE	FREQ	SPEAKERS	INDEX
I	347	7	569.1
J	256	8	479.9
YOU	211	8	395.5
A	178	8	333.6
ONE	177	8	331.8
WHAT	200	7	328.0
THAT	157	7	257.5
IT	135	8	253.0
S	137	7	224.7
THIS	128	7	209.9
NO	106	7	173.9
HE	92	8	172.4
HERE	90	8	168.7
THE	97	7	159.1
THERE	96	7	157.5
GO	83	8	155.6
OH	76	8	142.5
TO	69	8	129.3
LOOK	68	8	127.5
UP	63	7	103.3
IN	60	7	98.4
M	51	8	95.6
ON	60	6	84.3
AND	50	7	82.0
ME	50	7	82.0
-SP	43	7	70.5
YES	49	6	68.9
LITTLE	35	8	65.6
GO	45	6	63.3
TALK	45	6	63.3
FOR	37	7	60.7
NOT	32	7	52.5
ALL	30	7	49.2
DONT	30	7	49.2
TWO	39	5	45.7
AT	24	8	45.0
LIKE	30	6	42.2
PUT	35	5	41.0
GET	28	6	39.4
HAVE	24	7	39.4
IS	28	6	39.4
GOTTA	33	5	38.7
LI	33	5	38.7
GOING	26	6	36.6
SEE	26	6	36.6
WE	26	6	36.6
OUT	24	6	33.7
KNOW	28	5	32.8
THREE	28	5	32.8
THEY	23	6	32.3

SEQUENCE	FREQ	SPEAKERS	INDEX
WANT	23	6	32.3
SAY	33	4	30.9
HELLO	26	5	30.5
NOW	26	5	30.5
DO	21	6	29.5
CAN	25	5	29.3
FOUR	24	5	28.1
MY	24	5	28.1
OVER	19	6	26.7
HEY	16	7	26.2
COLOUR	22	5	25.8
COME	22	5	25.8
WHERE	18	6	25.3
BIG	15	7	24.6
CANT	17	6	23.9
OF	17	6	23.9
*	25	4	23.4
WANNA	19	5	22.3
GONNA	15	6	21.1
JUST	15	6	21.1
TAKE	15	6	21.1
THEM	15	6	21.1
EH	22	4	20.6
PLAY	22	4	20.6
OFF	17	5	19.9
THING	17	5	19.9
SIX	21	4	19.7
BACK	16	5	18.7
DOWN	16	5	18.7
FIVE	20	4	18.7
YOUR	11	7	18.0
EGG	18	4	16.9
SEVEN	18	4	16.9
SOME	14	5	16.4
ANOTHER	13	5	15.2
DID	13	5	15.2
PLAYING	13	5	15.2
VE	16	4	15.0
BALLOON	21	3	14.8
HIM	12	5	14.1
AH	14	4	13.1
WILL	14	4	13.1
YEAH	14	4	13.1
MINE	11	5	12.9
GIVE	18	3	12.7
HUH	27	2	12.7
ARE	10	5	11.7
AWAY	10	5	11.7
RE	10	5	11.7
RED	10	5	11.7

SEQUENCE	FREQ	SPEAKERS	INDEX
RED	10	5	11.7
RIGHT	10	5	11.7
THESE	10	5	11.7
TURN	10	5	11.7
AROUND	8	6	11.2
CAT	9	5	10.5
WAY	9	5	10.5
WHITE	9	5	10.5
-SS	11	4	10.3
UM	22	2	10.3
EIGHT	14	3	9.8
ER	14	3	9.8
AEROPLANE	10	4	9.4
GOOD	10	4	9.4
LEAVE	10	4	9.4
WITH	10	4	9.4
AGAIN	9	4	8.4
BLACK	9	4	8.4
CAR	9	4	8.4
CLOCK	18	2	8.4
DOG	12	3	8.4
HOME	9	4	8.4
MAN	9	4	8.4
COULD	7	5	8.2
TELL	7	5	8.2
AN	8	4	7.5
HAD	8	4	7.5
LOOKING	8	4	7.5
PIG	8	4	7.5
ROUND	8	4	7.5
TREE	8	4	7.5
AM	6	5	7.0
WILTS	15	2	7.0
GLITING	6	5	7.0
HAT	10	3	7.0
THEN	10	3	7.0
YELLOW	10	3	7.0
AFTER	7	4	6.6
RE	7	4	6.6
BRENDON	7	4	6.6
RIT	7	4	6.6
NOTHING	7	4	6.6
SHANE	14	2	6.6
BEEN	9	3	6.3
NIAONI	13	2	6.1
NINE	13	2	6.1
HOUSE	5	5	5.9
DOLL	6	4	5.6
DUCK	8	3	5.6
HURRY	12	2	5.6

SEQUENCE	FREQ	SPEAKERS	INDEX
IF	6	4	5.6
LAST	8	3	5.6
MAKE	8	3	5.6
MISS	8	3	5.6
MORE	6	4	5.6
NEVER	6	4	5.6
OTHER	8	3	5.6
OUR	8	3	5.6
PUSH	8	3	5.6
QNAME	8	3	5.6
TIME	6	4	5.6
WATER	6	4	5.6
WELL	6	4	5.6
COW	11	2	5.2
UH	11	2	5.2
BIRD	7	3	4.9
BROWN	7	3	4.9
COWBOY	7	3	4.9
TALKING	7	3	4.9
TRUCK	7	3	4.9
WHY	7	3	4.9
GUN	5	4	4.7
HICKORY	10	2	4.7
HIS	5	4	4.7
ISNT	5	4	4.7
MOUSE	10	2	4.7
BRETT	9	2	4.2
DICKORY	9	2	4.2
DOING	6	3	4.2
FISH	6	3	4.2
GIRL	6	3	4.2
GREEN	6	3	4.2
HER	6	3	4.2
HOLD	6	3	4.2
JENNIFER	6	3	4.2
SHIRT	6	3	4.2
SIDE	6	3	4.2
TURNIP	6	3	4.2
WENT	9	2	4.2
WHO	6	3	4.2
ARMSTRONG	8	2	3.7
COLD	5	3	3.5
DESERT	5	3	3.5
FALL	5	3	3.5
NAME	5	3	3.5
QUICK	5	3	3.5
RABBIT	5	3	3.5
RUN	5	3	3.5
THOSE	5	3	3.5
TYRE	5	3	3.5

SEQUENCE	FREQ	SPEAKERS	INDEX
TYRE	5	3	3.5
WAS	5	3	3.5
WHEN	5	3	3.5
HUMPTY	7	2	3.3
PUSSYCAT	7	2	3.3
YA	7	2	3.3
DUMPTY	6	2	2.8
FUCKING	6	2	2.8
OW	6	2	2.8
REALLY	6	2	2.8
TEAPOT	6	2	2.8
TEDDYBEAR	6	2	2.8
BASKET	5	2	2.3
BEAR	5	2	2.3
BUST	5	2	2.3
HARD	5	2	2.3
SITTING	5	2	2.3
WAIT	5	2	2.3

TWO WORD SEQUENCES OF THE ABORIGINAL SAMPLE OF UPPER FIVE YEAR
OLD CHILDREN IN ORDER OF INDEX VALUE

SEQUENCE	FREQ	SPEAKERS	INDEX
? WHAT	171	6	1368.1
I M	51	8	458.7
WHAT S	85	4	261.3
I LI	25	5	140.5
THAT ONE	36	6	130.3
THIS ONE	38	5	93.5
S THAT	40	4	84.2
? YOU	13	4	69.3
I WANT	12	5	67.5
S THIS	31	4	65.3
I CANT	10	5	56.2
? SEE	10	4	53.3
IT S	14	6	49.8
? WHERE	12	3	48.0
I DONT	10	4	45.0
ONE TWO	19	3	44.3
GOT A	20	6	43.7
A LITTLE	11	5	43.0
S A	16	5	42.1
THAT S	17	4	41.0
I CAN	9	4	40.5
LOOK AT	18	7	37.6
I KNOW	11	3	37.1
I AM	6	5	33.7
YOU GOT	9	4	33.4
I HAVE	5	5	28.1
I DID	6	4	27.0
A BALLOON	11	3	25.8
? CAN	6	3	24.0
A BIG	6	5	23.5
I JUST	5	4	22.5
YOU KNOW	8	3	22.2
I GOT	6	3	20.2
I VE	6	3	20.2
HE S	9	5	18.2
UP THERE	14	5	16.9
YOU CAN	6	3	16.7
YOU HAVE	6	3	16.7
YOU RE	6	3	16.7
YOU VE	6	3	16.7
NO YOU	10	4	16.3
YOU GOTTA	8	2	14.8
YOU WANNA	8	2	14.8
NO NO	9	4	14.7
YOU WANT	5	3	13.9
NO I	8	4	13.0
IN THE	14	4	12.9
IT IN	5	4	11.9
I WANNA	5	2	11.2
LOOK HERE	9	4	10.8

SEQUENCE	FREQ	SPEAKERS	INDEX
IT FOR	6	3	10.7
IT TO	6	3	10.7
M GOING	11	4	9.9
A COWBOY	6	2	9.4
THAT LITTLE	5	3	9.0
GOT TO	6	4	8.7
TWO THREE	19	4	8.1
OH NO	8	3	8.0
THE CLOCK	10	2	7.5
HE GOT	9	2	7.3
PUT IT	18	4	6.9
GOT IT	6	3	6.6
ALL THE	11	5	6.3
FOR YOU	11	4	6.3
THREE FOUR	20	4	6.1
THE LAST	5	3	5.6
IN HERE	6	4	5.5
AND A	6	4	4.6
FOR ME	8	4	4.5
THE MOUSE	6	2	4.5
TO THE	5	3	4.5
LIKE THAT	9	5	4.4
UP THE	9	2	4.4
LIKE THIS	10	4	4.0
OH OH	6	2	4.0
GOT ONE	5	2	3.6
ON THE	6	3	3.6
M GONNA	5	3	3.4
FOUR FIVE	16	3	3.2
M A	7	2	3.1
ME SOME	8	2	3.1
DONT KNOW	7	3	2.4
TALK TALK	8	2	2.4
GOING TO	9	3	2.3
IS A	6	4	2.2
M AFTER	5	2	2.2
FIVE SIX	15	3	2.0
GOTTA TALK	11	2	2.0
LL GO	6	3	1.6
OVER THERE	8	3	1.5
IS IT	5	3	1.4
OUT OF	6	3	1.4
VE GOT	10	4	1.4
WANT A	6	3	1.4
WANT TO	6	3	1.4
DO YOU	6	3	1.2
GIVE ME	14	3	1.2
SIX SEVEN	13	2	1.2
* ONE	9	2	1.0
ANOTHER ONE	7	4	1.0

SEQUENCE	FREQ	SPEAKERS	INDEX
ANOTHER ONE	7	4	1.0
COME ON	8	2	1.0
SEVEN EIGHT	13	2	1.0
HAVE A	5	2	0.9
HAVE TO	5	2	0.9
DO IT	6	2	0.8
COME BACK	6	2	0.7
SAY IT	5	2	0.7
COLOUR IN	5	2	0.6
ARE YOU	5	4	0.5
EIGHT NINE	10	2	0.5
THEM IN	5	2	0.5
WILL YOU	5	3	0.5
UM ?	8	2	0.4
BUT I	5	4	0.3
HAD A	5	3	0.3
LAST ONE	7	3	0.3
LEAVE IT	5	3	0.3
HICKORY DICKORY	8	2	0.2
MAKE A	6	2	0.2
MISS ARMSTRONG	7	2	0.2
HUMPTY DUMPTY	5	2	0.1
WENT UP	5	2	0.1

THREE WORD SEQUENCES OF THE ABORIGINAL SAMPLE OF UPPER FIVE YEAR OLD
CHILDREN IN ORDER OF INDEX VALUE

SEQUENCE	FREQ	SPEAKERS	INDEX
? WHAT S	79	4	1013.0
WHAT S THIS	31	4	75.9
I M GOING	11	4	47.3
WHAT S THAT	36	2	44.1
I M GONNA	5	3	16.1
I M A	7	2	15.0
? WHAT COLOUR	2	2	12.8
I M AFTER	5	2	10.7
I M NOT	4	2	8.6
I LL GO	6	3	5.9
ONE TWO THREE	18	3	5.6
I M UP	2	2	4.3
THAT S A	9	4	3.5
IT S A	5	4	2.3
I DONT KNOW	7	3	2.2
I LL TAKE	3	2	2.0
GOT A BALLOON	9	2	1.8
THAT ONE OVER	3	2	1.8
TWO THREE FOUR	18	4	1.4
I LL DO	2	2	1.3
I WANT TO	4	2	1.3
I VE GOT	6	3	0.9
I WANT A	3	2	0.9
? CAN I	4	3	0.7
I CAN SEE	3	2	0.6
I HAVE THIS	3	3	0.6
I WANT THAT	2	2	0.6
S A DOG	3	2	0.6
THREE FOUR FIVE	15	3	0.6
? SEE THAT	2	2	0.5
IT S CATCHING	2	2	0.5
IT S COLD	2	2	0.5
LOOK AT HIM	2	2	0.4
LOOK AT THE	2	2	0.4
S A GOAT	2	2	0.4
S A PIG	2	2	0.4
YOU HAVE TO	5	2	0.4
A BIG ONE	3	2	0.3
FOUR FIVE SIX	14	3	0.3
I AM GOING	2	2	0.3
M GOING TO	5	3	0.3
NO NO NO	5	2	0.3
YOU VE GOT	4	2	0.3
HE S NOT	2	2	0.2
I GOT A	2	2	0.2
IN THE CREEK	3	2	0.2
IT FOR ME	4	2	0.2
PUT IT IN	4	3	0.2
PUT IT ON	4	3	0.2
THE LAST ONE	5	3	0.2

SEQUENCE	FREQ	SPEAKERS	INDEX
UP THE CLOCK	8	2	0.2
UP THERE IN	2	2	0.2
YOU CAN HAVE	2	2	0.2
YOU VE GOTTA	2	2	0.2
FIVE SIX SEVEN	13	2	0.1
GOT TO RUN	2	2	0.1
IN THE AIR	2	2	0.1
NO I LL	2	2	0.1
NO I M	2	2	0.1
PUT IT BACK	2	2	0.1
PUT IT DOWN	3	2	0.1
PUT IT THIS	2	2	0.1
SIX SEVEN EIGHT	12	2	0.1
* ONE TWO	8	2	0.0
RUT I M	3	3	0.0
DO YOU WANT	4	3	0.0
GIVE ME SOME	8	2	0.0
LIKE THAT ?	2	2	0.0
M GONNA TO	3	2	0.0
OUT OF THE	3	2	0.0
SEVEN EIGHT NINE	10	2	0.0
UM ? WHAT	8	2	0.0
VE GOT A	4	2	0.0
WANT TO GET	2	2	0.0
WENT UP THE	5	2	0.0
WILL YOU DO	2	2	0.0

APPENDIX 11

ONE, TWO AND THREE WORD
SEQUENCES OF THE MACEDONIAN
SAMPLE OF UPPER FIVE YEAR
OLD CHILDREN IN ORDER OF
INDEX VALUE

EXPLANATORY NOTES

1. Full lists are also available of single words and sequences in alphabetic and raw frequency order. It is regretted that space forbids their inclusion in this report.
2. In order to provide important information on such matters as number of possessives, plurals, questions, etc. used, relevant non-alphabetic symbols are listed along with other words and morphemes on the following pages.
3. For details on the calculation of an Index Number for single words and two or three word sequences see pages 17 and 20 respectively. Index numbers for three word sequences should be multiplied by 100 to make them comparable with the index numbers used on the other lists.
5. For details on the interpretation of all symbols used in the Print out see Appendix 4: Instructions for Coding.

SINGLE WORDS USED IN THE MACEDONIAN SAMPLE OF UPPER FIVE YEAR OLD
CHILDREN IN ORDER OF INDEX VALUE

SEQUENCE	FREQ	SPEAKERS	INDEX
I	776	8	593.5
?	587	8	449.0
YOU	586	8	448.2
IT	319	8	244.0
THE	313	8	239.4
ONE	290	8	221.8
A	267	8	204.2
-SP	248	8	189.7
S	246	8	188.1
TO	241	8	184.3
THAT	234	8	179.0
THIS	225	8	172.1
NO	218	8	166.7
AND	209	8	159.9
WHAT	197	8	150.7
TWO	184	8	140.7
*	261	5	124.8
ME	160	8	122.4
YES	159	8	121.6
IS	157	8	120.1
DONT	152	8	116.3
M	150	8	114.7
LOOK	147	8	112.4
HERE	146	8	111.7
MY	145	8	110.9
CAN	132	8	101.0
DO	111	8	84.9
ON	111	8	84.9
UP	111	8	84.9
NOT	110	9	84.1
THREE	107	8	81.8
GOT	103	8	78.8
KNOW	99	8	75.7
IN	98	8	75.0
ARE	96	8	73.4
THERE	94	8	71.9
HAVE	93	8	71.1
WHERE	80	8	61.2
YOUR	80	8	61.2
WANT	79	8	60.4
COME	77	8	58.9
OH	76	8	58.1
SEE	75	8	57.4
WE	83	7	55.5
AT	72	8	55.1
LL	65	8	49.7
FOUR	64	8	48.9
GOING	63	8	48.2
HE	71	7	47.5
GO	60	8	45.9

SEQUENCE	FREQ	SPEAKERS	INDEX
MINE	60	8	45.9
GET	79	6	45.3
FOR	59	8	45.1
RE	66	7	44.2
THEY	55	8	42.1
FIVE	60	7	40.2
WITH	50	8	38.2
LIKE	46	8	35.2
VE	46	8	35.2
HOW	52	7	34.8
SAY	44	8	33.7
NOW	58	6	33.3
BOY	42	8	32.1
OF	42	8	32.1
PLAY	42	8	32.1
WAS	54	6	31.0
PUT	40	8	30.6
LITTLE	44	7	29.4
DID	43	7	28.8
OKAY	50	6	28.7
ALL	42	7	28.1
AM	41	7	27.4
DOWN	41	7	27.4
HEY	40	7	26.8
SAID	34	8	26.0
SHE	44	6	25.2
MAKE	37	7	24.8
SIX	43	6	24.7
CANT	36	7	24.1
SEVEN	42	6	24.1
MRS	41	6	23.5
WHEN	41	6	23.5
WHO	33	7	22.1
BE	32	7	21.4
HIM	32	7	21.4
TOO	28	8	21.4
SO	31	7	20.7
YOURS	31	7	20.7
TEACHER	26	8	19.9
BECAUSE	41	5	19.6
GOOD	29	7	19.4
GIRL	32	6	18.4
WILL	32	6	18.4
OUT	27	7	18.1
-SS	31	6	17.8
EAT	37	5	17.7
DIDNT	23	8	17.6
FINISHED	26	7	17.4
BOY	29	6	16.6
RUN	43	4	16.4

SEQUENCE	FREQ	SPEAKERS	INDEX
RUN	43	4	16.4
DRAW	34	5	16.3
GIVE	34	5	16.3
BIG	24	7	16.1
TEN	28	6	16.1
AWAY	33	5	15.8
SOME	23	7	15.4
YUK	26	6	14.9
JUST	31	5	14.5
GONNA	36	4	13.8
WELL	24	6	13.8
EIGHT	35	4	13.4
WHY	28	5	13.4
TALK	23	6	13.2
LET	27	5	12.9
BLACK	22	6	12.6
MUM	22	6	12.6
COLOUR	18	7	12.0
DOG	21	6	12.0
HAVENT	18	7	12.0
WENT	18	7	12.0
HAD	20	6	11.5
HORSE	30	4	11.5
OFF	20	6	11.5
TELL	20	6	11.5
MORE	29	4	11.1
BOAT	23	5	11.0
HIS	19	6	10.9
NOTHING	19	6	10.9
HOUSE	28	4	10.7
CAT	22	3	10.5
RIGHT	22	5	10.5
THANK	22	5	10.5
YELLOW	22	5	10.5
HER	18	6	10.3
JANET	26	4	9.9
TIME	17	6	9.8
IF	25	4	9.6
JOHN	25	4	9.6
THEM	20	5	9.6
AFTER	16	6	9.2
AH	16	6	9.2
LEAVE	16	6	9.2
NEVER	22	4	8.8
AEROPLANE	15	6	8.6
MUCH	18	5	8.6
NICE	18	5	8.6
TAKE	15	6	8.6
THEN	18	5	8.6
THESE	18	5	8.6

SEQUENCE	FREQ	SPEAKERS	INDEX
THOSE	18	5	8.6
FIRST	22	4	8.4
TREE	22	4	8.4
CARE	17	5	8.1
GOES	17	5	8.1
NINE	21	4	8.0
OUR	14	6	8.0
COMING	16	5	7.6
HELLO	16	5	7.6
LETS	20	4	7.6
ORANGE	16	5	7.6
BACK	13	6	7.5
BOOK	13	6	7.5
KITTEN	26	3	7.5
NAUGHTY	13	6	7.5
HIT	25	3	7.2
FAT	18	4	6.9
ONLY	12	6	6.9
WHICH	14	5	6.7
AGAIN	17	4	6.5
DIGGER	34	2	6.5
NAME	17	4	6.5
ANOTHER	11	6	6.3
BETTY	22	3	6.3
KEEP	11	6	6.3
RED	11	6	6.3
THINK	11	6	6.3
AN	13	5	6.2
CAR	13	5	6.2
MORNING	16	4	6.1
OR	16	4	6.1
SAYS	16	4	6.1
YEAH	21	3	6.0
BEAUTIFUL	20	3	5.7
NEED	12	5	5.7
OVER	12	5	5.7
SOMETHING	12	5	5.7
THING	10	6	5.7
TOOK	10	6	5.7
HA	14	4	5.4
SCHOOL	14	4	5.4
DOING	11	5	5.3
DUMB	11	5	5.3
HAS	11	5	5.3
PLAYING	11	5	5.3
TELLING	11	5	5.3
THROW	11	5	5.3
WATCH	11	5	5.3
WAY	11	5	5.3
CAME	9	6	5.2

SEQUENCE	FREQ	SPEAKERS	INDEX
CAME	9	6	5.7
HURRY	18	3	2.2
WANNA	18	3	2.2
EXCUSE	13	4	5.0
GUESS	13	4	5.0
SHUT	13	4	5.0
WATT	26	2	5.0
VERY	17	3	4.9
HEAD	10	5	4.8
BALL	16	3	4.6
BETTER	12	4	4.6
BLUE	12	4	4.6
FRIEND	12	4	4.6
MAMMA	24	2	4.6
SECOND	12	4	4.6
CATCH	9	5	4.3
DAY	15	3	4.3
DAD	11	4	4.2
EVERYBODY	11	4	4.2
UH	22	2	4.2
ALWAYS	14	3	4.0
FUCK	14	3	4.0
RIDE	14	3	4.0
UM	14	3	4.0
BREAK	10	4	3.8
CRAYON	10	4	3.8
FUNNY	10	4	3.8
HAND	8	5	3.8
LADY	10	4	3.8
LEG	8	5	3.8
LUNCH	10	4	3.8
OTHER	10	4	3.8
PAK	20	2	3.8
PAT	20	2	3.8
PINK	20	2	3.8
STUPID	8	5	3.8
HAT	13	3	3.7
LOVE	13	3	3.7
MAKING	13	3	3.7
SISTER	13	3	3.7
ABOUT	9	4	3.4
ANY	9	4	3.4
BROWN	9	4	3.4
CAUGHT	12	3	3.4
SHOULD	9	4	3.4
STILL	9	4	3.4
STORY	9	4	3.4
BABY	17	2	3.3
EVEN	7	5	3.3
STAY	7	5	3.3

SEQUENCE	FREQ	SPEAKERS	INDEX
TOP	7	5	3.3
FROM	11	3	3.2
DONE	8	4	3.1
GAVE	8	4	3.1
HAPPENED	8	4	3.1
HEAR	8	4	3.1
SIDE	8	4	3.1
SORRY	8	4	3.1
TOUCH	8	4	3.1
DRAWING	10	3	2.9
EGG	15	2	2.9
GRADE	10	3	2.9
HONG	10	3	2.9
MAMA	10	3	2.9
MAN	10	3	2.9
SAM	15	2	2.9
SHOW	6	5	2.9
WALKIE	10	3	2.9
WERE	10	3	2.9
AROUND	7	4	2.7
DORA	14	2	2.7
EATING	7	4	2.7
FIND	7	4	2.7
HOME	7	4	2.7
KID	14	2	2.7
US	7	4	2.7
KING	9	3	2.6
MADE	9	3	2.6
OPEN	9	3	2.6
TA	9	3	2.6
THIRD	9	3	2.6
HUH	13	2	2.5
APPLE	12	2	2.3
BEEN	6	4	2.3
CALL	6	4	2.3
CHAIR	6	4	2.3
CHOCOLATE	8	3	2.3
DICK	12	2	2.3
HAIR	6	4	2.3
HAVING	6	4	2.3
HOLD	8	3	2.3
HOT	6	4	2.3
JUMP	8	3	2.3
LOST	8	3	2.3
MAT	12	2	2.3
ROOM	12	2	2.3
SPEAK	6	4	2.3
THUMBKIN	8	3	2.3
TOILET	6	4	2.3
EASTER	11	2	2.1

SEQUENCE	FREQ	SPEAKERS	INDEX
EASTER	11	2	2.1
MATE	11	2	2.1
SHALL	11	2	2.1
BAND	7	3	2.0
DOOR	7	3	2.0
EASY	7	3	2.0
FULL	7	3	2.0
LOOKING	7	3	2.0
OUCH	7	3	2.0
OUTSIDE	7	3	2.0
OW	7	3	2.0
QNDUN	7	3	2.0
REALLY	7	3	2.0
SOMEONE	7	3	2.0
STOOD	7	3	2.0
BY	5	4	1.9
COULD	5	4	1.9
COUSIN	5	4	1.9
DRAGAN	10	2	1.9
FISH	10	2	1.9
GETTING	5	4	1.9
MEAN	5	4	1.9
QNAME	5	4	1.9
ROBERT	10	2	1.9
TONY	10	2	1.9
WAIT	10	2	1.9
ZIP	10	2	1.9
ALLOWED	6	3	1.7
BAG	6	3	1.7
BROKEN	6	3	1.7
BROTHER	6	3	1.7
DOES	6	3	1.7
FLOWER	6	3	1.7
KICK	9	2	1.7
LINE	9	2	1.7
PLEASE	6	3	1.7
PUTTING	6	3	1.7
QUICKLY	6	3	1.7
READ	6	3	1.7
STOP	6	3	1.7
TALLMAN	9	2	1.7
TURN	6	3	1.7
BOBBY	8	2	1.5
BOING	8	2	1.5
CHASEY	8	2	1.5
GREEN	8	2	1.5
MEN	8	2	1.5
SAT	8	2	1.5
SAN	8	2	1.5
STAR	8	2	1.5

SEQUENCE	FREQ	SPEAKERS	INDEX
BIN	5	3	1.4
BUM	5	3	1.4
FINISH	5	3	1.4
INSIDE	5	3	1.4
JENNIFER	5	3	1.4
NIGHT	5	3	1.4
OWN	5	3	1.4
PUSH	5	3	1.4
THAN	5	3	1.4
WRONG	5	3	1.4
BANANA	7	22	1.3
BIT	7	22	1.3
CARROT	7	22	1.3
DAVID	7	22	1.3
EYE	7	22	1.3
FOOTBALL	7	22	1.3
KONG	7	22	1.3
MILK	7	22	1.3
MONEY	7	22	1.3
STAMP	7	22	1.3
WALL	7	22	1.3
BASH	6	22	1.1
COLOURING	6	22	1.1
CUT	6	22	1.1
ENGLISH	6	22	1.1
FALL	6	22	1.1
HUNDRED	6	22	1.1
LEANNE	6	22	1.1
LIVE	6	22	1.1
MANY	6	22	1.1
READY	6	22	1.1
SCRIBBLED	6	22	1.1
SH	6	22	1.1
SHOT	6	22	1.1
SUPPOSED	6	22	1.1
WHITE	6	22	1.1
BEE	5	22	1.0
CLASS	5	22	1.0
DEIDRE	5	22	1.0
DRAGON	5	22	1.0
EH	5	22	1.0
ELVIS	5	22	1.0
MARK	5	22	1.0
MONKEY	5	22	1.0
NIP	5	22	1.0
O	5	22	1.0
PART	5	22	1.0
RABBIT	5	22	1.0
SIA	5	22	1.0
SUN	5	22	1.0

TWO WORD SEQUENCES OF THE MACEDONIAN SAMPLE OF UPPER FIVE YEAR OLD
CHILDREN IN ORDER OF INDEX VALUE

SEQUENCE	FREQ	SPEAKERS	INDEX
I M	140	8	676.4
? WHAT	122	8	418.9
I LL	52	8	231.0
? WHERE	62	8	212.9
I KNOW	40	8	181.6
I CAN	45	7	178.7
I DONT	41	7	162.8
I AM	38	7	150.9
I VE	32	7	127.1
ONE TWO	81	7	120.2
? HOW	39	7	117.2
IT S	60	8	112.0
THAT S	80	8	109.5
YOU RE	30	7	90.0
I WANT	26	6	88.5
I HAVE	26	5	73.8
YOU KNOW	26	6	66.8
? YOU	18	7	54.1
YOU CAN	16	7	48.0
? CAN	18	6	46.4
I WAS	27	3	46.0
YOU HAVE	15	7	45.0
LOOK AT	52	8	44.7
? DO	17	6	43.8
YOU ARE	17	6	43.7
YOU SAY	25	4	42.8
TWO THREE	45	7	42.4
? WHO	19	5	40.8
I GOT	14	3	39.7
I LIKE	10	7	39.7
YOU WANT	18	5	38.6
I HAVENT	11	6	37.4
I CANT	13	5	36.9
? WHY	21	4	36.1
YOU DONT	13	6	33.4
THIS IS	33	6	32.6
I NEVER	17	3	28.9
I DIDNT	10	5	28.4
DO YOU	40	8	26.0
THIS ONE	26	6	25.7
S A	18	7	22.7
YOU DO	10	5	21.4
DONT I NOW	26	7	20.2
WANT TO	42	8	19.4
M NOT	25	7	19.2
-SP AND	21	5	19.0
WHAT DO	22	6	19.0
? WHICH	11	4	18.9
YOU CANT	11	4	18.9
AND I	20	6	18.3

SEQUENCE	FREQ	SPEAKERS	INDEX
S NOT	17	6	18.3
ARE YOU	37	7	18.2
GOING TO	49	8	18.1
THREE FOUR	32	7	17.5
WHAT S	20	6	17.3
? ARE	8	5	17.2
IN THE	34	7	17.1
YOU GOT	8	5	17.1
I WANNA	10	3	17.0
TWO ONE	27	4	14.5
I DO	5	5	14.2
YOU DID	8	4	13.7
YOU DIDNT	8	4	13.7
YOU SAID	8	4	13.7
I HAD	6	4	13.6
NO I	17	5	13.5
THAT ONE	15	5	12.8
YOU GET	7	4	12.0
YOU GOING	7	4	12.0
I SEE	7	3	11.9
I GOING	26	4	11.4
I EAT	10	2	11.3
IT IN	8	6	11.2
YOU THIS	13	2	11.1
COME ON	32	6	10.8
S THE	12	5	10.8
FOUR FIVE	38	6	10.7
TO ME	12	5	10.6
? BUT	6	4	10.3
YOU LIKE	8	3	10.3
I NEED	6	3	10.2
THE TEACHER	11	4	10.1
S MY	11	5	9.9
DO IT	15	8	9.7
TO PLAY	13	4	9.2
I SAW	8	2	9.1
? IS	7	3	9.0
? WANT	7	3	9.0
ME TOO	11	7	9.0
S MINE	10	5	9.0
YOU VE	7	3	9.0
HAVE TO	26	5	8.8
ON THE	18	6	8.8
? I	5	4	8.6
I DID	5	3	8.5
TO THE	8	6	8.5
WHERE IS	29	5	8.5
HERE I	26	3	8.3
TO DO	11	4	7.8
? DID	6	3	7.7

SEQUENCE	FREQ	SPEAKERS	INDEX
? DID	6	3	7.7
IT IS	8	4	7.5
WHERE S	18	7	7.4
THE OTHER	3	4	7.3
VE GOT	31	7	7.3
GOT A	16	6	7.2
S WHAT	8	5	7.2
NO NO	11	4	7.0
CAN I	18	4	6.9
IS IT	15	4	6.9
THIS MORNING	14	3	6.9
TO BE	13	3	6.9
DONT WANT	12	5	6.7
IT ON	7	4	6.6
YOU DOING	5	3	6.4
YOU PUT	5	3	6.4
YOU THINK	5	3	6.4
WHAT IS	14	3	6.1
? SHALL	7	2	6.0
TO GO	8	4	5.6
M TELLING	10	5	5.5
THE LITTLE	6	4	5.5
THERE S	13	6	5.4
CAN SEE	11	5	5.3
? HUH	4	2	5.3
CAN YOU	9	6	5.2
WHAT ARE	9	4	5.2
IS MY	11	4	5.1
ONE ONE	6	4	5.1
THAT WE	6	5	5.1
YOU LL	6	2	5.1
YOU WERE	6	2	5.1
AND SEE	11	3	5.0
IT UP	7	3	4.9
TO GET	7	4	4.9
THEY RE	20	6	4.8
IT WAS	10	2	4.7
ONE FOR	11	2	4.7
IS THE	8	5	4.6
THE BIG	5	4	4.6
THE GIRL	10	2	4.6
DO THAT	11	5	4.5
ONE OF	7	3	4.5
UP THE E	14	4	4.5
* ONE	12	3	4.3
? UH	5	2	4.3
AND MY	7	4	4.3
S THIS	6	4	4.3
ONE RE	5	4	4.2
TO DRAW	12	2	4.2

SEQUENCE	FREQ	SPEAKERS	INDEX
TO HAVE	8	3	4.2
TO MAKE	6	4	4.2
A GIRL	7	3	4.1
SEE THE	15	5	4.1
THE FAT	6	3	4.1
GOT TWO	13	4	3.9
FIVE SIX	20	5	3.8
S GOING	7	3	3.8
AND ONE	6	4	3.7
* YES	15	2	3.6
KNOW WHAT	8	6	3.5
TO TELL	5	4	3.5
WHAT HAPPENED	6	4	3.5
COME HERE	12	5	3.4
THE CAT	5	3	3.4
THE DOOR	5	3	3.4
THE TOILET	5	3	3.4
YES I	7	4	3.3
YES YES	7	4	3.3
MY DAD	10	3	3.2
ONE NOW	5	3	3.2
S THAT	6	3	3.2
TO YOU	6	3	3.2
TWO KITTEN	12	2	3.2
WE RE	12	5	3.2
AND ME	5	4	3.1
DONT HAVE	7	4	3.1
HAVE A	9	5	3.1
THAT THAT	6	3	3.1
MY MUM	7	4	3.0
* I	8	3	2.9
A NICE	5	3	2.9
IS THAT	5	5	2.9
WHERE ARE	10	5	2.9
IS THUMBKIN	8	3	2.8
IT WENT	6	2	2.8
-SP ON	5	3	2.7
HERE S	5	5	2.7
S EASY	5	3	2.7
S GOT	5	3	2.7
THE FIRST	6	2	2.7
THE KING	6	2	2.7
CAN DO	9	3	2.6
GOT IT	7	5	2.6
HERE WE	6	4	2.6
* ?	7	3	2.5
LIKE THIS	15	5	2.5
ME A	7	3	2.5
THIS ON	5	3	2.5
THIS DAY	5	3	2.5

SEQUENCE	FREQ	SPEAKERS	INDEX
THIS WAY	5	3	2.5
AND AFTER	8	2	2.4
DO THIS	10	3	2.4
IS THIS	7	3	2.4
NO IT	6	3	2.4
NO YOU	6	3	2.4
TWO TWO	6	3	2.4
A CAR	6	2	2.3
A HORSE	6	2	2.3
A LITTLE	6	2	2.3
AND A	5	3	2.3
AND GET	5	3	2.3
DONT TOUCH	7	3	2.3
M MAKING	7	4	2.3
THE BALL	5	2	2.3
GO TO	10	5	2.2
GOT TO	7	4	2.1
IS TALLMAN	9	2	2.1
A TEACHER	5	2	2.0
DONT YOU	9	2	2.0
* DORA	8	2	1.9
* TWO	8	2	1.9
AT THE	6	6	1.9
AT THIS	18	2	1.9
GOT ONE	5	5	1.9
HOW MUCH	14	4	1.9
ME SEE	8	2	1.9
MY HOUSE	9	2	1.9
MY ROOM	9	3	1.9
ON IT	6	4	1.9
AND THE	6	2	1.8
S BEAUTIFUL	5	2	1.8
SIX SEVEN	19	4	1.8
DID YOU	15	4	1.7
DONT CARE	5	3	1.7
HAVE IT	5	5	1.7
ON A	7	3	1.7
WHAT IT	6	2	1.7
YES YOU	5	3	1.7
HE S	9	4	1.6
HERE YOU	5	3	1.6
LOOK LOOK	5	3	1.6
AND YOU	5	2	1.5
HE SAID	11	3	1.5
LIKE THAT	9	5	1.5
RE LOT	9	4	1.5
* JOHN	6	2	1.4
* NO	6	2	1.4
HOW ARE	14	3	1.4
IN OUR	5	4	1.4

SEQUENCE	FREQ	SPEAKERS	INDEX
IS A	6	2	1.4
KNOW I	5	4	1.4
WE ARE	9	3	1.4
AT MINE	5	5	1.3
KNOW WHERE	6	3	1.3
ON THIS	8	2	1.3
SHE S	11	5	1.3
* THREE	5	2	1.2
CAN RIDE	6	2	1.2
COME LOOK	7	3	1.2
FOR ME	7	4	1.2
GET IT	7	4	1.2
GET OUT	7	4	1.2
YOU	6	3	1.2
MAKE A	10	5	1.2
MRS WATT	26	2	1.2
NOT ALLOWED	5	3	1.2
PUT IT	10	4	1.2
UP THE	5	3	1.2
WANT ME	7	3	1.2
IN THERE	5	3	1.1
LOOK AND	5	2	1.1
M FINISHED	5	2	1.1
MY EASTER	5	2	1.1
OF THE	9	4	1.1
SEVEN EIGHT	16	3	1.1
THANK YOU	22	5	1.1
THEY ARE	7	4	1.1
GET DOWN	6	4	1.0
GIVE ME	13	5	1.0
LL GET	7	3	1.0
OH NO	6	3	1.0
THERE AND	5	3	1.0
IN MY	6	2	0.9
LL DO	6	3	0.9
WANT IT	5	3	0.9
WITH THE	6	4	0.9
ALL RIGHT	15	2	0.8
GOT ANY	5	2	0.8
HAVE SOME	6	2	0.8
HAVE THIS	6	2	0.8
OH I	5	3	0.8
RE GONNA	6	3	0.8
RUN AWAY	26	2	0.8
WE HAVE	5	3	0.8
BECAUSE I	9	4	0.7
EIGHT NINE	13	4	0.7
FOR YOU	8	2	0.7
GO IN	5	3	0.7
HAVENT GOT	10	6	0.7

SEQUENCE	FREQ	SPEAKERS	INDEX
HAVENT GOT	10	6	0.7
LL BE	5	3	0.7
LL TELL	5	3	0.7
WITH YOU	5	4	0.7
ALL THE	8	3	0.6
DOWN HERE	6	4	0.6
DOWN THERE	7	3	0.6
FOR THE	5	3	0.6
GET LOST	7	2	0.6
GET YOU	5	3	0.6
GO ON	7	2	0.6
LET ME	13	4	0.6
PLAY WITH	6	3	0.6
PUT THIS	7	3	0.6
RE A	5	3	0.6
SAY THAT	6	3	0.6
WHEN I	9	3	0.6
GO DOWN	6	2	0.5
HE HE	5	2	0.5
HOW COME	7	2	0.5
LL DRAW	5	2	0.5
NOW I	5	3	0.5
WE WENT	5	2	0.5
YOURS IS	8	3	0.5
GET YOUR	5	2	0.4
LEAVE IT	9	5	0.4
MINE S	5	2	0.4
RE COMING	5	2	0.4
WILL YOU	6	4	0.4
WITH A	5	2	0.4
BECAUSE WE	5	3	0.3
BECAUSE YOU	9	2	0.3
BUT I	5	4	0.3
DRAW A	7	3	0.3
IF YOU	12	3	0.3
LITTLE DOG	5	2	0.3
NINE TEN	10	4	0.3
NOW YOU	5	2	0.3
OF YOU	5	2	0.3
OUT OF	5	3	0.3
TELL YOU	7	4	0.3
BIG ONE	5	2	0.2
BOAT -SP	10	2	0.2
EXCUSE ME	13	4	0.2
GIRL -SP	7	2	0.2
GONNA HAVE	6	3	0.2
HA HA	8	4	0.2
JAN JOHN	8	2	0.2
JUST WHAT	6	2	0.2
KITTEN -SP	12	2	0.2

SEQUENCE	FREQ	SPEAKERS	INDEX
SAID YOU	5	2	0.2
SHOT UP	13	4	0.2
TELL ME	6	3	0.2
WELL I	8	2	0.2
WHICH ONE	8	4	0.2
BOOK -SP	5	3	0.1
BREAK IT	7	4	0.1
CRAYON -SP	6	3	0.1
DOG RUN	5	2	0.1
FAT CAT	5	2	0.1
HAD A	6	2	0.1
HAND -SP	5	4	0.1
HIT YOU	6	2	0.1
HURRY UP	13	2	0.1
JOHN CODE	7	2	0.1
KID -SP	14	2	0.1
LETS GO	5	2	0.1
MAKING A	5	3	0.1
MUCH IS	7	2	0.1
NAME IS	5	3	0.1
NAUGHTY BOY	5	2	0.1
OVER THERE	6	4	0.1
PAT DIGGER	9	2	0.1
THEN YOU	5	2	0.1
VERY WELL	14	2	0.1
WATT I	8	2	0.1
WATT MRS	7	2	0.1
WHY DID	5	2	0.1
BEE -SP	5	2	0.0
BOING BOING	5	2	0.0
CAUGHT ME	5	3	0.0
EGG -SP	6	2	0.0
EYE -SP	6	2	0.0
HONG KONG	5	2	0.0
KING -SS	5	2	0.0
MADE A	5	2	0.0
HAMMA HAMMA	5	2	0.0
OPEN IT	5	2	0.0
RIDE A	6	2	0.0
SHALL I	9	2	0.0
SUPPOSED TO	6	2	0.0

THREE WORD SEQUENCES OF THE MACEDONIAN SAMPLE OF UPPER FIVE YEAR
OLD CHILDREN IN ORDER OF INDEX VALUE

SEQUENCE	FREQ	SPEAKERS	INDEX
I M NOT	25	7	13.2
I M GOING	26	4	17.5
? WHAT S	14	6	30.6
I M TELLING	10	5	32.3
ONE TWO THREE	32	7	30.6
? WHERE IS	29	5	29.5
? WHAT DO	19	4	28.8
I DONT KNOW	19	7	20.7
I VE GOT	22	7	18.7
? WHERE S	13	7	18.5
I M MAKING	7	3	13.6
? WHERE ARE	10	5	10.2
ONE TWO ONE	26	3	9.0
TWO THREE FOUR	30	7	8.5
? WHAT ARE	7	3	8.4
I M FINISHED	5	2	5.5
I DONT WANT	10	4	6.2
I WANT TO	18	4	6.1
I M A	3	3	5.8
I M GETTING	3	3	5.8
? HOW MUCH	12	4	5.4
IT S NOT	12	4	5.1
? HOW ARE	14	3	4.7
I LL GET	7	3	4.7
I CAN DO	8	3	4.1
I LL DO	6	3	4.1
? WHAT HAPPENED	3	3	3.6
? WHAT IS	3	3	3.6
? DO YOU	14	6	3.5
I LL TELL	5	3	3.4
THAT S A	7	4	2.9
THAT S MY	7	4	2.9
THAT S WHAT	7	4	2.9
I M CAUGHT	2	2	2.6
I M DOING	2	2	2.6
? WHAT COLOUR	3	2	2.4
? WHAT DID	3	2	2.4
I DONT CARE	5	3	2.3
I LL D AW	5	2	2.3
THREE FOUR FIVE	25	5	2.1
I LL GO	3	3	2.0
I DONT HAVE	4	3	1.9
WHAT DO YOU	20	5	1.8
I KNOW I	3	3	1.6
THAT S EASY	3	3	1.6
I VE FINISHED	4	3	1.5
LOOK AT THIS	15	2	1.5
I LL PUT	3	2	1.4
THIS IS MY	11	4	1.4
YOU HAVE TO	8	4	1.4

SEQUENCE	FREQ	SPEAKERS	INDEX
IT S A	4	3	1.3
YOU RE A	5	3	1.3
? CAN I	9	3	1.2
FOUR FIVE SIX	20	5	1.0
I CAN MAKE	3	2	1.0
I CAN SEE	3	2	1.0
IT S MINE	3	3	1.0
I GOING TO	24	4	1.0
THAT S BEAUTIFUL	5	2	1.0
TWO ONE TWO	24	3	1.0
YOU KNOW WHAT	5	3	1.0
YOU RE GONNA	4	3	1.0
? HOW COME	4	2	0.9
? YOU KNOW	6	3	0.9
I LL MAKE	2	2	0.9
IT S GOING	4	2	0.9
IT S SUPPOSED	4	2	0.9
THAT S WHY	3	3	0.9
YOU WANT TO	6	4	0.9
I HAVE A	4	3	0.8
I HAVE SOME	6	2	0.8
I HAVE TO	4	3	0.8
THAT S MINE	4	2	0.8
? ARE YOU	8	5	0.7
DO YOU WANT	9	3	0.7
I CAN CATCH	2	2	0.7
I KNOW WHAT	2	2	0.7
I KNOW YOU	2	2	0.7
LOOK AT MINE	4	4	0.7
LOOK AT THAT	4	4	0.7
DO YOU KNOW	6	4	0.6
I DONT WANNA	2	2	0.6
THAT S NOT	3	2	0.6
THAT S THE	3	2	0.6
YOU SAY THAT	5	3	0.6
? CAN YOU	4	3	0.5
ARE YOU THIS	13	2	0.5
I HAVENT GOT	5	3	0.5
WANT TO PLAY	7	4	0.5
? WHICH ONE	6	4	0.4
HERE I AM	24	2	0.4
IT S COMING	2	2	0.4
NO I M	8	4	0.4
THAT S ALL	2	2	0.4
WHAT S THIS	6	4	0.4
YOU WANT ME	6	2	0.4
? WHY DID	4	2	0.3
CAN I HAVE	10	4	0.3
I LIKE THAT	3	3	0.3
I WANT A	2	2	0.3

SEQUENCE	FREQ	SPEAKERS	INDEX
I WANT A	2	2	0.3
I WANT SOME	2	2	0.3
VE GOT A	3	3	0.3
YOU DO IT	4	4	0.3
YOU DONT KNOW	4	3	0.3
YOU KNOW UM	2	2	0.3
YOU KNOW WHERE	2	2	0.3
YOU RE LOOKING	2	2	0.3
YOU RE NOT	2	2	0.3
YOU THIS MORNING	13	2	0.3
? CAN WE	2	2	0.2
ARE YOU DOING	3	3	0.2
FIVE SIX SEVEN	15	4	0.2
GOING TO GET	5	4	0.2
I DIDNT HEAR	3	2	0.2
I LIKE THE	2	2	0.2
I WAS PLAYING	2	2	0.2
LOOK AT HIM	2	2	0.2
LOOK AT IT	2	2	0.2
LOOK AT THE	2	2	0.2
NO I DONT	4	3	0.2
THIS IS MINE	3	2	0.2
WHAT ARE YOU	3	4	0.2
WHAT S THAT	4	3	0.2
WHERE IS THUMBXIN	3	3	0.2
YOU CAN KEEP	2	2	0.2
YOU CAN SEE	2	2	0.2
YOU HAVE A	2	2	0.2
YOU SAY I	2	2	0.2
YOU SAY WHAT	2	2	0.2
YOU VE GOT	7	3	0.2
? DID YOU	5	2	0.1
? SHALL I	7	2	0.1
? WANT TO	5	2	0.1
AND I KNOW	2	2	0.1
AND I LL	2	2	0.1
ARE YOU DRAWING	2	2	0.1
DO YOU LIKE	2	2	0.1
DO YOU THINK	3	2	0.1
DONT HAVE TO	7	4	0.1
DONT KNOW HOW	2	2	0.1
DONT NOW WHERE	3	2	0.1
DONT WANT TO	5	3	0.1
GOING TO MAKE	2	2	0.1
HOW ARE YOU	14	2	0.1
I HAVE TWO	2	2	0.1
I WANNA GO	4	2	0.1
IN THE BAG	3	2	0.1
IN THE BIN	3	2	0.1
IN THE DOOR	3	2	0.1

SEQUENCE	FREQ	SPEAKERS	INDEX
IS MY HOUSE	6	2	0.1
M NOT GOING	3	2	0.1
ON THE HEAD	3	2	0.1
S A BIG	2	2	0.1
SIX SEVEN EIGHT	14	3	0.1
THIS IS THE	2	2	0.1
TO PLAY WITH	5	3	0.1
TO THE TOILET	4	2	0.1
TWO KITTEN --SP	11	2	0.1
VE GOT TO	3	3	0.1
VE GOT TWO	8	2	0.1
WANT TO GO	2	2	0.1
WANT TO HAVE	2	2	0.1
WANT TO SEE	2	2	0.1
WE RE NOT	6	3	0.1
WHAT IS IT	11	2	0.1
WHERE IS TALLMAN	9	2	0.1
WHERE S MY	4	2	0.1
YOU DO THAT	3	2	0.1
YOU GOING TO	4	2	0.1
YOU GOT IT	2	2	0.1
YOU LIKE IT	3	2	0.1
YOU WANT A	2	2	0.1
* ? WHERE	4	3	0.0
* I CAN	5	2	0.0
* ONE TWO	3	2	0.0
--SP ON IT	4	2	0.0
A LITTLE BIT	4	2	0.0
AND AFTER I	3	2	0.0
AND GET IT	2	2	0.0
AND ONE FOR	3	2	0.0
AND SEE THE	4	2	0.0
BECAUSE YOU SAY	3	2	0.0
BUT I M	2	2	0.0
CAN DO THAT	2	2	0.0
CAN RIDE A	6	2	0.0
COME LOOK AND	4	2	0.0
DID YOU SAY	3	2	0.0
EIGHT NINE TEN	10	4	0.0
GIVE ME A	3	2	0.0
GIVE ME ONE	4	3	0.0
GO DOWN THERE	5	2	0.0
GO TO THE	2	2	0.0
GOT IT IN	3	3	0.0
GOT ONE OF	2	2	0.0
HA HA HA	4	2	0.0
HAND --SP UP	2	2	0.0
HAVE IT AT	2	2	0.0
HAVE TO BE	2	2	0.0
HAVE TO TELL	2	2	0.0

SEQUENCE	FREQ	SPEAKERS	INDEX
HAVE TO TELL	2	22	0.0
HAVENT GOT ANY	5	22	0.0
HERE S THE	2	22	0.0
HERE WE ARE	4	22	0.0
HOW MUCH IS	7	22	0.0
IN OUR CLASS	3	22	0.0
JANET JOHN COME	7	22	0.0
JOHN COME LOOK	6	22	0.0
KNOW WHERE IT	2	22	0.0
LET ME SEE	8	22	0.0
LL DO IT	3	22	0.0
LL DO THAT	3	22	0.0
LL GET YOU	4	22	0.0
LL TELL YOU	3	22	0.0
LOOK AND SEE	4	22	0.0
M TELLING THE	2	22	0.0
M TELLING YOU	2	22	0.0
M TELLING YOUR	2	22	0.0
MRS WATT I	8	22	0.0
MRS WATT MRS	7	22	0.0
MY DAD -SS	2	22	0.0
NO IT S	3	22	0.0
NOT ALLOWED TO	2	22	0.0
NOW I M	4	22	0.0
ONE OF THOSE	3	22	0.0
ONE ONE ONE	2	22	0.0
ONE ONE TWO	2	22	0.0
OUT OF THE	4	22	0.0
PAT DIGGER SAM	3	22	0.0
PUT IT ON	3	22	0.0
PUT THIS ON	2	22	0.0
RE GONNA HAVE	2	22	0.0
RIDE A HORSE	6	22	0.0
S GOING TO	5	22	0.0
S WHAT I	2	22	0.0
SEE THE BOAT	4	22	0.0
SEVEN EIGHT NINE	12	22	0.0
SHE S GOT	2	22	0.0
SHE S WATCHING	2	22	0.0
SUPPOSED TO BE	5	22	0.0
THAT THAT S	3	22	0.0
THE FAT CAT	4	22	0.0
THERE S A	3	22	0.0
THEY ARE NOT	3	22	0.0
THEY RE COMING	3	22	0.0
TO DO THAT	2	22	0.0
TO DRAW A	3	22	0.0
TO GO TO	2	22	0.0
TO TELL YOU	3	22	0.0
WANT ME TO	3	22	0.0
WATT MRS WATT	7	22	0.0
WHAT IS THAT	2	22	0.0
WHEN I WAS	5	22	0.0
WHERE ARE ALL	3	22	0.0
WHILE ARE THEY	2	22	0.0
WHERE IS IT	3	22	0.0
WHERE IS RINGMAN	2	22	0.0
WHERE S THE	2	22	0.0
WHY DID YOU	5	22	0.0
YOU GET A	2	22	0.0
YOU LL BE	2	22	0.0
YOU LL SEE	2	22	0.0
YOU PUT THIS	3	22	0.0
YOURS IS YUK	2	22	0.0

APPENDIX 12

ONE, TWO AND THREE WORD
SEQUENCES OF THE VIETNAMESE
SAMPLE OF UPPER FIVE YEAR
OLD CHILDREN IN ORDER OF
INDEX VALUE

EXPLANATORY NOTES

1. Full lists are also available of single words and sequences in alphabetic and raw frequency order. It is regretted that space forbids their inclusion in this report.
2. In order to provide important information on such matters as number of possessives, plurals, questions, etc. used, relevant non-alphabetic symbols are listed along with other words and morphemes on the following pages.
3. For details on the calculation of an Index Number for single words and two or three word sequences see pages 17 and 20, respectively. Index numbers for three word sequences should be multiplied by 100 to make them comparable with the index numbers used on the other lists.
4. For details on the interpretation of all symbols used in the Print out see Appendix 4: Instructions for Coding.

SINGLE WORDS USED OF THE VIETNAMESE SAMPLE OF UPPER FIVE YEAR
OLD CHILDREN IN ORDER OF INDEX VALUE

SEQUENCE	FREQ	SPEAKERS	INDEX
THE	622	10	624. 0
ONE	432	10	433. 4
YES	408	10	409. 3
TWO	365	10	366. 2
I	361	10	362. 2
NO	331	9	298. 9
THREE	294	10	295. 0
AND	287	10	287. 9
A	322	8	258. 4
?	253	10	253. 8
IS	249	9	224. 8
HERE	219	1	219. 7
FOUR	221	9	199. 6
-SP	211	9	190. 5
*	240	7	168. 6
FIVE	183	9	165. 2
COME	165	9	149. 0
YOU	160	9	144. 5
SEE	150	9	135. 4
S	166	8	133. 2
CAN	142	9	128. 2
LITTLE	139	9	125. 5
SIX	137	9	123. 7
LOOK	117	10	117. 4
THIS	123	9	111. 1
OH	135	8	108. 4
IT	104	8	83. 5
ON	94	8	75. 4
IN	82	9	74. 0
WHAT	101	7	70. 9
ME	69	9	62. 3
DOG	67	9	60. 5
UP	79	7	55. 5
BOY	58	9	52. 4
MY	64	8	51. 4
SEVEN	72	7	50. 6
HE	58	8	46. 6
AM	56	8	44. 9
WATER	52	8	41. 7
COLOUR	48	8	38. 5
ARE	46	8	36. 9
SITTING	44	8	35. 3
EIGHT	50	7	35. 1
GO	48	7	33. 7
DOWN	53	6	31. 9
OKA	52	6	31. 3
YELLOW	38	8	30. 5
HOUSE	43	7	30. 2
DUCK	37	8	29. 7
KNOW	42	7	29. 5

SEQUENCE	FREQ	SPEAKERS	INDEX
ELEPHANT	36	8	28.9
GOT	36	8	28.9
CAT	35	8	28.1
DONT	40	7	28.1
MAN	40	7	28.1
PLAY	40	7	28.1
BALL	46	6	27.7
RUN	46	6	27.7
M	34	8	27.3
VERY	37	7	26.0
WHERE	28	9	25.3
THAT	35	7	24.6
FOR	40	6	24.1
BLACK	28	8	22.5
GREEN	28	8	22.5
KITTEN	44	5	22.1
DOOR	30	7	21.1
GOOD	35	6	21.1
NINE	35	6	21.1
PLEASE	39	5	19.6
AT	24	8	19.3
NOT	24	8	19.3
EATING	38	5	19.1
MISS	38	5	19.1
FINISHED	27	7	19.0
HAVE	27	7	19.0
HORSE	31	6	18.7
HELLO	37	5	18.6
GIRL	26	7	18.3
CAR	36	5	18.1
TEN	30	6	18.1
MANY	25	7	17.6
QNOUN	35	5	17.6
RABBIT	34	5	17.1
TREE	26	6	15.7
FLOWER	25	6	15.0
TO	21	7	14.7
INSIDE	29	5	14.5
MORE	18	8	14.4
NOW	24	6	14.4
PIG	27	5	13.5
LION	19	7	13.3
LOOKING	19	7	13.3
QUICK	32	4	12.8
RED	18	7	12.6
YOUR	21	6	12.6
AEROPLANE	20	6	12.0
BIG	20	6	12.0
STANDING	20	6	12.0
AH	28	4	11.2

SEQUENCE	FREQ	SPEAKERS	INDEX
AH	28	4	11.2
JUMP	28	4	11.2
WANT	16	7	11.2
FISH	18	6	10.8
MOUSE	21	5	10.5
THANK	15	7	10.5
WITH	15	7	10.5
YEAH	17	6	10.2
GOING	14	7	9.8
-SS	16	6	9.6
PUT	24	4	9.6
BOAT	23	4	9.2
BIRD	29	5	8.7
SOME	17	5	8.5
PLAYING	14	6	8.4
RUNNING	14	6	8.4
BE	13	6	7.8
WE	13	6	7.8
MOTHER	19	4	7.6
KANGAROO	18	4	7.2
WALKING	12	6	7.2
BLUE	14	5	7.0
APPLE	17	4	6.8
HOW	17	4	6.8
WINDOW	17	4	6.8
BOOK	16	4	6.4
DO	16	4	6.4
AN	12	5	6.0
BICYCLE	15	4	6.0
COW	12	5	6.0
KOALA	20	3	6.0
LIKE	12	5	6.0
SUN	12	5	6.0
THEY	12	5	6.0
TAIL	11	5	5.5
OPEN	18	3	5.4
HIT	13	4	5.2
MRS	13	4	5.2
SNAKE	13	4	5.2
THERE	13	4	5.2
BABY	17	3	5.1
SHUT	10	5	5.0
ALL	8	6	4.8
MONKEY	12	4	4.8
OUT	12	4	4.8
STOP	12	4	4.8
DOING	9	5	4.5
FAT	15	3	4.5
GRASS	15	3	4.5
HOME	9	5	4.5

SEQUENCE	FREQ	SPEAKERS	INDEX
MAKE	15	3	4.5
RE	9	5	4.5
CHAIR	11	4	4.4
CHICKEN	11	4	4.4
FLYING	11	4	4.4
BEAUTIFUL	14	3	4.2
BROWN	14	3	4.2
ELEVEN	10	4	4.0
HUNG	20	2	4.0
ORANGE	10	4	4.0
PAPER	8	5	4.0
WAY	10	4	4.0
WHITE	10	4	4.0
BAA	13	3	3.9
FLOOR	13	3	3.9
WILL	13	3	3.9
CATCH	19	2	3.8
DOO	19	2	3.8
GIVE	9	4	3.6
OOH	9	4	3.6
TABLE	9	4	3.6
BACK	7	5	3.5
BAG	7	5	3.5
EAT	7	5	3.5
WEIR	17	2	3.4
QUICKLY	11	3	3.3
BOX	8	4	3.2
HA	8	4	3.2
HAS	8	4	3.2
MORNING	16	2	3.2
BEE	10	3	3.0
OLD	10	3	3.0
OUTSIDE	10	3	3.0
SIR	10	3	3.0
TAKE	6	5	3.0
BELL	14	2	2.8
BUS	7	4	2.8
HEY	7	4	2.8
LONG	7	4	2.8
STAR	7	4	2.8
SWAN	7	4	2.8
TWELVE	7	4	2.8
CUT	9	3	2.7
EMU	9	3	2.7
FIRE	9	3	2.7
HAND	9	3	2.7
HIS	9	3	2.7
ISNT	9	3	2.7
LADDER	9	3	2.7
SCISSORS	9	3	2.7

SEQUENCE	FREQ	SPEAKERS	INDEX
SCISSORS	9	3	2.7
CARD	13	2	2.6
GLUE	13	2	2.6
LE	13	2	2.6
BUTTERFLY	6	4	2.4
FLY	8	3	2.4
FROG	8	3	2.4
KICK	12	2	2.4
PENCIL	12	2	2.4
SHUTTING	6	4	2.4
TONGUE	8	3	2.4
TURN	12	2	2.4
TWENTY	8	3	2.4
FAST	7	3	2.1
FINISH	7	3	2.1
FOOT	7	3	2.1
FOURTEEN	7	3	2.1
GET	7	3	2.1
SHELL	7	3	2.1
TIME	7	3	2.1
VE	7	3	2.1
FOX	10	2	2.0
NAME	5	4	2.0
PUPPY	10	2	2.0
SHEEP	5	4	2.0
ANY	6	3	1.8
AWAY	6	3	1.8
BIN	6	3	1.8
CAKE	6	3	1.8
DONKEY	6	3	1.8
FATHER	9	2	1.8
FRIEND	6	3	1.8
NOSE	6	3	1.8
WOOL	6	3	1.8
CANT	8	2	1.6
COMING	8	2	1.6
FINGER	8	2	1.6
TOO	8	2	1.6
BIT	5	3	1.5
GONE	5	3	1.5
HOLDING	5	3	1.5
JUMPING	5	3	1.5
PARK	5	3	1.5
SH	5	3	1.5
SWIMMING	5	3	1.5
CRAYON	7	2	1.4
HIM	7	2	1.4
LEG	7	2	1.4
PURPLE	7	2	1.4
SAND	7	2	1.4

SEQUENCE	FREQ	SPEAKERS	INDEX
SAW	7	2	1.4
THANKYOU	7	2	1.4
TODAY	7	2	1.4
TUESDAY	7	2	1.4
BUCKET	6	2	1.2
HER	6	2	1.2
LIM	6	2	1.2
OFF	6	2	1.2
PICK	6	2	1.2
PINK	6	2	1.2
TEACHER	6	2	1.2
TEETH	6	2	1.2
TOILET	6	2	1.2
WET	6	2	1.2
WOMAN	6	2	1.2
BISCUIT	5	2	1.0
BYE	5	2	1.0
FEATHER	5	2	1.0
HO	5	2	1.0
JUMPED	5	2	1.0
KAY	5	2	1.0
LUN	5	2	1.0
MINE	5	2	1.0
MR	5	2	1.0
NINETEEN	5	2	1.0
NUMBER	5	2	1.0
PEN	5	2	1.0
QNAME	5	2	1.0
READING	5	2	1.0
SAME	5	2	1.0
SCHOOL	5	2	1.0
SEVENTEEN	5	2	1.0
TIGER	5	2	1.0
WENT	5	2	1.0

TWO WORD SEQUENCES OF THE VIETNAMESE SAMPLE OF UPPER FIVE YEAR OLD
CHILDREN IN ORDER OF INDEX VALUE

SEQUENCE	FREQ	SPEAKERS	INDEX
ONE TWO	247	8	859.2
TWO THREE	199	7	511.8
THREE FOUR	152	9	404.8
THE LITTLE	50	7	219.1
I CAN	67	8	194.8
FOUR FIVE	120	7	168.2
I AM	56	8	162.8
? WHAT	87	7	155.1
THE DOOR	27	7	118.3
FIVE SIX	84	7	97.5
I M	32	8	93.0
SEE THE	90	7	85.6
THE WATER	22	6	82.6
NO NO	48	5	72.0
I DONT	28	7	71.2
THE BOY	16	7	70.1
AND THE	31	7	62.7
THE BALL	23	4	57.6
THE DOG	14	6	52.6
? WHERE	21	8	42.8
AND A	28	5	40.4
SIX SEVEN	50	6	37.2
WHAT S	70	7	34.9
I SEE	22	4	32.0
A LITTLE	24	5	31.1
I HAVE	17	5	30.9
IN THE	49	8	29.1
ON THE	63	6	28.6
S THIS	35	6	28.1
THE TREE	11	4	27.5
THE GNOUN	14	3	26.3
I WANT	12	6	26.2
ONE MORE	10	6	26.1
COME AND	34	5	25.4
THE HORSE	8	5	25.0
YES I	10	6	24.6
? CAN	16	6	24.4
THE CAT	13	3	24.4
THE GRASS	13	3	24.4
IT S	40	7	23.4
AND SEE	20	4	23.1
TH MAN	7	5	21.9
IS A	15	6	20.3
THE BOX	8	4	20.0
THE CAR	8	4	20.0
AND DOWN	22	3	19.1
THE GIRL	6	5	18.8
MAN	18	4	18.7
AND ONE	12	5	17.3
ONE FOR	13	3	17.0

SEQUENCE	FREQ	SPEAKERS	INDEX
I KNOW	15	3	16.4
LITTLE DOG	26	5	16.4
THE KITTEN	8	3	15.0
COME ON	20	5	14.9
CAN SEE	19	6	14.7
THIS IS	22	6	14.7
THE THE	11	2	13.8
* YES	16	5	13.5
THREE ONE	15	3	13.3
? HOW	13	4	13.2
? IS	13	4	13.2
THE ELEPHANT	7	3	13.1
* NO	19	4	12.9
THE AEROPLANE	5	4	12.5
THE CHAIR	5	4	12.5
THE HOUSE	10	2	12.5
SEVEN EIGHT	35	7	12.4
YES SIR	10	3	12.3
I LIKE	8	4	11.6
OH NO	26	4	11.3
THE BIN	6	3	11.3
THE FLOOR	9	2	11.3
IS HERE	12	4	10.8
S A	20	4	10.7
COME HERE	14	5	10.5
IS IT	15	3	10.2
IS THE	9	5	10.2
YES YES	8	3	9.9
HERE HERE	11	4	9.7
THE FLOWER	5	3	9.4
A COW	7	5	9.1
IS SITTING	10	4	9.0
IS STANDING	10	4	9.0
THE BOAT	7	2	8.8
THE FAT	7	2	8.8
ONE THREE	5	4	8.7
YOU CAN	12	5	8.7
A BOY	11	3	8.6
A HOUSE	11	3	8.6
YES A	7	3	8.6
A RABBIT	8	4	8.3
LOOK UP	23	3	8.1
I WILL	11	2	8.0
THIS ONE	14	5	7.8
--SP ARE	8	5	7.6
I GO	7	3	7.6
I VE	7	3	7.6
A DOG	7	4	7.3
IS IN	8	4	7.2
THREE THREE	6	4	7.1

SEQUENCE	FREQ	SPEAKERS	INDEX
THREE THREE	6	4	7.1
A CAT	9	3	7.0
A DUCK	9	3	7.0
AND PLAY	8	3	6.9
CAN YOU	13	4	6.7
TWO ONE	9	2	6.6
TWO TWO	6	3	6.6
IS FLYING	7	4	6.3
THE BIRD	5	2	6.3
THE BUCKET	5	2	6.3
THE DUCK	5	2	6.3
THE MOUSE	5	2	6.3
THE TIME	5	2	6.3
A BIRD	8	3	6.2
HE S	22	6	6.2
SITTING ON	25	7	6.2
* COME	12	3	6.1
? I	6	4	6.1
? OH	6	4	6.1
AND YELLOW	5	4	5.8
IS EATING	6	4	5.4
AND I	6	3	5.2
AND JUMP	9	2	5.2
ONE ONE	6	2	5.2
* ONE	15	2	5.1
CAN GO	10	4	5.1
CAN RUN	10	4	5.1
CAN PLAY	13	3	5.0
FIVE ONE	15	2	5.0
FOUR ONE	12	2	4.8
S THE	12	3	4.8
A BIG	6	3	4.7
A PIG	9	2	4.7
-SP AND	8	3	4.6
? MISS	9	2	4.6
IS RUNNING	5	4	4.5
DONT KNOW	26	6	4.4
THREE BAG	5	3	4.4
ONE RABBIT	5	2	4.3
COME COME	7	4	4.2
EIGHT NINE	20	6	4.2
CAN T	8	4	4.1
IN H RE	11	5	4.1
LITTLE BOY	15	2	3.8
LOOK AT	8	4	3.8
TWO KITTEN	5	2	3.7
A GIRL	7	2	3.6
YOU SEE	6	4	3.5
IS VERY	5	3	3.4
ARE YOU	18	5	3.3

SEQUENCE	FREQ	SPEAKERS	INDEX
S YOUR	5	5	3.3
SEE YOU	6	4	3.3
BOY -SP	10	6	3.2
UP AND	19	3	3.2
? PLEASE	6	2	3.1
A WINDOW	6	2	3.1
HERE I	7	2	3.1
HERE S	7	2	3.1
* SEE	6	3	3.0
NO ONE	5	2	3.0
AND THIS	5	2	2.9
CAN COME	7	3	2.7
IS THIS	6	2	2.7
THIS WAY	6	4	2.7
A HORSE	5	2	2.6
A TREE	5	2	2.6
YOU GO	6	3	2.6
* LOOK	7	2	2.4
IS BIG	5	2	2.3
LITTLE CAT	6	3	2.3
YOU ANY	5	3	2.2
CAN JUMP	8	2	2.1
* AND	6	2	2.0
* I	6	2	2.0
IT IS	8	3	2.0
LITTLE GIRL	8	2	2.0
LITTLE KITTEN	8	2	2.0
THIS HERE	6	3	2.0
HE IS	10	4	1.9
SEE MY	7	2	1.9
SIX SIX	5	3	1.9
COME UP	6	2	1.8
* WATER	5	2	1.7
YOU TODAY	6	2	1.7
DOG CAN	9	3	1.6
COME SEE	5	2	1.5
DOG RUN	8	3	1.5
AT THE	12	6	1.4
LOOK SEE	6	2	1.4
NINE TEN	13	5	1.4
WHAT IS	5	4	1.4
YOU COME	5	2	1.4
OH LOOK	6	2	1.3
WHAT COLOUR	6	3	1.3
LOOK LOOK	5	2	1.2
BOY AND	5	4	1.1
EATING THE	15	4	1.1
THANK YOU	15	7	1.1
ELEPHANT AND	7	5	1.0
ME ME	8	2	1.0

SEQUENCE	FREQ	SPEAKERS	INDEX
ME ME	8	2	1.0
WHERE S	8	5	1.0
FOR THE	12	3	0.9
MAN IS	8	4	0.9
COLOUR -SP	7	3	0.8
MY MOTHER	5	3	0.8
UP THE	5	3	0.8
MAN AND	6	4	0.7
MISS WEIR	17	2	0.7
SITTING IN	5	4	0.7
THAT S	7	4	0.7
UP UP	6	2	0.7
GOOD MORNING	15	2	0.6
KITTEN -SP	9	3	0.6
RUN AND	7	1	0.6
THAT ONE	12	2	0.6
BOY IS	5	2	0.5
FOR ME	11	2	0.5
GO HOME	5	3	0.5
HAVE YOU	6	4	0.5
HOUSE -SP	5	3	0.5
KITTEN KITTEN	7	3	0.5
LOOKING AT	10	4	0.5
MY HOUSE	5	2	0.5
WHERE IS	5	4	0.5
GO UP	6	2	0.4
OKAY I	6	2	0.4
TEN ELEVEN	7	3	0.4
VERY FAST	5	3	0.4
YELLOW YELLOW	6	2	0.4
BALL -SP	6	2	0.3
CAR -SP	8	2	0.3
FLOWER -SP	5	4	0.3
GO AWAY	5	2	0.3
MANY COLOUR	6	3	0.3
QUICK QUICK	12	2	0.3
VERY GOOD	6	2	0.3
WANT TO	7	4	0.3
BIRD -SP	12	2	0.2
BLACK AND	5	2	0.2
HOW ARE	10	3	0.2
OPEN THE	10	3	0.2
RABBIT -SP	5	2	0.2
RABBIT AND	5	2	0.2
SHUT THE	7	5	0.2
TREE IS	5	2	0.2
AEROPLANE -SP	5	2	0.1
BAA BAA	8	3	0.1
BAG -SP	5	3	0.1
BE E	5	2	0.1

SEQUENCE	FREQ	SPEAKERS	INDEX
BOAT AND	5	2	0.1
CARD -SP	11	2	0.1
ELEVEN TWELVE	6	3	0.1
GIVE ME	8	3	0.1
HAND -SP	7	3	0.1
LIKE THE	5	2	0.1
SHUTTING THE	6	4	0.1
SOME MORE	5	2	0.1
FEATHER -SP	5	2	0.0
KICK IT	7	2	0.0
NINETEEN TWENTY	5	2	0.0
SAW A	6	2	0.0
SHELL -SP	7	3	0.0
SIR THREE	5	3	0.0
SIR YES	5	3	0.0
VE GOT	6	2	0.0

THREE WORD SEQUENCES OF THE VIETNAMESE SAMPLE OF UPPER FIVE YEAR OLD
CHILDREN IN ORDER OF INDEX VALUE

SEQUENCE	FREQ	SPEAKERS	INDEX
ONE TWO THREE	196	7	1182.7
TWO THREE FOUR	140	6	431.3
THREE FOUR FIVE	115	7	327.0
FOUR FIVE SIX	81	7	95.7
? WHAT S	59	6	55.1
FIVE SIX SEVEN	48	6	28.2
I CAN SEE	18	6	21.1
THE LITTLE DOG	19	4	16.7
SEE THE LITTLE	28	5	12.0
I DONT KNOW	24	6	10.3
SIX SEVEN EIGHT	32	6	7.2
WHAT S THIS	34	6	7.1
I CAN PLAY	9	1	5.3
NO NO NO	22	3	4.8
ONE TWO FIVE	2	2	3.4
ONE TWO FOUR	2	2	3.4
? WHAT COLOUR	6	3	2.8
THE LITTLE CAT	5	2	2.2
TWO THREE THREE	2	2	2.1
I AM PLAYING	4	3	2.0
I AM SITTING	4	3	2.0
? WHAT IS	4	3	1.9
THE LITTLE BOY	4	2	1.8
I CAN GO	4	2	1.6
I M VERY	4	4	1.5
SEVEN EIGHT NINE	20	6	1.5
SEE THE CAT	8	2	1.4
SEE THE DOG	5	3	1.3
? WHERE S	7	4	1.2
COME AND SEE	15	3	1.1
THE DOOR IS	3	3	1.1
I AM EATING	3	2	1.0
SEE THE BOAT	6	2	1.0
SEE THE KITTEN	6	2	1.0
? CAN YOU	12	3	0.9
I SEE THE	14	2	0.9
THE LITTLE APPLE	2	2	0.9
AND SEE THE	12	3	0.8
I CAN RUN	2	2	0.8
I CAN SHUT	2	2	0.8
I CAN TAKE	2	2	0.8
IT S A	12	3	0.8
I AM HITTING	2	2	0.7
I AM STANDING	2	2	0.7
SEE THE ELEPHANT	4	2	0.7
THE DOOR PLEASE	3	2	0.7
AND THE BOY	3	3	0.6
AND THE DOG	3	3	0.6
I M HERE	3	2	0.6
IN THE BOX	7	3	0.6

SEQUENCE	FREQ	SPEAKERS	INDEX
IN THE WATER	7	3	0.6
ON THE GRASS	11	2	0.6
ONE FOR THE	11	3	0.6
? WHERE ARE	4	3	0.5
? WHERE IS	4	3	0.5
CAN SEE THE	7	5	0.5
I WANT TO	6	3	0.5
IN THE DIN	6	3	0.5
ON THE FLOOR	8	2	0.5
SEE THE DUCK	3	2	0.5
SITTING ON THE	22	4	0.5
? HOW ARE	10	3	0.4
AND THE LION	3	2	0.4
I DONT THINK	3	2	0.4
I M EATING	2	2	0.4
I M LOOKING	2	2	0.4
THE BOY IS	3	2	0.4
THE MAN IS	5	4	0.4
THIS IS A	10	3	0.4
THREE ONE TWO	14	2	0.4
? IS IT	10	2	0.3
AND THE GIRL	2	2	0.3
AND THE LITTLE	2	2	0.3
AND THE GNOUN	2	2	0.3
COME AND PLAY	6	2	0.3
EIGHT NINE TEN	12	5	0.3
ON THE CHAIR	3	3	0.3
SEE THE AEROPLANE	2	2	0.3
SEE THE BOY	2	2	0.3
SEE THE BUTTERFLY	2	2	0.3
THE BOY --SP	2	2	0.3
THE DOG CAN	3	2	0.3
WHAT S THAT	3	3	0.3
AND A WOMAN	2	2	0.2
IT S AN	4	2	0.2
IT S THE	5	2	0.2
LITTLE DOG CAN	6	2	0.2
LITTLE DOG RUN	5	3	0.2
UP AND DOWN	19	3	0.2
YES SIR THREE	5	3	0.2
YES SIR YES	5	3	0.2
* ONE TWO	10	2	0.1
? CAN I	2	2	0.1
? HOW MANY	2	2	0.1
? I CAN	3	3	0.1
? OH WHAT	3	3	0.1
A MAN AND	3	2	0.1
AND ONE FOR	4	2	0.1
AND ONE RABBIT	2	2	0.1
CAN GO UP	6	2	0.1

SEQUENCE	FREQ	SPEAKERS	INDEX
CAN GO UP	6	2	0.1
CAN I HAVE	7	3	0.1
FIVE ONE TWO	13	2	0.1
FOUR ONE TWO	12	2	0.1
HERE HERE HERE	3	2	0.1
I HAVE A	2	2	0.1
I VE GOT	6	2	0.1
IN THE BUCKET	2	2	0.1
IN THE SAND	2	2	0.1
IN THE SKY	2	2	0.1
IS IN THE	5	4	0.1
IS SITTING ON	3	3	0.1
IT S HERE	2	2	0.1
IT S NOT	2	2	0.1
LITTLE DOG AND	2	2	0.1
ONE THREE FOUR	3	3	0.1
THE CAT AND	3	2	0.1
THE ELEPHANT AND	3	2	0.1
THE GIRL IS	2	2	0.1
THE HORSE IS	2	2	0.1
THE THE THE	5	2	0.1
THREE BAG -SP	5	3	0.1
TWO ONE TWO	8	2	0.1
WHAT S YOUR	2	2	0.1
* I CAN	5	2	0.0
* SEE THE	3	2	0.0
-SP AND THE	2	2	0.0
-SP ARE HERE	2	2	0.0
? MISS WEIR	5	2	0.0
AND THIS IS	3	2	0.0
ARE YOU GOING	2	2	0.0
ARE YOU TODAY	6	2	0.0
BAA BAA BLACK	3	2	0.0
BAG -SP FULL	3	2	0.0
BOY -SP AND	2	2	0.0
BOY AND A	2	2	0.0
BOY AND THE	2	2	0.0
CAN COME UP	5	2	0.0
CAN GO HOME	3	2	0.0
CAN JUMP AND	2	2	0.0
CAN RUN AND	4	2	0.0
CAN YOU SEE	3	2	0.0
COME SEF THE	5	2	0.0
DOG CAN COME	4	2	0.0
DOG CAN RUN	4	3	0.0
EATING THE BREAD	2	2	0.0
EATING THE FLOWER	3	2	0.0
EIGHT NINE NINE	3	2	0.0
ELEPHANT AND THE	3	2	0.0
FIVE ONE THREE	2	2	0.0

SEQUENCE	FREQ	SPEAKERS	INDEX
HAVE YOU ANY	5	3	0.0
HE IS IN	2	2	0.0
HE S EATING	3	2	0.0
HERE I GO	4	2	0.0
HERE S A	4	2	0.0
HOW ARE YOU	10	3	0.0
I GO TO	2	2	0.0
IS EATING THE	3	2	0.0
IS RUNNING VERY	2	2	0.0
IS SITTING IN	2	2	0.0
IS STANDING IN	2	2	0.0
IT IS A	2	2	0.0
KITTEN KITTEN KITTEN	4	2	0.0
LITTLE BOY AND	2	2	0.0
LOOK SEE THE	5	2	0.0
LOOK UP UP	2	2	0.0
LOOKING AT THE	6	3	0.0
MAN AND A	4	2	0.0
MANY COLOUR -SP	6	3	0.0
NINE TEN ELEVEN	6	3	0.0
ONE ONE TWO	2	2	0.0
OPEN THE DOOR	10	3	0.0
S YOUR NAME	2	2	0.0
SEE YOU TOMORROW	2	2	0.0
SHUT THE DOOR	5	4	0.0
SHUT THE WINDOW	2	2	0.0
SHUTTING THE DOOR	6	4	0.0
SIR THREE BAG	5	3	0.0
SIR YES SIR	5	3	0.0
SITTING IN THE	4	3	0.0
TEN ELEVEN TWELVE	6	3	0.0
THE DUCK RUN	2	2	0.0
TWO KITTEN -SP	5	2	0.0
TWO TWO FOUR	2	2	0.0
UP THE TREE	3	2	0.0
UP UP UP	3	2	0.0
WANT TO HIT	3	2	0.0
WHAT IS IT	2	2	0.0
WHERE IS YOUR	2	2	0.0
WHERE S MY	2	2	0.0
WHERE S THE	2	2	0.0
WHERE S YOUR	2	2	0.0
YELLOW YELLOW YELLOW	2	2	0.0
YOU ANY WOOL	2	2	0.0

WATCC

EDITH COWAN UNIVERSITY



AA5061612B