A LINGUISTIC DESCRIPTION AND COMPUTER PROGRAM FOR CHILDREN'S SPEECH

Geoffrey J. Turner and Bernard A. Mohan

ROUTLEDGE LIBRARY EDITIONS:
LINGUISTICS

ROUTLEDGE LIBRARY EDITIONS:
LINGUISTICS

Volume 34

A LINGUISTIC DESCRIPTION AND COMPUTER PROGRAM FOR CHILDREN'S SPEECH

This page intentionally left blank

A LINGUISTIC DESCRIPTION AND COMPUTER PROGRAM FOR CHILDREN'S SPEECH

GEOFFREY J. TURNER AND
BERNARD A. MOHAN

Routledge
Taylor & Francis Group
LONDON AND NEW YORK

First published in 1970

This edition first published in 2014
by Routledge
2 Park Square, Milton Park, Abingdon, Oxon, OX14 4RN

Simultaneously published in the USA and Canada
by Routledge
711 Third Avenue, New York, NY 10017

Routledge is an imprint of the Taylor & Francis Group, an informa business

© 1970 Geoffrey J. Turner & Bernard A. Mohan

All rights reserved. No part of this book may be reprinted or reproduced or utilised in any form or by any electronic, mechanical, or other means, now known or hereafter invented, including photocopying and recording, or in any information storage or retrieval system, without permission in writing from the publishers.

Trademark notice: Product or corporate names may be trademarks or registered trademarks, and are used only for identification and explanation without intent to infringe.

British Library Cataloguing in Publication Data
A catalogue record for this book is available from the British Library

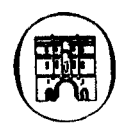
ISBN: 978-0-415-64438-9 (Set)
eISBN: 978-0-203-07902-7 (Set)
ISBN: 978-0-415-72402-9 (Volume 34)
eISBN: 978-1-315-85723-7 (Volume 34)

Publisher's Note
The publisher has gone to great lengths to ensure the quality of this reprint but points out that some imperfections in the original copies may be apparent.

Disclaimer
The publisher has made every effort to trace copyright holders and would welcome correspondence from those they have been unable to trace.

A Linguistic Description and Computer Program for Children's Speech

GEOFFREY J. TURNER
and
BERNARD A. MOHAN



London
ROUTLEDGE & KEGAN PAUL

First published 1970
by Routledge & Kegan Paul Ltd
Broadway House, 68–74 Carter Lane
London E.C.4
Printed in Great Britain
by Clarke, Doble & Brendon L'd
© Geoffrey J. Turner & Bernard A. Mohan 1970
No part of this book may be reproduced
in any form without permission from the
publisher, except for the quotation of
brief passages in criticism
I.S.B.N. 0 7100 6611 2

CONTENTS

Introduction by Professor Basil Bernstein vii

PART ONE

page

A LINGUISTIC APPROACH TO CHILDREN'S SPEECH 1

Geoffrey Turner

1 Aims and Approach 3

2 Outline of the Linguistic Theory 7

2.1 *Introductory* 7
2.2 *The categories of the grammatical theory* 8
2a *Unit* 8
2b *Structure* 10
2c *Class* 11
2d *System* 12
2.3 *Multivariate and univariate structures* 14
2.4 *Co-ordination and branching* 15
2.5 *Summary* 18

3 The Sentence 19

3.1 *Introductory* 19
3.2 *Definition of the sentence* 19
3.3 *Contextual function of the sentence* 22
3a *Response sentences* 23

3b *Textshifted sentences* page 31
3.4 *Sentence complexity* 34
3.5 *Coding information* 37

4 The Clause 40
4.1 *Introductory* 40
4.2 *Non-rankshifted clauses* 40
2a *Free clauses* 41
2a.i *Mood* 41
2a.ii *Contextual functions of free clauses* 42
2a.iii *Elected subclassification of free clauses* 43
2a.iv *Social class influences on the use of free clauses* 45
2b *Bound clauses* 50
2b.i *Subclassification of bound clauses* 50
2b.ii *Social class influences on the use of bound clauses* 54
2c *Tag clauses* 56
4.3 *Rankshifted clauses* 60
4.4 *Clause elements* 62
4.5 *Clause complexity* 63
4.6 *Systemic choices in the clause* 64
4.7 *Coding information* 65

5 The Nominal Group 68
5.1 *Introductory* 68
5.2 *The primary structure of the nominal group* 69
5.3 *The modifier* 70
5.4 *The head* 75
5.5 *The modifier and the head related* 77
5.6 *The qualifier* 82
5.7 *The modifier and the qualifier related* 83
5.8 *Nominal group complexity* 88
5.9 *Coding information* 89

6 The Verbal Group 93
6.1 *Introductory* 93
6.2 *Tense* 93
6.3 *Voice* 107
6.4 *Modality* 111
4a *Modals of 'possibility'* 113
4b *Modals of 'obligation'* 113
6.5 *Polarity* 115
6.6 *Aspect* 118
6.7 *Lexical element* 119
6.8 *Coding information* 122

7 The Adjunctival Group *page* 126
7.1 *Introductory* 126
7.2 *Prepositional group* 126
7.3 *Adverbial group* 129
7.4 *Vortex* 129
7.5 *The temperer and the finisher* 135
7.6 *Adverbial and prepositional group complexity* 136
7.7 *Linker* 136
7.8 *Binder* 137
7.9 *Coding information* 139

8 Mazes 143
8.1 *Introductory* 143
8.2 *Maze types* 145
8.3 *Coding information* 152

9 Concluding Remarks 153

PART TWO

A COMPUTER PROGRAM FOR ASSISTING THE LINGUISTIC ANALYSIS OF CHILDREN'S SPEECH 155

Bernard A. Mohan

Bibliography 175

ACKNOWLEDGEMENTS

We should first like to acknowledge the sympathy, encouragement and specific assistance of Professor M. A. K. Halliday of University College, London. The present authors alone, however, are responsible for the linguistic description presented in this monograph, which is based on the early work of Professor Halliday.

Secondly, we must thank our colleagues, Dr. W. P. Robinson (Deputy Head of the Sociological Research Unit 1965–6), Misses J. A. C. Nicholls, S. J. Rackstraw, A. C. Scott, and Mr. J. N. H. Pellowe, for their useful advice and criticism at the time of the compilation of the coding frame.

We should also like to express our gratitude to our present colleagues for their comments on earlier drafts of parts of this monograph: Dr. R. Hasan, Miss A. Hakulinen, Mrs. S. Taylor, and Messrs. P. R. Hawkins and R. E. Pickvance. Special thanks are due to Mr. B. N. Lewis (Deputy Head of the Sociological Research Unit 1966–8) for many enjoyable discussions from which we have greatly benefited.

We wish to thank the Director and staff of the University of London Institute of Computer Science for providing facilities which have enabled us to develop this work, and particularly Mr. D. F. Hendry for his encouragement and advice in the application of BCL to linguistic analysis.

Finally, we must express our deep indebtedness to Professor B. B. Bernstein for his inspiration and guidance throughout the period of the work.

INTRODUCTION

In this volume we are presenting a detailed account of the application of Professor M. A. K. Halliday's Scale and Category Grammar to the analysis of the speech of five-year-old children. This application was carried out by G. J. Turner in association with colleagues in the linguistic section of the Sociological Research Unit. This volume also contains the computer program designed by B. A. Mohan, who also worked with Mr. Turner on the coding frame for the linguistic analysis. Mr. Mohan would wish me to acknowledge the Unit's debt to Mr. Hendry of the Institute of Computer Science who acted as consultant throughout the period of the writing of the program. Mr. Brian Lewis, Deputy Head of the S.R.U., worked closely with Mr. Mohan in the final stage of writing up the program so that it could be made available in its present form. In terms of the various Unit activities we would all agree that the development of the computer program by Mr. Mohan and the grammatical analysis by Mr. Turner (in association with his fellow researchers) represent major achievements in the history of our research.

I would like in this introduction to outline briefly some of the problems which lay behind the development of the grammatical analysis. When we started in 1963 we were faced with the initial difficulty that there existed no precedents for the work we were doing. Very little was known about the speech of five-year-old children and certainly no major *grammatical* analysis had been reported in the literature of the speech of English children. We also had little idea of the influence of social class, sex of the child and ability upon such children's speech. We were also confronted with the problem of the selective effect of the various tasks which comprised the first language sample (see Brandis W. and Henderson D., Social Class, Language and Communication; Routledge and Kegan Paul, 1969) upon the nature of the choices the five-year-old children would take up. Finally we were also concerned to produce a grammatical description which would enable us to examine the speech of the children when they were older.

The only guides we had were the limited studies of the speech and written language of ten-year-old and fifteen-year-old children. A Nuffield Grant (1963–5) provided the Sociological Research Unit with the opportunity to carry out a grammatical analysis, using Halliday's Scale and Category Grammar, of sixty children's written language. Mr. Turner joined this project in 1964 and his work on the grammatical description became the basis of his report in this volume. We were also concerned to ensure that the grammatical analysis would enable us to relate the children's choices to various social antecedents. The developing theory with which I had been struggling had never been applied to the speech of very young children. Indeed even in the case of older children it had yielded only very general expectancies of the grammatical choices which would be taken up in different social contexts. I am distinctly conscious of the bewilderment I felt at the onset of this research when I was called upon to make decisions about the inclusion or exclusion of sub-systems within the grammar and even more to judge the degree of delicacy of the analysis at any level of the rank scale. I write this so that the reader is aware not only of the nature of the unmapped territory we were about to explore but also so that the reader will be aware of the significance of Mr. Turner's and Mr. Mohan's contribution.

We decided to set up a fairly broad and in part delicate grammatical description and apply this to all tasks within the first speech schedule. We used as our sample, the children in the factorial design reported elsewhere. This sample contained 110 children matched for various attributes like social class, sex, ability and according to an index of the mother's verbal communication to her child. This analysis together with the construction of the grammatical description took Mr. Turner and four assistants one year to carry out. We considered that such an intensive analysis might open up the directions to follow in more specialised analyses.

Mr. Mohan was faced with the problem of writing a program which included about 250 speech elements and also numerous types of possible maze phenomena in speech, (false starts, grammatical and lexical substitutions etc.). The precision required by the program brought about a continuous dialogue between Mr. Turner and Mr. Mohan.

In this account the reader will find reference to the 'middle-class' and the 'working-class'. We would not like to give the impression that the linguistic behaviour of either parents or children are as homogeneous or as different as these terms would appear to indicate. Indeed a paper written (unpublished) in 1962 and further papers since that date have moved further and further away from the crude ascription of particular uses of language to broad social class groups. This is *not* to say that broad social class correlates of non-dialect uses of language do not exist (our research is witness to their existence) only to say that as the theory developed in generality, the social antecedents of these uses of language became rather more precise, and modes of restriction and of elaboration were introduced.

It remains to be seen whether the speech of mothers and children we have collected will allow these later theoretical developments to be tested for their usefulness.

We hope that the approach to the grammatical analysis of children's speech, although developed for a specific research inquiry and following a specific linguistic theory, will be of interest to any workers concerned with the speech of children.

Basil Bernstein

This page intentionally left blank

PART ONE

A LINGUISTIC APPROACH TO CHILDREN'S SPEECH

This page intentionally left blank

Chapter 1

AIMS AND APPROACH

In a paper called 'Syntax and the Consumer', Halliday (1964) puts forward the view that 'different coexisting models in linguistics may best be regarded as appropriate to different aims, rather than as competing contenders for the same goal'. The aim of the grammar developed by Halliday and his colleagues is 'to show the patterns inherent in the linguistic performance of the native speaker'. The description involves 'a characterisation of the special features, including statistical properties, of varieties of the language used for different purposes ("registers"), and the comparison of individual texts, spoken and written, including literary texts. This in turn is seen as a linguistic contribution towards certain further aims, such as literary scholarship, native and foreign language teaching, educational research, sociological and anthropological studies and medical applications.' As Halliday points out: 'The interest is focused not on what the native speaker knows of his language but rather on what he does with it; one might perhaps say that the orientation is primarily textual and, in the widest sense, sociological.'

Our own particular aim was to make a statistical comparison of linguistic features in the speech of a sample of five-year-old working-class and middle-class children. Our main interest was in isolating social class differences in usage, those social class differences which Bernstein's theory of elaborated and restricted codes (e.g. Bernstein, 1965) implies are of educational consequence. It was necessary to take fairly large samples of children (450 children in all) in order to be able to say anything useful about the influence of social class. Furthermore, it was necessary to obtain speech from each subject in a variety of situations, in order to obtain some kind of check on the influence of the situation and to see whether particular situations favoured one or other of the social classes. It can be seen that there is a basic compatibility between our aims and those for which Halliday regards his type of grammar as appropriate.

Marshall and Wales (1966), in a critique of Halliday's paper, argue that the study of linguistic performance should be preceded by a study of linguistic competence, competence being the tacit knowledge of the language user. They write: 'It has always been accepted that, once the outlines of a competence model are fairly clear, we can then explore the

manner in which this competence is realised in performance (Chomsky and Miller, 1963).' If we carry on performance studies without a competence theory, 'we are left without a norm against which to evaluate the results of such performance studies'. For a study of certain aspects of language performance a prior study of competence would indeed be necessary or at least highly desirable, for example, for a full examination of 'maze' behaviour, that is, false starts, substitutions etc., of grammatical deviations, and so forth. However, for most of our work it is not necessary to attempt to characterise the underlying competence of the speaker. Bernstein (1965) writes: 'The code which the linguist invents in order to explain speech events is capable of generating n number of speech codes.' The language code is a set of options. It is 'a set of rules to which all speech codes must comply, but which speech codes are generated is a function of the system of social relations'. The form of the social relation regulates the options which the speakers take up. Speech codes are distinguished in terms of the relative frequencies with which particular options available in the language code are taken up. As Bernstein (1964b) puts it: 'Speech . . . is constrained by the circumstances of the moment, by the dictate of the local social relation and so symbolises not what can be done, but what *is* done with different degrees of frequency.' The social class differences in speech and writing that have been found by Bernstein (1962b) and Lawton (1963, 1964) have been differences in relative frequency. There has been no suggestion in the above studies that there is any fundamental difference in the tacit knowledge of the middle-class and working-class language user. What is of primary importance to Bernstein is not the difference between 'competence' and 'performance' but the difference between performances that have been influenced by different social relations.

We might point out, theoretical considerations apart, that it would have been an enormously difficult task, in terms of time and manpower, to have attempted a systematic study of underlying competence in samples of subjects large enough to be representative of the different social classes. The difficulties of obtaining evidence about underlying competence may be gathered from this quotation from Chomsky (1964): 'If anything far-reaching and real is to be discovered about the actual grammar of the child, then rather devious kinds of observation of his performance, his abilities and his comprehension in many different kinds of circumstances will have to be obtained, so that a variety of evidence may be brought to bear on the attempt to determine what is in fact his underlying linguistic competence at each stage of development.' It was not feasible for us to approach large samples of subjects in this way.

Almost all of the linguistic description to be presented in this Monograph consists of a modified and much simplified version of the description made and at present being developed by Professor M. A. K. Halliday, University College, London. It should be pointed out that the

categories for the present description were worked out in 1965. For practical purposes it was necessary to 'freeze' the description at that time; as we were coding a large body of data there was not time available for incorporating revised or new categories into the description and for re-working the data once the main analysis had begun. Considerable changes have been made in Halliday's description since that time. Some suggestion of these changes may be gauged from the name that is now used to refer to the description, 'systemic', rather than 'scale-and-category' (relating to the scales 'rank', 'exponence', 'delicacy' and later 'depth' and the categories 'unit', 'structure', 'class' and 'system') or 'system-structure'. Briefly, much greater emphasis has been placed on paradigmatic relations (the realm of the category 'system') and much less emphasis on syntagmatic relations ('structure'), the latter now being fully derived from the former. Our work came too early to draw fully on the new forms of description. So whilst we describe systems, not all the structural representations we give are related to underlying systemic choices. For this reason we shall refer to the grammatical description we used as 'scale-and-category' rather than 'system-structure' or 'systemic' description. Although our main purpose is to describe the grammatical categories that we used, we shall mention, wherever it seems appropriate, the more recent categories, and comment on their usefulness.

The modifications and simplifications mentioned above were made for four main considerations, namely, our sociological interests, computational convenience, the age of the subjects, and coder reliability. From the point of view of the sociological theory certain parts of the grammar were of much more interest than other parts, so, for example, the choices of number and person in the personal pronouns are much more important than, say, the choice of number in the noun. One great advantage of 'scale-and-category' grammar is that it does allow the linguist to vary the amount of detail he goes into in his description by making use of the scale of delicacy. The second consideration was that we intended to make the frequency counts with the aid of a computer. In order to keep the program of a reasonable length, a number of economies were made in the description; for example, tables were restricted to ten entries. It should be added, however, that these restrictions were quite consonant with our general aim which was to make a broad general survey of the data in order to see which were the special areas of promise, worthy of a subsequent more delicate analysis. Some modifications were made because we were dealing with five-year-old children. According to Inhelder and Piaget (1958) children of this age are still generally at the stage of 'preoperational intuitive thinking'. They do not reach the stage of 'formal propositional thinking' until they are eleven years and upwards. We used the divisions of Piaget as a guideline. So, for example, in our sub-classification of binders (subordinating conjunctions), we did not consider it necessary to have separate sub-classes for *granted (that), insofar as, provided that, assuming*

(that), etc. Instead, a general category, 'other binders', was set up, just in case binders such as these occurred. In general, then, we were less detailed in our description of those areas of grammar that we did not consider would be exploited by five-year-old children. It should be stressed, though, that we used the divisions from Piaget's theory of cognitive development only as a guideline. We do not mean to imply a complete acceptance of Piaget's position, which would seem to be that language follows cognition rather than precedes it; that language depends on the development of thought processes. We would stress the interdependence of language and cognition, and argue that certain varieties of language may facilitate or inhibit the child's cognitive development. As Bernstein (1961) writes: 'The Piagetian development sequence from concrete to formal operations may not be inevitable for a child restricted to a public language. The child may well be limited to limited concrete options.' Finally, we had to consider the question of coder reliability; there were in fact five researchers coding the data. We did not use any category which did not achieve 80% coder agreement in the tests which were run before the main analysis began. Final coder agreement over categories varied between 80% and 100%.

The chapter which follows gives a brief outline of the linguistic theory on which the present description was based.

Chapter 2

OUTLINE OF THE LINGUISTIC THEORY

2.1 Introductory

The linguistic theory underlying this study is that constructed by Professor M. A. K. Halliday (see especially Halliday, 1961). The theory requires that linguistic events should be comprehended at several different levels as different kinds of patterning are involved. There are three primary levels: substance, form and context. The substance is the raw material of language: 'phonic' (audible noises) or 'graphic' (visible marks). The form is the internal structuring of the substance; its organisation into meaningful events. The context is the relation of form to non-linguistic features in the situations in which language is used. We reproduce below the table given in Halliday, McIntosh and Strevens (1964) which summarises the complete framework of levels:

<table>
<tr><td>Subject concerned</td><td colspan="2">Phonetics</td><td colspan="2">Linguistics</td><td></td></tr>
<tr><td>Level (general)</td><td>SUBSTANCE (phonic or graphic)</td><td>relation of form and substance</td><td>FORM</td><td>CONTEXT (relation of form and situation)</td><td>situation (non-linguistic phenomena)</td></tr>
<tr><td rowspan="2">Level (specific)</td><td>PHONETICS</td><td>PHONOLOGY</td><td rowspan="2">GRAMMAR & LEXIS (vocabulary)</td><td rowspan="2">SEMANTICS</td><td rowspan="2"></td></tr>
<tr><td>SCRIPT</td><td>'GRAPHOLOGY' (writing system)</td></tr>
</table>

Form, it may be observed, is made up of two 'demi-levels': grammar and lexis. The distinction between these two requires some comment. It is based on recognition of the fact that at different places in the language there are different ranges of possibilities. Grammar accounts for those places where there are closed system choices, that is, those where there is a

small fixed number of contrastive possibilities and a clear line between what is possible and what is not; the choice of 'active' or 'passive' in the verbal group is an example of a closed system. Lexis accounts for those places where there are open set choices, that is, those where there is a very large number of possibilities and no clear line separating what is possible from what is not; in a clause such as 'she is a very . . . girl', the blank may be filled by a choice from a very large range of items, including old, sensitive, athletic, domineering, etc.: it is items such as these that are handled by lexis.

Our description is grammatical in the majority of cases; just occasionally it relates to other levels, particularly those of phonology and context. It will be enough to describe the phonological and contextual categories as they occur in the coding frame, but the description of the grammatical ones will benefit from a prior consideration of the grammatical theory underlying them.

2.2 The categories of the grammatical theory

'Scale-and-category' grammar has four theoretical categories, and from these theoretical categories are derived the descriptive categories which are needed for the description of a particular language. The four fundamental categories are: unit, structure, class and system. Each of these will now be explained and illustrated with examples from English grammar, particularly aspects of grammar that are relevant to our work.

2.2a Unit

The category 'unit' is set up to account for the stretches of language that carry grammatical patterns. The units of grammar form a taxonomic hierarchy. In such a hierarchy there is a fixed relation among the members and each member is assigned a place in order in the hierarchy. The relation is one of constituency, such that, going from the highest (largest) to the lowest (smallest), an occurrence of a unit consists of one, or more than one, complete occurrence of the unit next below it. For a description of the English language we require five types of unit: sentence, clause, group, word and morpheme. The sentence consists of one or more complete clauses, the clause of one or more complete groups, and so on. The units that are required for the description of a language are in fact ranged on a scale, the scale of rank. We shall have more to say about this scale, especially when we discuss the categories 'structure' and 'class'.

Now we shall exemplify the five types of unit that are needed for a description of English. We shall take the following sentence:

(1) That dog is trying to escape.
111 That dog 1 is try+ing 1 to escape 111

The boundary symbols are those given by Halliday *et al.* (1964). Briefly, 'that dog is trying to escape' is a sentence which consists of one clause. The clause consists of three groups 'that dog', 'is trying' and 'to escape'. Each group consists of two words. Each word consists of one morpheme, except 'trying', which comprises two morphemes 'try' and 'ing'.

Our system for denoting unit boundaries is based on the Hallidayan system. The two are reproduced alongside below:

Halliday et al.	*Type of Boundary*	*Present writers*
111	sentence boundary	/3
11	clause boundary	/2
1	group boundary	/1
(space)	word boundary	(no space) or/
+	morpheme boundary or fusion of morphemes	not analysed
[[ ]]	boundary of rankshifted clause	⟨ /2⟩
[ ]	boundary of rankshifted group	⟨ /1⟩
(( ))	boundary of included clause	(/2)

It should be remembered that in both forms of notation a boundary symbol of higher rank dominates one of lower rank, so, for example, /3 dominates /2. That is why in the example above the clause boundary symbol does not appear: in a sentence which consists of only one clause, the clause boundary symbol is dominated by the co-extensive sentence boundary symbol, so:

not	11	111	or	/2		/3
but	111				/3	

In our analysis we carried these boundary symbols two stages further. We recognised /4 'task boundary' and /5 'interview boundary'. So the last sentence in a task received the /4 boundary and the last one of the interview received /5.

Usually, then, a unit operates as part of the unit next above it on the rank scale. However, the theory allows for rankshift. This happens when a unit is shifted down the rank scale, to operate as part of a unit of rank lower than or equal to itself. In the example which follows, 'who are getting married' is a clause acting in group structure:

Those are the people who are getting married.

The clause in question does not add a comment on the whole of the first clause, but rather defines which people are being referred to.