Aston University

Some pages of this thesis may have been removed for copyright restrictions.

If you have discovered material in Aston Research Explorer which is unlawful e.g. breaches copyright, (either yours or that of a third party) or any other law, including but not limited to those relating to patent, trademark, confidentiality, data protection, obscenity, defamation, libel, then please read our Takedown policy and contact the service immediately (openaccess@aston.ac.uk)

A LINGUISTIC STUDY OF INTERVIEW AND CONVERSATION

VOLUME 1

SUBHI SHAMOON POLIS ZORA

Doctor of Philosophy

THE UNIVERSITY OF ASTON IN BIRMINGHAM

JANUARY 1990

This copy of the thesis has been supplied on condition that anyone who consults it is understood to recognise that its copyright rests with its author and that no quotation from the thesis and no information derived from it may be published without the author's prior, written consent.

THE UNIVERSITY OF ASTON IN BIRMINGHAM

A LINGUISTIC STUDY OF INTERVIEW AND CONVERSATION

SUBHI SHAMOON POLIS ZORA Ph.D 1990

THESIS SUMMARY

The present thesis focuses on the overall structure of the language of two types of Speech Exchange Systems (SES) : Interview (INT) and Conversation (CON). The linguistic structure of INT and CON are quantitatively investigated on three different but interrelated levels of analysis : **Lexis, Syntax and Information Structure**. The corpus of data investigated for the project consists of eight sessions of pairs of conversants in carefully planned interviews followed by unplanned, surreptitiously recorded conversational encounters of the same pairs of speakers. The data comprise a total of approximately 15.200 words of INT talk and of about 19.200 words in CON.

Taking account of the debatable assumption that the language of SES might be complex on certain linguistic levels (e.g. syntax) (Halliday 1979) and might be simple on others (e.g. lexis) in comparison to written discourse, the thesis sets out to investigate this complexity using a statistical approach to the computation of the structures recurrent in the language of INT and CON. The findings indicate clearly the presence of linguistic complexity in both types. They also show the language of INT to be slightly more syntactically and lexically complex than that of CON. Lexical density seems to be relatively high in both types of spoken discourse. The language of INT seems to be more complex than that of CON on the level of information structure too. This is manifested in the greater use of Inferable and other linguistically complex entities of discourse. Halliday's suggestion that the language of SES is syntactically complex is confirmed but not the one that the more casual the conversation is the more syntactically complex it becomes.

The results of the analysis point to the general conclusion that the linguistic complexity of types of SES is not only in the high recurrence of syntactic structures, but also in the combination of these features with each other and with other linguistic and extralinguistic features.

The linguistic analysis of the language of SES can be useful in understanding and pinpointing the intricacies of spoken discourse in general and will help discourse analysts and applied linguists in exploiting it both for theoretical and pedagogical purposes.

KEY WORDS :
INTERVIEW CONVERSATION LEXICAL DENSITY
SYNTACTIC COMPLEXITY INFORMATION STRUCTURE

ACKNOWLEDGEMENTS

I am greatly indebted to a number of people without whose help this work would not have been possible.

First of all, I would like to thank my supervisor Catherine M. Johns-Lewis for her continued advice and support. Her constructive comments have been helpful and illuminating during the various stages of this work.

My sincerest thanks are due to Professor Frank Knowles and Mr Tom Bloor for the great help and support they provided me with at the most crucial stage of the work. Their insightful comments and suggestions encouraged me greatly and helped me to get the thesis completed.

I would also like to thank Rev. John Austen for the great help he offered me in collecting the data and for the sixteen students who acted as informants during the interviews and conversations.

To Professor Dennis Ager, Dr. John Fletcher, Dr. Peter Coxhead, Keith Richards, John Stagg, of Aston University, and Jean Ure of the University of Edinburgh, I owe special gratitude for their great help, advice and constructive comments.

I am also indebted to the authorities at Mosul University for granting me the necessary study leave to undertake this work. To my colleagues and friends at the College of Arts I owe special acknowledgement for their confidence in me, their continued support and understanding, and for actually covering my teaching responsibilities during the years I spent doing this thesis.

To my wife JANET and my four daughters DIYAN, JIYAN, EAVAN, and BASMAN who saw very little of me during the past four years, I am particularly indebted. I thank them for the immense support they have always provided me with. To them, and particularly to the one born on the very day the study leave was granted to me, Diyan, I dedicate this work.

LIST OF CONTENTS

Title Page		1
Thesis Summary		2
Acknowledgements		3
List of Contents		4
List of Tables		13
List of Figures and Diagrams		17
Contents of Volume Two		18
List of Abbreviations and Participant Initials		20
Transcription Conventions		21
INTRODUCTION		22
0.1	Introduction to the area	22
0.2	Introduction to the present study	25
0.3	Overview of the Study	29
0.4	Some general findings of the thesis	31
	Footnotes to INTRODUCTION	35
CHAPTER ONE : REVIEW OF THE RELEVANT LITERATURE		36
1.0	Introduction	36
1.1	'Speech Exchange Systems : Definition, types and characteristics	37
1.1.1	A preamble	37
1.1.2	Some definitional comments	38
1.1.3	Interview and Conversation : Two types of Speech Exchange Systems	40
1.1.3.1	Interview	40
1.1.3.2	Conversation	42
1.1.4	Organisational and structural differences between interview and conversation	44

1.1.4.1	On the organisational level	44
1.1.4.2	On the linguistic/structural level	46
1.2	Linguistic complexity	48
1.2.1	The notion of 'complexity'	48
1.2.2	Research work on linguistic complexity	51
1.2.2.1	Spoken versus written discourse research	51
1.2.2.1.1	An overview	51
1.2.2.1.1	A critique	54
1.2.2.2	Text-type differentiation between spoken or written discourse	56
1.2.3	Towards a Model for the analysis of linguistic complexity of Speech Exchange Systems	57
1.2.3.1	Two contradictory views on the complexity of spoken discourse	57
1.2.3.2	Insights and Implications	62
1.2.3.3	A model for linguistic complexity	63
1.2.3.3.1	On the level of Lexis	65
1.2.3.3.2	On the level of Syntax	68
1.2.3.3.3	On the level of Information Structure	69
1.2.3.3.3.1	The Textually-Based Approaches	71
1.2.3.3.3.2	The Cognitively-Based Approaches	73
1.2.3.3.3.3	The Discoursally-Based Approaches	74
1.3	Approaches to the analysis of Speech Exchange Systems	76
1.3.1	Qualitative Approaches	76
1.3.1.1	The philosophically oriented approaches	78
1.3.1.2	The sociologically oriented approaches	80
1.3.1.2	The Discourse Analysis Approaches	81
1.3.2	Quantitative Approaches	84
1.3.3	Choosing the approach for the analysis of Speech Exchange Systems	85
1.4	The present investigation	88

1.4.1	Aims and Purposes	88
1.4.2	Field of research	90
1.5	Conclusion	91
	Footnotes to CHAPTER ONE	93

CHAPTER TWO : THE DATA : COLLECTION, DESCRIPTION AND ANALYSIS 95

2.0	Introduction	95
2.1	Collecting the Corpus of Data	96
2.1.1	Some Preliminary Considerations	96
2.1.2	The Interviewer	98
2.1.3	The Interviewees	98
2.1.4	The Setting	100
2.1.5	The Topics	101
2.1.6	The Transcription	103
2.2	The Analysis	104
2.3	Conclusion	106
	Footnotes to CHAPTER TWO	107

CHAPTER THREE : THE UNITISATION OF SPOKEN DISCOURSE 109

3.0	Introduction	109
3.1	The unitisation of spoken discourse	110
3.1.1	Problems and perspectives	110
3.1.2	Types of spoken discourse modes and the unit of analysis	110
3.1.3	Analytical approaches and trends	117
3.1.3.1	Approaches based on prosodic cues	118
3.1.3.2	Approaches based on syntactic cues	120
3.1.3.3	Approaches based on cognitive cues	125

3.1.4	Discrepancies between prosodic and syntactic units	126
3.1.5	Uses of units in the literature	127
3.2	Choosing a unit of analysis for spoken discourse	131
3.2.1	Free Standing Syntactic Unit	133
3.2.1.1	Major Units	135
3.2.1.2	Minor Units	139
3.2.1.3	Examples from the data	140
3.3	Quantification of the data : Words, turns and units	142
3.4	Conclusion	148
	Footnotes to CHAPTER THREE	150

CHAPTER FOUR : LEXICAL STRUCTURE OF INT AND CON 151

4.0	Introduction	151
4.1	Lexical Studies in the Literature	152
4.1.1	Lexical complexity and lexical density	152
4.1.2	Importance of lexical studies	154
4.2	Lexical Density	157
4.2.1	Definition, technique and methodologies	157
4.2.2	Some theoretical and practical problems	158
4.2.2.1	Lexical items and words	158
4.2.2.2	Lexical Vs Grammatical Words	159
4.2.2.3	Other more practical problems	161
4.2.3	Previous research on LD in types of discourse	164
4.3	Inter-text lexical variability in LD	168
4.3.1	Method of Analysis	168
4.3.2	The Results	169
4.3.2.1	Overall LD in INT and CON	169
4.3.2.2	Repetition of L and G words in INT and CON	170
4.3.2.3	LD in INT and CON excluding repetition and interviewer speech	171

4.3.2.4	Postgraduate/Undergraduate variation in LD	172
4.3.2.5	LD in the individual's output	173
4.3.3	General discussion of results	175
4.3.3.1	Differences from other findings	175
4.3.3.2	Inter-text/subject variability in LD	178
4.4	Intra-Text lexical variability in LD	180
4.4.1	Some preliminary points	180
4.4.2	A functional analysis	182
4.4.3	Kinds of functional units	183
4.4.4	The analysis	189
4.4.5	Results of the investigation	192
4.4.6	General discussion	194
4.5	Summary and Conclusions	200
	Footnotes to CHAPTER FOUR	204

CHAPTER FIVE : THE SYNTACTIC STRUCTURE OF INT AND CON 205

5.0	Introduction	205
5.1	Syntactic complexity in spoken discourse	206
5.1.1	Some preliminary considerations	206
5.1.2	Indices for syntactic complexity/simplicity	208
5.2	Features of syntactic complexity	210
5.2.1	Syntactic form and function	210
5.2.2	Embedding	212
5.2.3	Kinds of interdependency between clauses	215
5.3	Coordination in INT and CON	217
5.3.1	'Strong' and 'Weak' coordination	217
5.3.2	Coordination and syntactic complexity	220
5.3.3	Results	222
5.3.4	Discussion	226
5.4	Subordination in INT and CON	230
5.4.1	Some introductory remarks	230

5.4.2	The analysis	232
5.4.2.1	Nominal Clauses	233
5.4.2.1.1	'that' clauses (Subordinate Declarative Clauses)	234
5.4.2.1.2	Nominal Relative Clauses	237
5.4.2.1.3	Subordinate Interrogative Clauses	239
5.4.2.1.4	Extrapositional Nominal Clauses	242
5.4.2.1.5	Pro-forms in Nominal Clause position	244
5.4.2.1.6	to-Infinitive Nominal Clauses	246
5.4.2.1.7	-ing Nominal Clauses	246
5.4.2.1.8	Others	247
5.4.2.1.9	Discussion	249
5.4.2.2	Adjectival (Relative) Clauses	254
5.4.2.2.1	Restrictive Relative Clauses	255
5.4.2.2.2	Nonrestrictive Relative Clauses	257
5.4.2.2.3	Sentential Relative Clauses	258
5.4.2.2.4	The analysis	260
5.4.2.2.5	Discussion	263
5.4.2.3	Appositive Clauses	265
5.4.2.3.1	The analysis	270
5.4.2.3.2	Discussion	273
5.4.2.4	Adverbial Clauses	275
5.4.2.4.1	Characterisation of Adverbial Clauses in the data	277
5.4.2.4.2	Discussion	283
5.4.2.5	Comment Clauses	284
5.4.2.5.1	Results	286
5.4.2.5.2	Discussion	288
5.4.3	Subordination and syntactic complexity in SES	289
5.5	Elliptical Units	290
5.5.1	Results	294
5.5.2	Discussion	296
5.6	Some other features of relevance	298
5.6.1	Discourse Markers and Reaction Signals	298

5.6.2	Results and Discussion	301
5.7	Inter-individual differences in SC in INT and CON	304
5.8	Summary and conclusion	306
	Footnotes to CHAPTER FIVE	312

CHAPTER SIX : THE STRUCTURE OF INFORMATION IN INT AND CON 313

6.0	Introduction	313
6.1	Some preliminary considerations	314
6.1.1	Terminology	314
6.1.2	Information status in SES : An illustrative example	315
6.1.3	The problem	317
6.1.4	Levels and technique of analysis	320
6.2	The linguistic representation of information structure	323
6.2.1	Linguistic forms and information structure	323
6.2.2	Reference and referential entities in spoken discourse	327
6.3	Approaches and arguments of relevance	330
6.3.1	Arguments within the 'textual approaches' to IS	331
6.3.2	Arguments within the psycholinguistic approaches	333
6.3.2.1	Chafe's account of information structure	333
6.3.2.2	Other psycholinguistic accounts of relevance	335
6.3.3	Implications for an extended theory	335
6.4	The 'ASSUMED FAMILIARITY' taxonomy	337
6.4.1	Why the 'ASSUMED FAMILIARITY' taxonomy	337
6.4.2	General characteristics	338
6.4.3	A critical assessment of the taxonomy	342
6.4.4	Prince's application of the taxonomy	344
6.4.5	Some necessary modifications to the AF taxonomy	346
6.5	The present investigation	348

6.5.1	Methodology	348
6.5.2	New Entities	350
6.5.2.1	Brand New Entities	350
6.5.2.2	Unused Entities	354
6.5.2.3	New Entities in INT and CON : Discussion	358
6.5.3	Inferable Entities	360
6.5.3.1	Comparative Inferables (CIs)	361
6.5.3.2	Inferables from previous and subsequent discourse	364
6.5.3.3	Inferables from the 'Outside World'	366
6.5.3.4	Inferables from Logical Interpretation	369
6.5.3.5	Generic Inferables	371
6.5.3.6	Descriptive Inferables	373
6.5.3.7	Other Inferables	375
6.5.3.	Containing Inferables	377
6.5.3.9	Inferable Entities in INT and CON : Summary and discussion	379
6.5.4	Evoked Entities	383
6.5.4.1	Textually Evoked Entities	383
6.5.4.1.1	Third Person Pronouns	384
6.5.4.1.2	Extrapositional 'it'	388
6.5.4.1.3	Repeated Entities	390
6.5.4.1.4	Relative Evoked Entities	395
6.5.4.1.5	Textual Deictic Demonstratives	397
6.5.4.1.6	'Others' Textually Evoked Entities	401
6.5.4.1.7	Textually Evoked Entities in INT and CON : Discussion	402
6.5.4.2	Situationally Evoked Entities	404
6.5.4.2.1	First Person Pronouns	405
6.5.4.2.2	Second Person Pronouns	409
6.5.4.2.3	'Others' Situationally Evoked Entities	411
6.5.4.2.4	Situationally Evoked Entities in INT and CON : Discussion	415

6.6 Information Structure in INT and CON : Summary and general discussion 416
6.7 Conclusion 421
Footnotes to CHAPTER SIX 422

CHAPTER SEVEN : SUMMARY AND CONCLUSION 424

7.0 Introduction 424
7.1 Organisational features of INT and CON 425
7.2 Structural features of INT and CON 426
7.3 Conclusions 437
7.4 Some theoretical and applied implications 441
7.5 Limitations of the study 443
7.6 Suggestions for future research 444

REFERENCES 448

LIST OF TABLES

Table	2.1	The participants in the INTs and CONs	99
Table	3.1	Total number of words in INT and CON	143
Table	3.2	Number of words for each subject in INT and CON	144
Table	3.3	Number of turns and their OPTWs in INT and CON	145
Table	3.4	Major and Minor Units in INT and CON	146
Table	3.5	OPTWs of Major and Minor Units in INT and CON	146
Table	3.6	Distribution of Major Units in terms of words, clauses, clause complexes and ellipticals in INT	147
Table	3.7	Distribution of Major Units in terms of words, clauses, clause complexes and ellipticals in CON	147
Table	4.1	Lexical Density of six subtypes of spoken English texts (from Stubbs 1986)	166
Table	4.2	Lexical Density in five types of spoken discourse as reported by Hassan (1988 : Unpublished data)	166
Table	4.3	Overall Lexical Density in INT and CON (including repetition and interviewer speech)	169
Table	4.4	Lexical Density in INT and CON excluding repetition and interviewer speech	171
Table	4.5	Lexical Density in Interviewer input in INT and CON	173
Table	4.6	Lexical Density in subject's output in INT and CON	174
Table	5.1	OPTWs of clause complexes having 'and', 'but' and other coordinating conjunctions in INT and CON	224
Table	5.2	OPTWs of 'strong' and 'weak' coordinated clause complexes in INT and CON	225
Table	5.3	OPTWs of all coordinatives in INT and CON	226
Table	5.4	OPTWs of 'that' and 'zero' NCls in INT and CON	236

Table 5.5 OPTWs of Nominal Relative Clauses in INT and CON 238
Table 5.6 OPTWs of Subordinate Interrogative Nominal Clauses in INT and CON 241
Table 5.7 OPTWs of Extrapositional Nominal Clauses in INT and CON 243
Table 5.8 OPTWs of Proforms Nominal Clauses in INT and CON 245
Table 5.9 OPTWs of Nonfinite Nominal Clauses (-ing, to-Inf. and 'others' in INT and CON 248
Table 5.10 OPTWs of all Finite and Nonfinite Nominal Clauses in INT and CON 251
Table 5.11 OPTWs of all Nominals (finite and nonfinite) in INT and CON 252
Table 5.12 OPTWs of Restrictive and Nonrestrictive Relative Clauses in INT and CON 260
Table 5.13 OPTWs of all types of Relative Clauses in INT and CON 261
Table 5.14 OPTWs of Nonrestrictive and Restrictive Apposition in INT and CON 271
Table 5.15 OPTWs of all Appositive Clauses in INT and CON 273
Table 5.16 OPTWs of Adverbial Clauses in INT and CON 282
Table 5.17 OPTWs of Comment Clauses in INT and CON 287
Table 5.18 OPTWs of Elliptical Units in INT and CON 296
Table 5.19 OPTWs of Discourse Markers and Reaction Signals in INT and CON 301
Table 5.20 MEAN OPTWs of all syntactic features for Postgraduate and Undergraduate Students in INT and CON 305
Table 5.21 Total OPTWs of all syntactic features in INT and CON 307
Table 6.1 Raw figures and percentage of occurrence of each

of the AF Categories in relation to each other in INT and CON 350
Table 6.2 OPTWs of Brand New Entities in INT and CON 354
Table 6.3 OPTWs of Unused Entities in INT and CON 357
Table 6.4 OPTWs of Comparative Inferables in INT and CON 364
Table 6.5 OPTWs of Adjacent Inferables in INT and CON 366
Table 6.6 OPTWs of Situational Inferables in INT and CON 368
Table 6.7 OPTWs of Logical Inferables in INT and CON 370
Table 6.8 OPTWs of Generic Inferables in INT and CON 372
Table 6.9 OPTWs of Descriptive Inferables in INT and CON 374
Table 6.10 OPTWs of Other Inferables in INT and CON 376
Table 6.11 OPTWs of Containing Inferables in INT and CON 378
Table 6.12 Third Person personal, reflexive and possessive pronouns 384
Table 6.13 OPTWs of Third Person Pronouns in INT and CON 387
Table 6.14 OPTWs of Extrapositional 'it' in INT and CON 390
Table 6.15 OPTWs of Repeated Evoked Entities in INT and CON 394
Table 6.16 OPTWs of Relative and Zero Relative Pronouns in INT and CON 397
Table 6.17 OPTWs of Evoked Deictic Demonstratives in INT and CON 400
Table 6.18 1st Person and 2nd Person personal, reflexive and possessive pronouns 405
Table 6.19 OPTWs of 1st Person Situationally Evoked Entities in INT and CON 408
Table 6.20 OPTWs of 2nd Person Situationally Evoked Entities in INT and CON 411
Table 6.21 OPTWs of Time Deixis, Place Deixis and Others Situationally Evoked Entities in INT and CON 414

LIST OF FIGURES AND DIAGRAMS

1. FIGURES

Fig. 1.1	A Model for the analysis of linguistic complexity In INT and CON	64
Fig. 2.1	Diagram showing the sitting posture of the participants in the interview encounters	101
Fig. 4.1	Percentage of repeated G and L words in INT and CON	170
Fig. 4.2	Mean percentage of values of Lexical Density of Postgraduates and Undergraduates in INT and CON	172
Fig. 4.3	Percentage of frequency of occurrence of Major Units according to function	190
Fig. 4.4	Percentage of 'Lexical Density' Units in INT and CON	191
Fig. 4.5	Percentage of Low Lexical Density Units in INT and CON	192
Fig. 4.6	Percentage of Medium Lexical Density Units in INT and CON	193
Fig. 4.7	Percentage of High Lexical Density Units in INT and CON	193
Fig. 5.1	Percentage of frequency indices for coordinating conjunctions in INT and CON	223
Fig. 5.2	Percentage of subordinate clauses in INT and CON	233
Fig. 5.3	Percentage of all subcategories of Nominal Clauses in INT and CON	250
Fig. 5.4	Percentage of occurrence of each Relative Pronoun in INT and CON	262
Fig. 5.5	Frequency of occurrence of all Adverbial Clauses in INT and CON	277
Fig. 5.6	Percentage of types of Comment Clauses in INT and CON	286
Fig. 5.7	Percentage of occurrence of Elliptical Units in	

	INT and CON	295
Fig. 6.1	Percentage of occurrence of Inferable Subcategories in relation to each other	380
Fig. 6.2	Percentage of Relative and Zero Relative Pronouns in INT and CON	396
Fig. 6.3	Percentage of occurrence of Textually Evoked Subcategories in INT and CON	403
Fig. 6.4	Percentage of distribution of 'Generic' and 'Non-Generic' 'YOU' in INT and CON	410
Fig. 6.5	Percentage of Situationally Evoked Entities in INT and CON	415
Fig. 6.6	Percentage of occurrence of each of the AF categories in relation to each other in INT and CON	417

2. DIAGRAMS

Diagram 6.1	Phases of Information Production and Exchange	316
Diagram 6.2	Prince's 'ASSUMED FAMILIARITY' Taxonomy	339
Diagram 6.3	Brown and Yule's modification of Prince's ' ASSUMED FAMILIARITY' Taxonomy	341
Diagram 6.4	Prince's 'ASSUMED FAMILIARITY' Taxonomy with the proposed modifications	347

CONTENTS OF VOLUME TWO

APPENDIX ONE : THE DATA 3

INT I 3

CON I 10

INT II 32

CON II 40

INT III 52

CON III 60

INT IV 70

CON IV 79

INT V 93

CON V 102

INT VI 107

CON VI 121

INT VII 135

CON VII 142

INT VIII 154

CON VIII 163

APPENDIX TWO : NUMERICAL DATA FOR THE FIGURES IN THE THESIS 180

1. Numerical data for Fig. 4.1 180

2. Numerical data for Fig. 4.2 180

3. Numerical data for Fig. 4.3 181

4. Numerical data for Fig. 4.4 181

5. Numerical data for Fig. 4.5 182

6. Numerical data for Fig. 4.6 182

7. Numerical data for Fig. 4.7 183

8. Numerical data for Fig. 5.1 184

9. Numerical data for Fig. 5.2 184

10. Numerical data for Fig. 5.3 185
11. Numerical data for Fig. 5.4 185
12. Numerical data for Fig. 5.5 186
13. Numerical data for Fig. 5.6 186
14. Numerical data for Fig. 5.7 187
15. Numerical data for Fig. 6.1 188
16. Numerical data for Fig. 6.2 188
17. Numerical data for Fig. 6.3 189
18. Numerical data for Fig. 6.4 189
19. Numerical data for Fig. 6.5 190
20. Numerical data for Fig. 6.6 190

APPENDIX THREE : LIST OF AUDIO-TAPES (Attached to Volume Two)

	Side One	Side Two
CASSETTE I	INT I	INT II
	CON I	CON II
CASSETTE II	INT III	INT V
	CON III	CON V
	INT IV	INT VI
	CON IV	CON VI
CASSETTE III	INT VII	INT VIII
	CON VII	CON VIII

LIST OF ABBREVIATIONS AND PARTICIPANT INITIALS

A. ABREVIATIONS

ADCL	Adverbial Clause
AI	Adjacent Inferable
AF	Assumed Familiarity
APPCL	Appositive Clause
ARG	Argumentation
BNE	Brand New Entity
CA	Conversation Analysis
CCL	Comment Clause
CI	Comparative Inferable
CON	Conversation
CONT. I	Containing Inferable
DESC	Description
DA	Discourse Analysis
DI	Descriptive Inferable
DM	Discourse Marker
ELI	Elicitation
EU	Elliptical Unit
EXP	Exposition
FSP	Functional Sentence Perspective
FSSU	Free Standing Syntactic Unit
G	Grammatical (Word)
GI	Generic Inferable
IE	Inferable Entity
IG	Information (General)
INT	Interview
IP	Information Personal
IS	Information Structure
IU	Idea Unit
L	Lexical (Word)
LB	Lower Boundary
LD	Lexical Density
LI	Logical Inferable
NAR	Narration
NCL	Noun Clause
NP	Noun Phrase
OI	Other Inferable
P	Pretonic
PG	Postgraduate
OPTWs	Occurrence Per Thousand Words
RCL	Relative Clause
RP	Relative Pronoun
RS	Reaction Signal
SAT	Speech Act Theory
SC	Syntactic Complexity

B. PARTICIPANT INITIALS

AB	*INTERVIEWEE*
AM	*INTERVIEWEE*
BG	*INTERVIEWEE*
CB	*INTERVIEWEE*
DD	*INTERVIEWEE*
GM	*INTERVIEWEE*
HC	*INTERVIEWEE*
HK	*INTERVIEWEE*
JA	***INTEVIEWER***
JC	*INTERVIEWEE*
JH	*INTERVIEWEE*
KS	*INTERVIEWEE*
KSH	*INTERVIEWEE*
KW	*INTERVIEWEE*
PM	*INTERVIEWEE*
RF	*INTERVIEWEE*
RH	*INTERVIEWEE*

SD	Standard Deviation
SEE	Situationally Evoked Entity
SES	Speech Exchange System
SI	Situational Inferable
TEE	Textually Evoked Entity
TU	Tone Unit
TG	Tone Group
T	Tonic
UB	Upper Boundary
UE	Unused Entity
UG	Undergraduate
V	Coefficient of Variation
WSRT	Wilcoxon's Signed Rank Test

TRANSCRIPTION CONVENTIONS

1) **General transcription conventions :**

..	a short unfilled pause
...	a lengthy pause
e::m	a filled pause
e:::m	a lengthy filled pause
::	a lengthened syllable
[	interruption
--	vocalised sound such as coughing, clearing throat, etc. Also used to denote unclear speech.
(he he)	laughter

2) **Special Conventions for CHAPTER SIX**

underlining	:	**Brand New Entities**
/ /	:	**Unused Entities**
( )	:	**Inferable Entities**
[ ]	:	**Situationally Evoked Entities**
{ }	:	**Textually Evoked Entities**
**	:	**Zero Relative Pronoun**

INTRODUCTION

0.1 INTRODUCTION TO THE AREA

One of the main fields of inquiry of current (socio)linguistic research, normally covered within the general rubric 'DISCOURSE ANALYSIS' is the study of 'natural' language, natural from the point of view of its **"spontaneity, unplannedness and casualness"** (cf. Stubbs 1983:33) as opposed to 'unnatural' language (i.e. **"planned, rehearsed, thought about, altered and edited"** (ibid) language. This distinction and the linguistic research covering its major themes and practices brought about a line of research specialising in language variation and language variability within and across language types (i.e. 'registers' in Halliday et al's 1964 terminology). Studies of this sort are mainly covered by the general field of **'SOCIOLINGUISTICS'**(1) and rely heavily on the social situation and context of use of utterances including the participants themselves, the subject matter, medium and function of discourse(2). These represent dimensions along which discourse types (spoken and/or written) are differentiated and compared. The outcome of this upsurge is a bulk of studies on text type differentiation on spoken versus written discourse, spoken versus spoken discourse and written versus written discourse.

Most of the studies on language variation within the different fields of (socio)linguistic research are carried out on spoken versus written discourse types (cf. Drieman 1962; De Vito 1964, 1965, 1966, 1967; Golub 1969; Ure 1971; O'Donnell 1974; Poole and Field 1976; Kroll 1977; Ochs 1979; Stubbs 1980; Chafe 1982; Akinnaso 1981, 1982; Farag 1986; Johns-Lewis 1987; see also the collection of papers in Tannen 1982; among many others). Some of these studies look at the differences and similarities of spoken and written discourse in terms of their internal linguistic structure (lexical, grammatical, discoursal) (e.g. Poole and Field op.cit.; Ure op.cit.; Kroll op.cit.), others look at the strategies employed by speakers and

writers (cf. Lakoff 1982, Farag op.cit.) while others focus their research on cognitive factors of discourse organisation and processing in both modes of language (e.g. Chafe 1982). From the linguistic point of view, very limited research work has been done on spoken versus spoken discourse types especially on the types of dialogic discourse often referred to as **"SPEECH EXCHANGE SYSTEMS"** (cf. Sacks et al. 1974) (Henceforth SES), such as conversations, interviews, meetings, debates, ceremonies and the like.

The general findings of the above-mentioned research work from the point of view of the linguistic structure and the conveyance of information are in the main controversial, inconsistent and varied though rich in size and insights (see CHAPTER ONE for review and discussion). Some have found that the language of SES is on the whole structurally complex, involving the use of complex clause type structures with low lexical densities and information packaging (cf. Halliday 1979, 1985b; Beaman 1984), compared to written discourse types that are generally agreed to have a more consistent syntactic structure with higher lexical densities and more information packaging and structure. Others have found the language of such systems to be syntactically less complex than written discourse types, less organised and integrated and more fragmentary (cf. Chafe 1982; Brown and Yule 1983a).

Drawing conclusions on the true nature of the various types of spoken discourse (especially SES) on the level of structure and information packaging by comparison with written discourse is not without its own shortcomings. This will be elaborated on in CHAPTER ONE below. Suffice it here to draw attention to the fact that both discourse modes display quite different features and employ greatly different strategies, such as those related to 'planning' (Ochs 1979), 'spontaneity and forethought' (Lakoff 1982), 'formality' (Crystal and Davy 1969) and so on. Each mode has certain characteristic features that might not be shared with the other mode (e.g.

phonological and extralinguistic features of spoken discourse). These features and strategies make each of the two modes unique in certain aspects and, on the whole 'unalike' and 'incomparable' at these levels. There are of course certain features of language structure and organisation that can be compared between the two modes of discourse. But for these to be used appropriately as **"Tertium Comparationis"** (James 1980) for making valid comparisons, there should be a tight control on such variables as participants, topic, formality of the situation and so on. Unfortunately, this is not well established in the present literature and most of the comparisons made lack such control. Even when such control is present, the findings concerning the characteristic features of each mode are only applicable in relation to each other. This means that they cannot be taken single sidedly to characterise the particular mode they belong to without being juxtaposed to the other mode. In other words, one cannot make generalisations about one single mode (spoken or written) from findings obtained from comparisons of both types. One cannot, for example, generalise from Halliday's conclusion stated above that spoken discourse is syntactically complex without relating this finding to the complexity of written discourse and vice versa. What is needed, then, is comprehensive research work on the linguistic structure of either type of discourse (spoken or written) each in its own right. Variety differentiation within and across varieties of one particular mode is more illuminating and insightful to the recognition of the true nature of that mode. Such comparisons will give a clearer picture of the structure of discourse, its real complexity and behaviour. Only after this is achieved can one make comparisons of the complexities involved in both types of discourse on comparative and contrastive bases.

The present knowledge of the nature of spoken interactional discourse is richer on the level of its overall organisation than on the level of its structure. This is due to the great amount of research on the organisational features of conversational discourse conducted within the

developing line of ethnomethodological research pioneered by Sacks and his associates during the past three decades or so. This school has developed a deep interest in the language of conversation and its main features such as turn-taking, repair, sequences, closings and openings and so on. Comparatively, less is known on the structure of speech exchange systems and the level of its complexity and variability across speech types. Traditional linguistic and stylistic research (see, for example, Crystal and Davy 1969; Quirk et al. 1985) has provided some rich insights into how language varieties should be dealt with within the current lines of research. There are still many unresolved issues related to the complexity of sequential speech exchange types from the point of view of their linguistic structure both in lexis and grammar and most of all in the representation of information units within discourse. Despite the fact that it is sometimes difficult to divide SES in sentential terms, the tendency in traditional research was to treat spoken discourse like any other piece of language away from its real context. This has the disadvantage of producing very general and abstract descriptions of language in its spoken form. It does not reflect in any real sense the characteristics of specialised types of spoken discourse and their variability from one situation to another. Also, it does not inform the analyst of the factors influencing speakers' speech in various situations. This is one of the reasons why this research has been undertaken.

0.2 INTRODUCTION TO THE PRESENT STUDY

The present research sets out to investigate empirically the structure of two types of SES, interview (INT) and conversation (CON)[3] produced by the same speakers, from the point of view of linguistic complexity on three levels of language expression and organisation : lexis, syntax and information structure. Two sets of data were recorded for the analysis. The first set was obtained from conducting preplanned, relatively formal interviews with pairs of students and the second set from unplanned,

informal conversations of these students with each other. The general aim is to investigate in depth the structure of speech exchange types of discourse in informal, naturally and spontaneously occurring conversation and formal, information-eliciting interview. The specific aims are, first of all, to investigate the nature of variability between the two types of speaker output on the level of linguistic structure and complexity and secondly to examine the type of inter-individual variability in these issues.

A second aim is to explore the characteristics of the dynamic interaction in the two types of SES under investigation in relation to such participant variables as age, and level of education, which cast explanatory light on the linguistic variation encountered. Again, both inter- and intra-subject variability will be looked at in the analysis chapters. The hope is to present a contribution towards the analysis of spoken interactional discourse in terms of its internal structure and the ways in which information is conveyed and organised.

The analysis of SES carried out in this thesis can be taken to fall under the general line of sociolinguistic investigation of the type referred to by Trudgill (1978:12) as **"secular linguistics"** which mainly aims to examine linguistic structure and variability across language types. Clearly studies of this type are basically empirical in nature involving direct (mostly statistical) analysis of data of language used in its social context. The incorporation of quantitative techniques into the study of the language of SES has some advantages. First of all, it helps to provide a clearer picture of the structures used and their functions in the stream of speech, especially in so far as the organisation and packaging of information is concerned. This area of linguistic research is thought to be considerably lacking in present day research. Secondly, it helps to provide **"a more objective approach since the methods involved are essentially inductive"** (Farrugia 1988:20). Thirdly and perhaps most importantly, it may give

researchers valuable insight into making valid generalisations which can be exploited for qualitative research on spoken discourse in general and the language of SES in particular.

As will be made clear in CHAPTER TWO below, in the collection of data, two factors were borne in mind. First of all, interviewing total strangers would not give the typical kind of data (especially for the conversational encounters), that could be taken as 'natural', 'informal' and 'spontaneous'. This is why pairs of student friends (both males and females) were asked to participate in the INT and CON encounters. Secondly, and perhaps most importantly, the principle of controlling certain variables that might influence the output of speakers had to be taken into consideration so that we could get the desired comparable sets of data. The factors that were thought to be most obviously influencing the data were those of **'participants, 'plannedness', 'spontaneity', 'topic', and 'formality of the speech situation'**.

Accordingly, it was necessary, at first, to take great care in choosing the right interviewer for performing the complex task of interviewing and the right students to be interviewed. For this purpose an interviewer was chosen, one of the chaplains at Aston University, who had previous experience and training in the art of interviewing (see CHAPTER TWO below). A population of 16 students, eight males and eight females, grouped in eight pairs, each pair carefully chosen to represent intimate friends, was asked to participate in the interviews. The interviewees were all native speakers of English, coming originally from different parts of Britain. Their age range was between 20 and 30 years. Six of them (the first three pairs) were postgraduate students and the remaining five pairs were undergraduates. This classification helped in examining the influence of educational maturity on the choice of particular lexical and syntactic structures. Care was also taken to choose students who were educated in the fields of religion or politics. This proved to be an important factor in

having certain features of structure tested, differentiated and accounted for.

The interviews were preplanned in the sense that the students were told in advance that they would be asked some questions relating to their previous life experiences, social and religious relations and preferences, present life at the university and new friends and relations. Elicitation techniques were all left to the experience of the interviewer and his ability to handle interviewing situations. The conversation encounters consisted of the casual chat that followed the interview sessions during which the conversants were not aware that they were still being monitored and recorded. The analysis is based on recordings obtained from the eight sessions of interviews and eight encounters of casual conversations. The total number of words of both types of speech mounts to approximately 35000 words. The analysis comprises calculations of the interviewees' output only, each pair in relation to each other (i.e. with the interviewer input excluded from the calculations) in both the interview talk (with which the interviewer was directly involved) and the conversation talk (in the recording of which the interviewer was not directly involved and was not physically present in most of the encounters) and thereby to maintain identity of speaker for INT and CON. More about the process of data collection will be given in CHAPTER TWO below.

The method used for analysing the data is mainly statistical involving the use of computer programmes, tables, graphs, figures and statistical tests of significance. It is worth mentioning here that all the tables, graphs and figures as well as the examples from the data are numbered according to the chapter they happen to be in. So for example in **e.g. 2.9** the number '**2**' refers to **CHAPTER TWO** and the number '**9**' refers to the sequence of the example within the chapter. This is applicable to the tables and figures as well.

0.3 OVERVIEW OF THE STUDY

The thesis is planned to consist of two volumes. The first volume represents the main body of the thesis and bibliography and the second volume contains the appendices which consist of the data transcripts of the eight INTs and CONs and also some statistical tables which represent the numerical values of the figures used in the main body of the text . Audio cassettes of the eight encounters of INT and CON are also attached to the second volume of the thesis.

The main body of the thesis is planned as follows :

CHAPTER ONE presents a review of the literature on the structure of spoken interactional discourse, particularly those types of SES that have relevance to this study. Findings of previous work on text type differentiation of spoken versus written discourse and spoken versus spoken discourse are reviewed and discussed. Then a review of the notion of linguistic complexity follows with relation to three main types of complexity : lexical, syntactic and informational. This will be followed by a critique of the quantitative and qualitative techniques for sociolinguistic research of text type differentiation and description and the choice of the right method for the research at hand.

CHAPTER TWO describes the methods and techniques used in data collection and elaborates on the choice of the interviewer and interviewees, their general social and educational backgrounds and their role in making this research possible.

CHAPTER THREE reviews critically the methods used in the literature for the segmentation into units (*'unitisation'* for short) of spoken discourse for the purpose of (socio)linguistic analysis. It shows the problems and difficulties encountered by analysts in segmenting spoken discourse

(especially SES types) in terms of the traditional grammatical notion of 'sentence' and other grammatical notions of 'clause', 'T-Unit', etc. Then a method of segmenting INT and CON data into units, based on both semantic and syntactic criteria is presented and discussed in relation to the other methods available in the literature. The analysis of the data of INT and CON presented in CHAPTERS FOUR, FIVE and SIX is mainly based on this unit. All the data is segmented into these units following certain criteria which will be elaborated on in the chapter. The final results represent calculations of all lexical and syntactic features contained within these units. Linguistic complexity is assessed on the basis of the comparatively high occurrence of certain features over others, as will be seen in the analytical chapters.

CHAPTER FOUR presents an overall investigation of the lexical structure of the data of INT and CON and the variability in lexical density on the intra and inter- individual levels. A critique of the technique of lexical density as a measure of text type differentiation is presented within the chapter. A formal and a functional analysis will also be produced in the chapter to assess which types of functional units of discourse carry more lexical densities than others, under what conditions and what factors influence these densities.

CHAPTER FIVE focuses on the important issue of syntactic complexity in INT and CON. This is arrived at by computing the frequencies of occurrence and presence of certain features of syntactic structure (rather than others) that are taken to be indices of syntactic complexity. Such features include the use of embedding/subordination and complex clause structures. It also covers other internal features within the units of discourse such as the greater use of apposition, right and left dislocations, word order, ellipsis and so on. Discussion of the use of each of these features is provided at the end of each section and comparisons with other (comparable) research work in the literature are also presented within

these discussions.

CHAPTER SIX looks in considerable detail at the notion of information structure and packaging in the data of INT and CON. It combines features of CHAPTERS FOUR and FIVE as both lexis and syntax have a role in the expression of information within discourse. The aim is to find out the types of complexities involved within INT and CON in terms of the linguistic representation of information structure. Prince's analytical model of information structure is modified and applied for that purpose.

Finally **CHAPTER SEVEN** presents a summary and conclusion of all the chapters and of the general and more specific results of the analysis. It also provides a section on the limitations of the work and suggestions for future research on the notion of linguistic complexity of speech exchange types of discourse.

0.4 SOME GENERAL FINDINGS OF THE THESIS

The general findings of the research point to two main conclusions :
A. That spoken interactional discourse is on the whole complex on the three levels of analysis considered for this work (i.e. lexis, syntax and information structure). First of all, higher lexical densities are obtained from the data of INT and CON than are reported by other researchers who investigated lexical density in spoken discourse. Secondly, on the level of syntax, the results show the language of both INT and CON to contain a great number of syntactically complex units having multiple embedding and subordination. They also contain other features which point to the presence of syntactic complexity in the two types of SES under examination.

On the level of the linguistic representation of information structure and organisation, the results also show the language of both INT and CON

to be relatively complex. Besides the implication this phenomenon might yield for the issue of linguistic complexity of discourse, it might also give implications for the presence of another type of complexity and that is complexity on the level of information processing, an issue which is not going to be tackled in this thesis. The packaging of 'Inferable' entities (see Section 6.5.3 below) within discourse which cause a greater processing load on the language processor is clearly recurrent in the data. Referential entities, which also bear on the issues of linguistic complexity and information processing, are frequently used in the data, sometimes with greater repetition and longer span of use than one would expect to find in any other types of discourse (spoken or written).

Turning now to the issue of intra-individual variability in the use of the above-mentioned criteria, the results show consistently that the language of INT is more complex than that of CON on the three levels of analysis. There is a slightly higher lexical density in the INT sessions than in the CON encounters. The language of INT has been observed to contain more syntactically complex features and inferable entities (cf. Prince 1981) than that of CON. There is some inconsistent variability in the use of New and Given information in INT and CON but the difference is not statistically significant.

These results confirm the claim often made in sociolinguistic research (cf. Crystal and Davy 1969; Hudson 1980; Coates 1986) that speakers vary their use of language from one speech exchange encounter to another and according to the social situation they are using the language in. The most obvious factors that might have their influence on this variation are related to what Crystal and Davy (op.cit:74) group under the dimension of **'STATUS'**. This includes such factors as formality/informality, social relations between communicators (e.g. intimacy relations, business relations and so on). The purpose of the encounter and the topics discussed might have their role as well. Also of relevance here is the factor

of how much the topics under discussion are preplanned and prepared as opposed to the natural, unplanned conversational discourse. All these factors seem to have their influence on the differences between the two types of talk under investigation, INT and CON.

Another interesting finding which the results of this research indicate is that with certain features there appears to be some inter-subject variability in the use of the above-mentioned criteria. This has been noticed to increase with those features that have a higher degree of occurrence than others. The higher Standard Deviation (SD) and Coefficient of Variation (V) values shown with the tables express the degree and range of variation between speakers. Interestingly, more educationally mature students display higher linguistic complexity results especially so on the lexical and syntactical levels. This confirms the effect of educational maturity on the complexity of the linguistic output in sequential discourse, as suggested by some researchers (e.g. Stubbs 1986).

The issue of inter-subject variability which has been noticed to occur between speakers might be due to factors related to the social and educational status of the participants, previous training, etc. Evidence for this has been provided by comparing the results obtained from the postgraduate and undergraduate students and those students who are more orientated towards politics and religion. Previous training in these fields seems to have an effect on the linguistic output of speakers.

Another possible explanation for the phenomenon of inter-subject variability might be the amount of data obtained from each speaker. Obviously, those speakers who produced more spoken output gave higher OPTWs than others and showed higher significance in the results. This confirms the need for a bigger population of speakers to get more valid results.

The data of INT and CON analysed for the present study represent two varieties of spoken interactional discourse. The findings obtained can only reflect the general trends of speakers to vary their speech according to the difference in the speech situation. Examining other types of spoken interactional discourse such as debates, meetings, ceremonies, discussions, and other types of interviews and conversations, might confirm the findings of the present research and might add to our understanding of the language of speech exchange systems. This is going to be emphasised within the section on suggestions for future research. One important implication, however, which is yielded by the present investigation, is that in an analysis of big corpora of speech types features of linguistic complexity examined should be kept to a minimum so as to be dealt with in depth. Only through this can one make generalisations to include types of spoken discourse of the type exchanged by speakers in everyday communicative situations.

FOOTNOTES TO INTRODUCTION

1. See Trudgill (1978) for an elaboration on the different use and definitions of the term 'sociolinguistic' in current research.

2. These features are well elaborated on in Halliday et al's work of (1964) under the rubrics of **"Field, Mode and Style of Discourse"**, and later expounded by Halliday (1977), Gregory (1967), Gregory and Carroll (1976).

3. All throughout the thesis, the abbreviations INT and CON will only be used to refer to the data of 'interview' and 'conversation' recorded and analysed in the present work, respectively. In all other cases, the full terms will be employed.

CHAPTER ONE

REVIEW OF THE RELEVANT LITERATURE

1.0 Introduction

The study of the linguistic structure of **"SPEECH EXCHANGE SYSTEMS"** (SES) has proved to be neither easy nor straightforward hence not adequately covered by linguistic investigation. Most researchers take it for granted that the language of SES is linguistically 'simple' (cf. Kroll 1977; Chafe 1982), 'fragmentary' (cf. Chafe op.cit.), 'incomplete' (cf. Gregory and Carroll 1976:39), 'chaotic' and 'syntactically untidy'. Precisely these features made such types of interaction uninteresting for some researchers (e.g. Chomsky 1965) but a valid field of enquiry for others (e.g. Garfinkel 1967; 1972; Sacks 1972; Sacks et al. 1974; Sacks and Schegloff 1979; Tannen 1982, 1984, Chafe op.cit.). With the advent of sociolinguistics, ethnomethodology, discourse analysis and conversation analysis, which partly or wholly take spoken discourse to be their prime tool of research, the language of 'SES' has come to be treated as an area of enquiry in its own right, an area which is **"open to analysis"** (Stubbs 1983:37). However, considerably little research work is found within linguistics proper concerning the language of SES from the point of view of structural complexity, especially on the level of differentiation between one variety type and another, except perhaps for those research works that have been carried out within the field of text type differentiation between spoken and/or written discourse (see, for example, Drieman 1962; de Vito 1965, 1966, 1967; Labov 1972b; Crystal and Davy 1969; Poole and Field 1976; Kroll 1977; Halliday 1979; Beaman 1984; Farag 1986), whose findings are highly interesting and relevant to the present investigation.

The major task of the present chapter is to review the relevant literature on SES from the point of view of definition, characterisation,

structure and linguistic complexity in general and in the context of text type differentiation in spoken discourse in particular. The second task is to provide an overall characterisation of terminology used throughout this thesis. However, the chapter does not attempt to summarise what has been written on discourse, since a) this has been done elsewhere (e.g. Brown and Yule 1983a; Stubbs 1983; Levinson 1983) and b) this would not be entirely relevant to the issue on which the thesis focuses. Rather, the chapter specifically orientates itself to the literature on spoken discourse variations, SES structure and linguistic complexity.

1.1 'Speech Exchange Systems' : Definition, types and characterisation

1.1.1 A preamble

Since the type of language used by any two speakers, whether in formal or informal settings, takes place in variable situations in real time and place and is mostly spontaneous and unplanned, there is bound to be some degree of variability in its use from one context to another, the level of variability being dependent on many factors which are related to the speakers themselves and to the social context which they happen to be exchanging talk within. The most obvious markers of this variability are manifested in the use of formal structures of the language exchanged i.e. in its phonology, lexis and syntax. The present research looks at the type of variation used by speakers in interview and casual conversation situations on two of these levels : lexis and syntax. Since these levels are the surface carriers of information in running discourse, a third dimension has been added to the analysis, which is that of information structure and speakers' variability in its employment in interview (INT) and conversation (CON). The major aim is to investigate the issue of linguistic complexity and to see how far the language of INT differs from that of casual conversation on these three levels. However, before elaborating on these levels of analysis, it is essential at first to discuss the major terms

concerning types of spoken discourse which are going to be investigated in this research.

1.1.2 Some definitional comments

Initially, spoken interactional discourse, as a medium of language in use varies in accordance with a number of situational and contextual factors. First of all, it may vary in accordance with the number of its users e.g. whether it is produced by one speaker in a particular situation **(monologuing)** or whether it is exchanged between two or more speakers in normal everyday interactional situations **(conversing)** (cf. Gregory 1967:188). Secondly, it may vary according to three major factors that are related to the situation itself such as the level of formality/informality, spontaneity/non-spontaneity (ibid), and plannedness/unplannedness (cf. Ochs 1979). These are not going to be elaborated on here as this has been done elsewhere (cf. Gregory and Carroll 1976; Baka 1989). The distinction monologuing/conversing is an important one since it involves the use of what Sacks et al (1974) refer to as the turn-taking mechanism. During monologues language users produce speech on their own, i.e. with the other participants if any, just listening and not joining in, such as in most types of classroom teaching or sports commentaries (cf. Gregory and Carroll op.cit.:40). 'Conversing', on the other hand, involves at least two speakers exchanging turns of talk in a highly organised manner. The ordering of turns among speakers is analysed in terms of '**adjacency pairs**' (Sacks et al. op.cit.). The length, type and ordering of these turns depend mostly on the type of conversational setting and the purpose of the communicative situation. The types of spoken interactional discourse that fall under the rubric of 'conversing' are the ones that are of interest in this thesis.

There seems to be some confusion in the literature regarding the use of terminology to refer to the conversational varieties of spoken discourse.

Levinson (1983:318) refers to them as **"conversational activities"**; others have used the terms **"conversational discourse"** or **"conversational interaction"** (e.g. Brown and Yule 1983a). Most of the confusion seems to arise from the use of the term 'conversational' although most researchers distinguish between their use of the terms above and that of 'conversation'. However, as will be seen below, the term 'conversation' per se has also caused some confusion. In addition to that, most of these terms are too general in that they cover a multiplicity of features that might be linguistic or extralinguistic. So for the sole purpose of avoiding this confusion, it seems reasonable to adopt a term that has no direct reference to 'conversation' and reserve the latter to its original meaning of 'chat' (see below). This we find available in Sacks et al's (op.cit.) use of **"Speech Exchange Systems" (SES)**, which is going to be adopted throughout the thesis to describe all kinds of formal and informal talk exchanged by speakers via the use of turns and adjacency pairs. Examples of such systems include casual conversation, interviews, meetings, ceremonies, debates, discussions and the like.

It is obvious that the distinction between SES types is apparent from the setting in which they take place, participants, topics, levels of formality, plannedness and spontaneity of talk. What is also reasonably clear is that they all involve the use of certain interactional features by which they can be distinguished without reference to direct linguistic analysis, for example length and organisation of turns, simultaneous and interruptive speech, repair, feedback etc. (see Sections 1.1.4 and 7.2 below).

What is not directly distinguishable, however, is the nature and complexity (as will be outlined in Section 1.2.1 below) of the linguistic output within these types of SES. This of course needs detailed controlled analysis of the language of each type individually and against other types of SES or other variations of language use (e.g. written discourse (cf. Chafe 1982), for the purpose of seeing whether there are any systematic

differences between them on the various levels of analysis chosen for making the comparison.

1.1.3 Interview and Conversation : Two types of Speech Exchange Systems

Interview and conversation, in the most general terms, are two 'varieties'(1) of what has been referred to above as **"speech exchange systems"**. The distinction between them as two general **"conversational activities"** (Levinson 1983:318) is clearly apparent from the presence of the interviewer in the interview situation as well as from the other situational and contextual criteria mentioned above. They both involve the use of direct face to face interaction with the exchange of turns and other necessary features of verbal and non-verbal interaction.

In the following sections the different definitions and realisations of interview and conversation in the literature and in relation to the data of INT and CON of the present study will be presented. Then a discussion of the few attempts that have tried to characterise them in relation to each other before moving to discuss the more relevant issue of linguistic complexity in SES in general and in the data of INT and CON in particular.

1.1.3.1 Interview

In its simplest forms the term 'interview' is used to describe a variety of direct person to person social encounters which are specially planned and organised for particular well defined purposes. It often entails an interviewer who is responsible for the act of initiating a piece of conversation mainly in the form of direct questions to be answered by another person (or persons), often referred to as 'respondents', 'interviewees' or 'subjects'. The time required for a single interviewing situation ranges from a few seconds, whereby two or three short sentences

are exchanged to an hour or more in extended interviews. Gorden (1969:3) explains that the term 'interviewing'

> "evokes images of a person taking a census, doing a sample opinion poll, interviewing prospective employees -all done in some setting in which the person's major activity is interviewing".

Like casual conversation, interviewing seems to be an inevitable act of social life which demands a rather wide variety of conversational events from the individual, who must be particularly prepared to meet the specific requirements of each and carry out the necessary stylistic and behavioural activities demanded by that event.

It is not possible and indeed not at all necessary to cite here all the definitions and characteristic features of 'interview' as used in the literature simply because this activity has been researched not only in linguistics but also in sociology, psychology, social psychology and in many other fields and practices (cf. Benney and Hughes 1956; Gorden 1969; Denzin 1970; Silverman 1971; Brenner 1978, 1981)(2).

A definition which seems to be the nearest possible one to the purposes of the present study comes from Denzin (op.cit.:123) who quotes Maccoby and Maccoby (1954:499) characterising the interview as :

> "face-to-face verbal exchange in which one person, the interviewer, attempts to elicit information or expressions or opinions or belief from another person or persons".

Some of the features implied in the above definition are also found in other types of SES including casual conversation. The main differences though lie in the levels of formality of the encounter, pre-planning of the encounter and the degree of spontaneity of speech (see discussion below).

The interviews recorded for the present analysis have many features in

common with Maccoby and Maccoby's interviews cited above. The main task of the interviewer in the present data was also to elicit as much information as possible from the interviewees, to know about their opinions and beliefs by asking them questions related to their social and educational life experiences. I have found this type of talk suitable for sociolinguistic analysis in comparison to the same speakers' natural, unmonitored type of chat, which, in most cases, the interviewees were fortunately unaware of. Besides, the interviewer's position as an active chaplain at the university and his relationship with the students through religious contact justified the need for the interview. These points will be elaborated on in CHAPTER TWO below.

1.1.3.2 Conversation

There has been considerable confusion concerning the term 'conversation' in the literature due to various, sometimes conflicting, interpretations of the term in disciplinary and interdisciplinary research. There seems to be also a general absence of a clear-cut taxonomy of spoken discourse modes into which 'conversation' can fit (Johns-Lewis 1986:208). However, there does not seem to be a great fusion of ideas of what goes on in conversation. The definitions and the descriptions of 'conversation' and its structure, which each of the disciplines offers seems to be distinct and not fully integrated with concepts from other disciplines, although there may be cross-reference.

Mcgregor (1984:208) expresses a similar point of view, as follows :

> "Yet definitions of 'conversation' are rare, and general remarks about it point uncomfortably to the dimly recognised complexities of multi-faceted behaviour which occurs in face-to-face interaction".

However, in spite of the absence of a convincing taxonomy of verbal interaction activities, there seem to be certain criteria which can be used

for the categorisation of conversation. Some of these criteria have been mentioned above. They include (among other things) 'plannedness', 'spontaneity' and 'formality'. Levinson's (1983) distinction between **'conversational activity'** and **'conversation'** and Sacks' et al's **'Speech Exchange Systems'** and the other notions concerning language variation mentioned above can also be helpful in establishing a rather clear-cut distinction between 'conversation' as an important activity of human interaction in its own right and the other types of SES. And although these types share with conversation such organisational features as 'turn-taking' (Sacks et al. 1974), 'closings' (Schegloff and Sacks 1973), 'repair' (Schegloff 1979), 'simultaneous speech' (Coates 1986) etc., they can in no way be called 'conversation' in the full sense of the term as they lack other features which are characteristic of everyday natural, informal and spontaneous conversation. This is defined by Levinson (op.cit.:284) as follows :

> ".. that familiar predominant kind of talk in which two or more participants freely alternate in speaking, which generally occurs outside specific institutional settings like religious services, law courts, classrooms and the like".

As such, conversation entails the presence of a number of 'interactants', at least two, who socially exchange turns of talk about a particular topic of interest. It is a social activity which, in turn, involves a number of subactivities such as cooperation, tolerance, mutual interest, common knowledge and a capacity to interpret speakers' meanings and intentions. It is conditionally unplanned, unthought of, spontaneous and natural which, **"although it may look chaotic when closely transcribed, is , in fact, highly ordered"** (Stubbs 1983:19) though it might not be ordered in the same ways as, for instance, written texts usually are.

1.1.4 Organisational and structural differences between interview and conversation

It seems clear enough now, looking at the selected definitions of interview and conversation discussed above, that the two varieties of spoken discourse are different on many levels, some of which are linguistic, others sociological or cognitive. This has been confirmed by researchers of various disciplines who seem to agree that, although the two speech modes share certain interactional features in common, they are quite different especially on the social and behavioural levels. This is clearly confirmed by the following quote from Ferguson (1975:10) who thinks that interviews, in general tend to be structured in a way that casual conversations are not. He states that :

Aston University

Content has been removed for copyright reasons

So it seems plausible to classify the differences between the two varieties of SES according to two levels. The first level is concerned with the social and behavioural organisation of speech in the two types and the second one is related to the structural features of the talk, which include the formal linguistic realisations of the language being used with all the features and categories entailed herein. Following is a brief outline of each of these levels of differences between the two types of SES :

1.1.4.1 On the organisational level

Included under this level are all features of turn taking, mechanism, adjacency pairs, side sequences, openings, closings, repairs, etc. (cf. Sacks et al 1974 and also see Section 1.3.1.2 for elaborative comments). Obviously there seems to be some difference on all these levels, especially so with the length of turns exchanged by the interactants. Interruptive and

self-interruptive talk seem to be more prevalent in the conversational type of talk than in the interview type. This can obviously be confirmed by looking at the transcripts of INT and CON in the appendices in VOL.II of this thesis.

Some analysts have talked about the differences between interview and conversation in terms of temporal variation. Milroy (1980:66) shows that what she terms the **"interview style"** is characterised by :

Content has been removed for copyright reasons

As far as the organisation of topic is concerned, research work has shown that normal conversations give each interactor freedom to choose topics and to range widely in discussions (Denzin 1972:142); interview talk, on the other hand, violates both of these assumptions. The respondent is not free to carry a topic through to completion; the interviewer controls what is discussed. Wolfson (1976:192) thinks that the differences between interviews and conversations lie in (1) the fact that the subject does not have the right to introduce new topics during an interview severely restricts his opportunities for introducing narratives and (2) the subject in interview knows that what he is expected to do is give answers to a series of questions. As a result his narratives when he does tell them, are usually in the form of a summary- short and to the point with little as to the interaction of the participants. Conversational narratives, on the other hand, are usually full of such detail. This detail appears in a number of features- grammatical, phonological and stylistic which are absent from the summary. This latter point is confirmed in this analysis as well.

Pearce (1973:4) reports that Khan and Cannell (1957) conceive of the difference between interviewing and casual conversation as being essentially that of compression. The interviewer's problem is to bring about a sequence of communication in a relatively short time span

consisting as far as possible of items relevant to the topic under investigation. He points out that the broadcast interview is compressed in relation to ordinary conversation, which is also confirmed in the interview part of the informants' talk in this study. This is important to the understanding of the interview structure and function. Another difference lies in the **"relationship between the participants"** (ibid:23) ; whereas two people conversing ordinarily are talking for their own and each other's benefit, the participants in the interview are oriented primarily towards the audience (who may be absent, but active auditors).

Some researchers refer the difference between conversation and interview to the degree of "planning" each of the speech events has been going through (Lang 1984, Zora 1986). Lang (op.cit:19) notices that his interviewee's speech seems to be **"remarkably coherent and well-structured"**. His conclusion is that it is when discourse has been planned to some extent, that **"fluency"** occurs. Zora (op.cit.) refers the higher frequency of pausing in interview than in conversation to the amount of time available to interviewees to plan their utterances before delivering them as opposed to its relative lack to conversants. He also talks about what he terms '**monitoring**'; interviewees in general are always aware that they are being monitored which makes them always conscious of what they say whereas in natural conversation the participants are less conscious of their speech.

1.1.4.2 On the linguistic/structural level

Some of the theoretical comparisons made between interview and conversation in the literature hint at the existence of linguistic variation between the two modes, though these have not been reported in current linguistic research.

Milroy (op.cit.) mentions that the interview style has **"structural simplicity"** in comparison to the other conversation styles whose structure, she thinks, seems to be of a more complex nature. The findings of this research do not seem to support this view as will be seen towards the end of the thesis.

Implicit in Halliday's (1979, 1985b) treatment of the linguistic complexity of spoken discourse in general is the idea that the more informal and spontaneous the conversational discourse is, the more 'syntactically complex' and 'lexically sparse' it becomes (see Section 1.2.3.1 for elaboration on these claims). So if one considers the language of interview on the formal side of the formality continuum (which it normally is), then the language of interview is simple from the structural point of view, which goes in line with Milroy's suggestions stated above. However, these are all speculations based on particular types of interviews which might not be representative of all types of interviewing situations in general.

Evidence to the structural differences between the language of INT and that of CON on the levels of lexis, syntax and information structure is provided by the statistical analyses conducted in this thesis. There will be full discussions and exemplifications of these differences in the analytical chapters, FOUR, FIVE and SIX respectively.

It is not claimed that the foregoing remarks on features of interview as opposed to casual conversation represent a complete list of all that is known about the contrasted styles. Rather, the discussion should be seen as attempting to highlight the most significant features which have been studied by workers in the field, whose findings and results will be cross-examined against the findings of the analysis in the following chapters.

As has already been mentioned, the differences between the language of interview and conversation on the structural levels and from the point of view of linguistic complexity is not clearly distinguishable the way the other social and behavioural differences are. The former need a comprehensive sort of analysis between large corpora of the two types of talk to be checked and established as existing, whereas the latter can easily be depicted by deep observation of the social and cognitive processes involved in the articulation of talk in the forms of speech. The analysis reported in this thesis is wholly devoted for the investigation of the differences on the level of structural complexity in selected data of interview and conversation specifically organised and recorded for the project. The next step in this review will be an overall discussion of what is meant by linguistic complexity in various linguistic research work and in this research. This is needed for pinpointing the various aspects of the linguistic structure investigated by studies on language variation in current research.

1.2 Linguistic complexity

1.2.1 The notion of 'complexity'

In searching for a satisfactory definition of complexity, it must first of all be acknowledged that the definition any researcher adopts will reflect his/her own purposes. Those who study the syntax of discourse, for example, relate it to the type, frequency and distribution of clauses within a particular type of discourse (cf. O'Donnell 1974, Kroll 1977, Beaman 1984). This is normally determined by the frequency of subordinate clauses and their complexity as regards that of coordinate clauses and phrases. It is also determined by the depth and amount of 'embedding' of phrases and clauses, to a great extent. Beaman (op.cit:46) adopts this view although she notes that the presence of dependent clauses in a discourse is insufficient to account for complexity in language. A combined functional and formal

classification of subordinate clauses would reveal the real complexity of discourse.

Quirk et al. (1985:987) point out that subordination is not the only factor that determines the 'complexity' of linguistic units (e.g. the sentence) **"when 'complexity' is understood in a nontechnical sense"**. Other factors such as the degree of modification, the obscurity and compression of vocabulary, nominalisations and so on may have their role in making smaller units such as phrases as difficult or even more difficult to understand than corresponding subordinate clauses. Sometimes the coherence of the sentence as a whole may be difficult to understand.

Halliday (1979) also shares the view that complexity is not just related to the size and amount of clauses and their combinations ('clause complexes' in his terminology). His treatment of the question of linguistic complexity is unique in that he combines three levels of analysis for its definition and characterisation : lexis, syntax and information structure. Therefore, when he talks about the complexity of spoken or of written language he considers all these three issues together.

Others though have treated the problem of complexity in a much narrower sense considering it from a purely lexical level (see, for example, Ure 1971, Stubbs 1986, Farrugia 1988 'though the latter suggests considering discourse subordination and coordination in treating the issue of text type differentiation and complexity in addition to that of lexical density'). Farag (1986) added another dimension of comparability between spoken and written narratives which is that of referentiality.

As far as spoken discourse is concerned there are of course some further criteria which can affect the complexity of the spoken message, such as prosody and disfluencies (Ferguson 1975) in speech production

including filled and unfilled pausing and silences. Prosody is used here in the sense of Crystal and Davy (1969) to cover the whole range of non-segmental effects distinguished as being in some way systematic in speech. These include variations in stress, pitch, loudness, speed, pause, rhythmicality and tension. The issue of pausing in spoken discourse also has an important contributory role as regards the complexity of spoken discourse. Many researchers have studied the relationship between pausing and features of linguistic complexity such as syntactic complexity (See for example, Maclay and Osgood 1959; Boomer 1965; Boomer and Dittman 1962, 1964; Barik 1968, 1979; Hawkins 1971, Beattie 1979, 1980, 1983; Stenström 1986b; Zora 1986).

Prosody is woven into the fabric of spoken discourse, and an exhaustive study of the complexity issue would have to look at the interrelationship between syntactic form and prosodic characteristics. In the present project, prosody, although frequently mentioned, is not fully investigated in its own right, not because it is not directly linked to the issue of linguistic complexity as such but because of limitations of time and resources(3).

An important question, which will not be resolved in this thesis though and will be left for future research (see discussion below), arises at this point and that is : has 'complexity', as defined by the various researchers mentioned above any significance for the listener/reader's processing of text? If it has, what is the role of each type of complexity (lexical, grammatical or informational) in facilitating or inhibiting the process of comprehension on the part of the message decoder ? If we knew the answers to these questions, a number of issues still under debate would be nearer resolution. Among these would be : the desirability of text authenticity for specialist learners of language ; and the importance of "natural" (i.e. structurally more complex) text for early readers.

It must be emphasised that the research presented in this thesis limits itself to the description of textual features of complexity. It makes absolutely no claims as to the psycho-linguistic status of these features for text processing, recognising that what has been called features of complexity may not in fact represent any increase in processing load for the hearer; and conversely that features which would be termed non-complex may in reality represent a significant processing difficulty. This type of complexity has been well researched in the fields of cognitive discourse and text processing (see for example, Haviland and Clark 1974; Schank 1975; Schank and Abelson 1977; de Beaugrande 1980; van Dijk and Kintsch 1983; see also the collection of papers in Freedle 1979 and those in Just and Carpenter 1977).

Until controlled experimentation is carried out, it must remain an open question where complexity, as treated here, has processing significance. Whatever the answer to this question, it must be acknowledged that an explanation is required for the phenomenon that Halliday (op.cit.) commented on. If he is right that an increase in grammatical complexity is offset by a decrease in lexical complexity, then we have to ask : why does grammatical complexity decrease? These questions are pursued at various points in this thesis (Chapters FOUR, FIVE, and SIX).

1.2.2 Research work on linguistic complexity

1.2.2.1 Spoken versus written discourse research

1.2.2.1.1 An overview

Most of the work done on language complexity is to be found in studies of text type differentiation, particularly that of spoken versus written discourse. This type of work, whether focusing on complexity or

other issues, lies clearly within sociolinguistics if it is quantitative. However there are other traditions of research on spoken and written language which are in the main non-grammatical, those which are based on the culture-related matters of orality and literacy (cf. see for example , Gumperz 1984 and also the collection of papers in Tannen 1982, 1984) or those that have an educational strand of research (e.g. Golub 1969). But these are not of concern to the purposes of this study. The concern at present is with the type of work that is linguistically or sociolinguistically orientated of the type concerned with language structure and behaviour in running discourse (see for example O'Donnell 1974; Poole and Field 1976; Goody 1977; Halliday 1979, 1985b; Akinnaso 1981, 1982; Johns-Lewis 1987 and see also Farag 1986 for a review). The relevant findings of these studies will be discussed in the following sections and will be referred to in the analytical chapters below.

Generally speaking the linguistic features which most studies of spoken versus written discourse differentiation have centred around include : lexical density and packing of information (see for example, Ure 1971; Halliday 1979, 1985b; Stubbs 1986), features of English syntax including subordination and coordination (see for example, O'Donnell 1974; Kroll 1977; Beaman 1984; Chafe 1982; Farag 1986; Bilton 1987), language elaboration, verb complexity and personal reference (see for example, Horowitz and Newman 1964; Poole and Field 1976), referential expressions (Farag 1986).

Drawing on previous studies of spoken versus written discourse differentiation, Johns-Lewis (1987:2) neatly summarises what she thinks to be the features of spoken and written discourse that have attracted researchers' attention. She distinguishes the following as characteristics of spoken versus written language :

1. Subjectless verbs
2. Reformulation
3. Fragmentariness
4. Repetition
5. Left dislocation
6. Right dislocation
7. Direct quotation
8. Parenthetical remarks
9. Present tense for past event
10. Deictic usage
11. Fuzzy expressions
12. Zero clause connection

Of course, certain of these features can occur with written discourse as well (ibid:4). For example, subjectless verbs are possible in written prose, reformulation in lectures and repetition in certain types of written discourse.

As has already been mentioned above some researchers confined themselves to specific features for making their comparisons, for example, 'lexical density' (Ure 1971; Stubbs 1986), coordination and subordination (Kroll 1977; Beaman 1984).

All these features represent in one way or another dimensions of the linguistic structure which characterise a particular type of discourse as being simple, complex, fragmentary, incomplete, etc. However, in the state they are found in the literature they are scattered and disorganised. So there seems to be a need for a model of linguistic complexity for text type differentiation. This will be outlined in the few coming sections after some critical statements on the work on spoken versus written discourse variation described above are offered and discussed.

1.2.2.1.2 A critique

While it is true that **"Written and spoken language will never be totally alike"** (cf. Halliday 1985b:44), which shows that there is always some degree of variation in language between its written and spoken types, the results of spoken versus written discourse differentiation reported above draw attention to apparent contradictions. First of all, though most of these studies clearly offer some rich insights into the complexity issue, they have a tendency to offer global descriptions of discourse. Researchers base their conclusions on findings drawn from analysing sets of data which represent specific varieties of discourse, yet claims and generalisations are made to include all spoken or written discourse. Sometimes even the comparisons made between the varieties are based on sets of data which are not strictly compatible, for example, Chafe's (1982) after-dinner conversations are compared with lectures. Most clearly, it is not just the language of the two mediums that is different and variant; **"the devices utilised in the two media for maximum effect can be expected to be different"** (Lakoff 1982:44). Constant use of language means constant change and variation. But it is not only 'use' that causes variation to occur between spoken and written discourse. Lakoff (1982:237) points out that :

Aston University

Content has been removed for copyright reasons

So strategies can also be different. Both situational and contextual factors might be different. Some researchers have made their comparisons without paying attention to the problem of controlling the factors which might influence the linguistic output. Others, though, controlled specific variables such as speaker, context of production and topic, ensuring that the discourse modes they are differentiating are indeed comparable (see for example, Beaman 1984; Farag 1986; Bilton 1987).

It is also clear that the more variables one can control the better results one may get and the more authentic one's own generalisations would become. Research on language variation has shown that topic, participants, planning, spontaneity and formality of the situation are the most important factors to be controlled. Johns-Lewis (1977:9) states that:

Aston University

Content has been removed for copyright reasons

The tendency to compare language variations which are not strictly compatible and comparable on such levels as medium, participants and topics is often misleading. What is more misleading though is to treat the compared sets of language data or text types as 'discrete entities' (cf. Widdowson 1973; Hatim 1981), which in a way means to treat them as separate 'dichotomies' that contain certain shared characteristics to be compared with each other. Some linguists who have worked on language variations, especially on spoken versus written discourse variations, have recognised the danger of this tendency in research and have thus opted for treating those text types along a 'continuum' rather than a dichotomy (e.g. Tannen 1982; 1984; Lakoff 1982; Chafe 1982; Farag 1986). Talking about her two sets of spoken and written data, Lakoff (op.cit.:241) expresses the view that :

Aston University

Content has been removed for copyright reasons

Another point that can be made about the research work reviewed above is that conversation is largely absent from it; with the exception of investigations such as those of (e.g. O'Donnell 1974; Chafe 1982) and those

on communicative narrative discourse (e.g Tannen 1982; Farag 1986). Even when an SES is compared with a written event the two events compared are not, as discussed earlier, strictly speaking comparable as communicative event types. Chafe (op.cit.) is aware of this problem admitting that :

Aston University

Content has been removed for copyright reasons

The lack of research on conversational data from studies of spoken versus written language is precisely because researchers have been concerned to analyse comparable event types. It is for this reason that oral and written narratives have been favoured : both event types are authentic event types in their own right, both occur naturally in spoken and in written form. The difficulty, if one selects conversation, is what to compare it to. There is no written counterpart to conversation. Nor is it immediately obvious what kinds of conversational events can be compared. Same participants, different settings ? Same topic, different participants ? (etc.) The research reported in this thesis opts for same subjects, same setting, but different modes of interaction. This will be elaborated on in CHAPTER TWO below.

1.2.2.2 Text-type differentiation within spoken or written discourse

Variationists' studies which deal with features of discourse within one medium (i.e. spoken versus spoken or written versus written discourse) and are related to the problem of linguistic complexity are relatively few compared with those between spoken versus written discourse discussed above. Most of the work done has been carried out within the general field of sociolinguistics dealing mainly with linguistic variables in terms of

phonology, lexis or syntax within and between speech communities (cf. Labov 1972a, 1972b, the collection of papers in Trudgill 1978, Milroy 1985). Some of these studies are confined to variations within one particular variety of language, either spoken or written(4).

Studies featuring linguistic differentiation between one type of spoken discourse and another are particularly interesting though they are relatively rare, and the present available studies are either confined to investigating narrative discourse (cf. Labov 1972b, Chafe 1980, the collection of papers in Tannen 1982, Rimmer 1984, 1988) or are more related to variations within phonology such as those studies on the prosody of discourse (cf. Brazil 1975, 1978, 1981; the collection of papers in Johns-Lewis 1986), studies on pausing in spoken discourse (cf. Goldman-Eisler 1958; Maclay and Osgood 1959, 1972; Beattie 1983; Stenström 1986b; Zora 1986; Johns-Lewis 1987). Some of these studies constitute an important source for the present state of knowledge of spoken discourse variations.

1.2.3 Towards a model for the analysis of linguistic complexity of Speech Exchange Systems

1.2.3.1 Two contradictory views on the complexity of spoken discourse

In general terms, the major outcome of the studies reviewed above as far as linguistic complexity of spoken discourse(5) is concerned, is a bulk of contradictory findings and remarks especially about which of the two discourse modes, spoken or written, is more lexically or syntactically complex than the other, which is per se an important issue to be drawn upon in this research in relation to SES. The general findings of the research work on language variations yield in the main two contradictory but interrelated points of view about the complexity of spoken discourse types. While Halliday's theoretical suggestions (1979, 1985b) (see discussion

below), supported by the findings of some researchers (e.g. Poole and Field 1976, Beaman 1984) point to the claim that spoken discourse is more syntactically complex than written discourse as it has relatively more complex sentences with simple words (See CHAPTERS FOUR and FIVE below), others have shown written discourse to be more syntactically and lexically complex than spoken discourse (e.g. O'Donnell 1974; Kroll 1977; Brown and Miller 1980; Chafe 1982; Brown and Yule 1983).

Examining such indices as structural elaboration, nesting and embeddedness and uncommon clause usage, Poole and Field (1976) point out that their findings show that spoken discourse (oral interview) is more structurally complex than written discourse (group testing situation). They state that :

Content has been removed for copyright reasons

They refer these findings to the less time for what they call the "sifting" of verbal expression of ideas and for the "little or no opportunity for corrective self-feedback because of the limitations of short-term memory spans" (ibid:309).

In his theoretically-based examination of the complexity of spoken discourse in relation to written discourse Halliday (1979, 1985b) reports similar findings to those of Poole and Field cited above. His main claim is that the complexity of any type of discourse is linked not just to the number, size and nature of clauses and clause complexes but to other, lexical and semantic, factors which are inherent within the clause. He views both written and spoken discourse as complex each in its own right, but the type of complexity with which each is associated is quite different :

Aston University

Content has been removed for copyright reasons

He considers the complexity of written language to be **"lexical"** whereas that of spoken language to be **"grammatical"**. He states that of all types of spoken discourse the more informal and spontaneous the spoken language is the more 'grammatically' complex it becomes. This is obviously confirmed in the following quote :

Aston University

Content has been removed for copyright reasons

The quote above is very rich in its implications especially for the purposes of this study. It, first of all, refutes many of the other researchers' findings that spoken discourse is 'syntactically simple' (see discussion below) as it shows greater reliance on coordinate structures (cf. Crystal and Davy 1969; Kroll 1977; Chafe 1982; Brown and Yule 1983). Secondly, as it has already been hinted above, it confirms clearly that it is not just syntax that is involved in the complexity of discourse but also other factors such as lexical content and organisation :

Aston University

Content has been removed for copyright reasons

Thirdly, Halliday's views about the relationship between the issues discussed above and the controversial issue of packaging of information and its structure in the clause is also related to linguistic complexity in one way or another. Any increase in lexical density necessarily means a higher informational load (Halliday: ibid) and vice versa. It seems also important to study the effect of increase in syntactic complexity on the information status and on the encoding and decoding of this information (see Halliday

1967, 1985a for his views about the information structure in discourse and discussion of these views in CHAPTER SIX below).

The results of Beaman's (1984) analysis on variation between spoken and written English narrative discourse are also supportive of the argument put forward by Halliday (op.cit.) especially the part of it which is related to the differences in the type of complexities involved in the two modalities. Her investigation, which centres around subordination and coordination, shows that :

Aston University

Content has been removed for copyright reasons

She refers these results to the increased **planning time"** (Ochs:1979) allowed to the writer and to the lack of **"visibility"** (Lakoff:1979) and absence of extralinguistic factors between writer and audience. The differences in complexity between the spoken and written modalities which previous studies have found, Beaman believes to be due to the differences in the formality and purpose or register of the discourse rather than true differences between spoken and written language.

Looking at the second point of view which mainly maintains through evidence obtained from linguistic analysis that spoken discourse is more structurally complex than written, one can easily notice a diversity of opinions about this structural 'complexity', and about the methods and measures used for its specification and interpretation (see Farag 1986:29 for a detailed critique). O'Donnell (op.cit.) bases his claims about the syntactic complexity of written discourse in relation to spoken discourse on counting in his (single) informant's spoken output a hundred T-Units (see Section 3.1.3.2 for discussion of this grammatical unit of analysis) as against another hundred T-Units of written newspaper columns produced

by the same informant. Kroll, (op.cit.), Chafe (op.cit.), in addition to the theoretically based views of Brown and Yule (op.cit.), when looking at differences between types of spoken and written discourse, all provide evidence to the effect of the presence of more syntactic complexity of the latter than the former. All of them look at complexity from the point of view of the increase in subordinated structures and decrease in coordinated structures within discourse.

The following quote from Brown and Miller (1980:11) summarises the major trends of the proponents of this view and shows indices of linguistic comparison other than the mere counting of grammatical units :

> "Written language has a greater range of vocabulary, a more complex syntax, except perhaps in hastily composed letters, none of the stops, starts and repetitions of spoken language."

While all these studies confirm the complexity of one type of discourse mode rather than the other, some researchers found no differences between the two modes of discourse (cf. Blankenship 1962). Farag (1986) thinks that it is wrong to associate complexity and simplicity with writing and speech respectively, since individuals differ considerably in their strategies. Although she found differences between speech and writing, she tends to refer them to individual style (i.e. style constant for speech or for writing). Similar results concerning style constancy were earlier reported by Blankenship (op.cit.).

One research which is mainly concerned with investigating linguistic variation between two types of spoken discourse namely narrative and ordinary conversation, and hence merits special attention in this study is that of Labov (1972b) in which he finds that narrative discourse shows a more complex structure than ordinary conversation. He states that :

Aston University

Content has been removed for copyright reasons

This shows that although Labov takes embedding and clause structure into account when considering the structural complexity of spoken discourse, there are other issues like the structure of the verb phrase, the modals, and so on which have to be considered as well.

1.2.3.2 Insights and implications

To sum up the discussion so far, it seems that some linguists have talked about the complexity of SES to the exclusion of everything that cannot be stated in terms of the number and size of phrases, clauses and clause complexes (cf. Halliday 1979, Beaman 1984). As has already been explained above, for Halliday complexity is a complex of grammatical and lexical information. However, there is a trade-off between complexity at the level of syntax, and complexity at the level of lexis. Halliday stresses that lexical items are of low occurrence in spoken discourse (including conversational discourse), and that it is only complex at the grammatical level but not at the lexical one. He describes this complexity as **"intricacy"** (1985:62), **"the intricacy with which the information is organised"**. Conversely, written discourse can be expected to be more complex than spoken at the level of lexis, but possibly less complex at the level of syntax. This, however, is an issue of contention, as will be implied by the results of the analysis of types of SES conducted in this thesis (see especially CHAPTERS FOUR and FIVE).

Intuitively, it would seem that complexity in SES, as outlined above, is actually increased by features which are particularly characteristic of SES. Here we are thinking of such features as ellipsis within grammatical and lexical constituents, inversion, left and right dislocation, structural revision, interruption, overlapping speech, topics and subtopics etc. Even such non linguistic features as voice quality, gaze direction, head nods can have their role in increasing or otherwise decreasing the complexity of a linguistic message. In addition to that, the status of information and its packaging in certain discourse types rather than others or in certain parts of discourse rather than others seems in the opinion of the present researcher to have its role in the specification of complexity as a whole. These are all speculations which need to be verified in subsequent research. The concern here is to investigate the role played by lexis and information structure in manifesting the complexity of SES, in addition to that of syntactic complexity mentioned by Halliday (op.cit.) and others who have done research on these lines (cf. Poole and Field, 1967, Beaman 1984, Farag 1986, Bilton 1988).

1.2.3.3 A Model for linguistic complexity

The criteria set by Halliday for the description and characterisation of the linguistic complexity of spoken and written discourse along with his substantial work on information structure and Given/New distinction in spoken discourse can be used as a point of departure for the analysis of SES in this thesis. Although some of the claims he makes, especially regarding the issue of lexical density in SES (see CHAPTER FOUR) are not substantiated by the findings of this thesis, they are still plausible as criteria for the analysis of discourse. One of the main reasons for this lies in the fact that most of Halliday's claims about the complexity of discourse discussed above, can be taken single-sidedly to be applied to one particular discourse mode rather than in contrast with another. This seems more suitable with the data used here in this project as it is not an examination

of speech in comparison to writing but actually one variety of speech against another variety of speech, so to speak. All research work on language variation (other than Halliday's cited above) yields results of complexity of one discourse mode in relation to the other.

Accordingly, and taking insights from the discussion above, a model for the analysis of linguistic complexity is proposed which covers the three major areas of linguistic structure of discourse : lexis, syntax and information structure (see Fig. 1.1 below). The analysis is carried out on the lower units of discourse which the data of INT and CON have been segmented into, moving up to the Major units then to the total occurrences of the specific categories under investigation within the whole body of data. The methodology and the criteria of data segmentation into units will be discussed in CHAPTER THREE below.

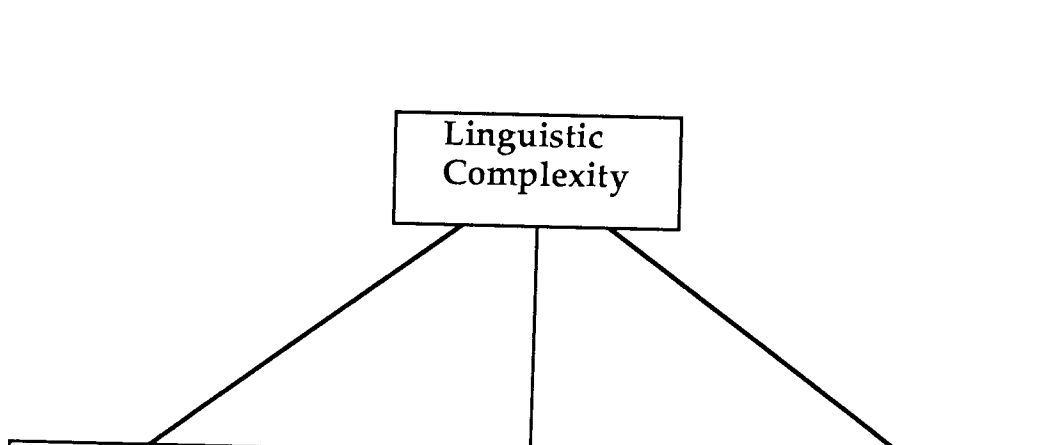


Figure 1.1 A model for the Analysis of Linguistic Complexity in Interview & Conversation

Besides the fact that these three criteria represent the 'gist' of any linguistic investigation of the structure of discourse, they can be used, as the research work reviewed above has shown, as levels of comparison between text-type varieties of spoken versus written discourse mediums

and between varieties of one medium. This is why they seem to fit the type of analysis presented here. Secondly, all of these levels can readily be subjected to computational techniques of analysis and statistical tests of frequencies and significance using big corpora of data for the analysis (see Section 1.3.2 below). Following is a brief characterisation of each of these three levels in relation to the analysis in the present thesis. More elaboration on each will be provided in the analysis chapters of the thesis.

1.2.3.3.1 On the level of Lexis

The issue of lexical structure in SES, and indeed in any piece of spoken or written discourse, is manifested in the packaging and density of 'lexical' items as regards 'grammatical' items within discourse (see Stubbs 1986 and CHAPTER FOUR below for the distinction between 'lexical' and 'grammatical' items). As shown above, a great number of researchers refer the complexity of written discourse to its heavy reliance on the use of more lexical than grammatical items (cf. Ure 1971, Halliday 1979, 1985b). Indeed in certain varieties of written texts (e.g. newspaper editorials (cf. Farrugia 1988), the ideas are expressed in highly specialised lexical items, with the grammatical items being kept to the minimum, as in the following two texts taken from Farrugia (op.cit.:73) (newspaper editorial) and Prince (1981:234) (recipe instructions), respectively :

1.1 (Newspaper Editorial)(6)

Football Crazy

A RIGHT bunch of crackpots are running European football

Their latest idea on "foreign" players means English clubs would have to dump many of their Scottish, Irish and Welsh stars. Good job Mike England isn't manager of Wales any more.

That would have had the Europrats in a straitjacket.

1.2 (Recipe Instruction)(7)

> Cover the ears and the tail with aluminium foil. Place the pig in a pan uncovered in the oven for 15 minutes. Reduce the heat to 325° and roast until tender, allowing 30 minutes to the pound. Baste every 15 minutes with....

In both of the examples above the ratio of Lexical to Grammatical items is (57% : 43%) (See CHAPTER FOUR below). Within SES, however, the general belief is that lexical density is much lower than that of written discourse, thus making texts of the spoken type less 'complex' and less 'informationally loaded' than written texts (cf. Halliday op.cit.). Ure (1971:445) provides an example of a type of spoken discourse representing one of her language-in-action texts which has a very low lexical density judged by the low occurrence of 'lexical' as opposed to 'grammatical' items (see Section 4.2.2.2 for definitional comments). The text reads as follows :

1.3 (from Ure 1971:443)

Aston University

Content has been removed for copyright reasons

But can this extract be said to represent types of spoken discourse in general ? The answer is definitely no. The following extract is taken from one of the Interviews recorded for the present analysis. Notice the higher occurrence of 'lexical' over 'grammatical items' which indicates that the extract is highly lexically dense in comparison to Ure's text cited above.

1.4 (INT III)

RH. 83.1 the the way the the::.. residences are organised as well..
2 makes fo::r a very.. e:::m.. oppressive atmosphere I think..
84. e:::m... let's expand on that..
85.1 the..
2 I think it's very difficult for s.. for somebody not to conform..
3 to::.. what they believe is the general.. view of things
86.1 and e:::m.. walking around in a flat of nine people..
2 wearing CND badges..
3 when when you.. believe everybody else is.. anti CND
4 or not.. supporting CND..
5 is a difficult thing to do I think
6 when you've just.. left home
7 and.. e:::m.. you're.. you're trying to fit in..
8 trying to.. form groups of friends..

This type of conversation shows that there are certainly many types of SES that are as lexically dense and hence as much informationally loaded as any written text. Gregory (1976:38) mentions that **"We can recognize people who 'talk like a book' as well as people who 'write like they talk'"**, which seems to suggest that there are variations within SES which, like those of written discourse variations, fall on a scale of complexity regarding the employment of lexical items. This complexity can be manifested both in the text as a whole and at the lower levels in the smaller units of discourse structure as will be seen from the analysis in CHAPTER FOUR below.

It is not only the ratio of lexical to grammatical items that determines the lexical complexity or otherwise of a text and its variability in different types of discourse but also the information load each of the lower units of discourse has according to its function in running speech. The hypothesis to be tested in this concern is whether certain units of discourse with specific functions can be more lexically dense (hence more informationally loaded) than other units. This hypothesis needs to be tested against the

lower units of the data of INT and CON in this thesis as well. An investigation of this phenomenon might shed some light on the issue of complexity of SES on the lexical level. The analysis in CHAPTER FOUR is directed mainly towards providing evidence (with statistical frequencies and exemplification) for the presence of such variation between the two types of SES under investigation (INT and CON).

1.2.3.3.2 On the level of Syntax

Related to the notion of lexical structure in discourse is that of syntactic structure. This is normally manifested in the use of clauses and multi-clausal combinations formed by such syntactic processes as 'coordination', 'subordination', 'apposition', 'ellipsis' and other characteristics of the English grammatical system mentioned in section 1.2.2 above. This issue represents a major field of inquiry in the study of language structure since the advent of linguistics as the science of language investigation and has been of concern more recently to studies on language variation (see Section 1.2.2.1 above). Within SES, this issue is, too, of great importance since, as will be seen in CHAPTER FIVE below (see also CHAPTER THREE, Section 3.1.2 for discussion and examples), the syntax of any conversational activity, whether formal or informal, planned or unplanned, is not as tidy and organised as, for example, written discourse or other types of spoken discourse such as monologic discourse. However, it is intuitively suggested that the degree of syntactic 'untidiness' is not just variable according to the type of SES at hand but also to the degree of its formality. Apparently, it seems that the more informal the discourse is the more syntactically untidy it becomes. Certain SES types (e.g. broadcast interviews, especially the types which show characteristics of 'monologic' rather than 'dialogic' discourse) display highly organised syntactic tidiness comparable sometimes to that of written discourse. This does not mean, however, that the syntax of informal conversation is chaotic or disorganised as its transcript might

show it to be. It is, on the contrary, highly organised but this organisation is configurated in such a way that it is rather shown to be **'intricate'** (Halliday 1985b:62). Again, this intricacy is variant from one type of conversation to another and from one speech situation to another even among the same speakers. This is precisely what makes its study rather interesting and beneficial.

Some of the features of the language of SES that have some bearing on its syntactic complexity either because of their high level of occurrence or because of the manner they behave within running discourse need to be examined along with the other features of English syntax described above. These features include types of elliptical units, discourse markers and what Quirk et al. (1985) refer to as 'Reaction Signals'. It is realised here that these features are not directly related to syntax (cf. Stenström 1984b). However, they seem to characterise speech styles giving them a distinctive characteristic structure which is not found in other types of spoken or written discourse; hence their inclusion in the analysis of syntactic features of SES adopted in this thesis. Details of the syntactic analysis and examples on each of the features examined in the data will be presented in CHAPTER FIVE below.

1.2.3.3.3 On the level of information structure

The third perspective in the study of the internal structure of discourse (including SES types), one which has relevance to both lexical and syntactic organisation, is that of the structure of information organisation within the units of discourse (cf. Halliday 1967) and the linguistic representation of these units. Lexical and syntactic units of discourse combine to form 'units' or 'chunks' of information, the size and nature of which are determined by the speaker who has the option of distributing the quanta of information he/she wishes to convey to his addressee(s) (ibid). Such units have been extensively researched by scholars of various

disciplines. Different methodologies and criteria for the unitisation and description of information within sequential discourse have been offered. Some insist that these units are realised intonationally as 'tone groups' that consist of focused elements representing what the speaker assumes would be new information to the listener and unfocussed elements representing the old information or what the speaker thinks to be already known to the listener (cf. Halliday ibid and in 1985a). Others attach other prosodic features to the realisation of these units, such as pitch, stress and pausing as in Crystal's (1975) 'tone unit'. Grimes (1975) uses the term **'information blocks'** to describe his units of discourse structure while Chafe (1980) uses the **'Idea Unit'** as a unit of information that can be realised semantically, syntactically and intonationally. Chafe uses this unit to study, on the psycholinguistic level, chunks of information as **'focuses of consciousness'** and **'centres of interest'** (ibid). Cognitivists use what they refer to as **'propositional units'** and other 'propositional-based units' (cf. van Dijk 1980; van Dijk and Kintsch 1983; Rumelhart 1977; Schank 1975). Others though, have used just syntactic and semantic criteria to distinguish these units (e.g. Kroll's (1977) **'idea unit'**; also used by Farag (1986), and have studied them in relation to their internal linguistic structures rather than the type and nature of information produced within them.

Mainly, definitions and characterisations of these units in the literature depend on the discipline within which the study of information is being dealt with and the purpose of the analysis per se (see CHAPTER SIX below). For the present study a unit for the analysis of SES is proposed depending on a combination of criteria which are thought by the present researcher to be applicable to most (if not all) types of spoken discourse whether of the formal or the extreme informal type as in casual chat. This unit, which is going to be termed 'FREE STANDING SYNTACTIC UNIT (FSSU), represents the basis for the analysis of the data of INT and CON. There will be a detailed account of this and other types of units in

CHAPTERS THREE and SIX below. For the time being, it is important to comment albeit generally and briefly on the approaches that have dealt with the issue of information structure in text and discourse and the way(s) they have handled this issue from the point of view of information description, distribution and organisation within the lower and/or higher units of discourse.

There are generally three types of approaches that can be recognised in the literature concerning the analysis of information structure in discourse. They can be grouped under three major headings : 'TEXTUAL', 'COGNITIVE' and 'DISCOURSAL' respectively. Each of these will be outlined in the following three subsections.

1.2.3.3.3.1 The textually-based approaches

These are represented by the work of the Prague School of *linguistics* (e.g. Mathesius 1941, 1942; Dahl 1974; Danes 1974; Firbas 1964, 1974, 1975; Vachek 1975) and the later work of other linguists related to it (cf. the work of systemic linguists such as Halliday 1961, 1967, 1979, 1985a, 1985b; Halliday and Hasan 1976; Hasan 1981, 1985; Berry 1975, 1981; Fawcett 1980; Martin 1983, 1984). Also included within these approaches is the various research work that has been done under the rubric **'TEXT LINGUISTICS'**, mainly represented by the work of de Beaugrande (1978, 1979, 1980, 1981, 1984); van Dijk (1977); Petofi (1974, 1979) and see also de Beaugrande and Dressler (1981) Al-Jabr (1987) and Al-Jubouri (1988) for description and more references.

The main preoccupation within the textual approaches is with the characteristic structure of the text and its main textual and functional features, such as 'thematic structure (i.e. Theme/Rheme), and 'informational structure' (i.e. Given/New) (cf. Danes 1974; Halliday 1985a) (see also discussion below), 'cohesion' and some major issues in the

linguistic structure of discourse which some approaches attach to cohesion, such as 'reference', 'substitution', 'ellipsis', and so on (cf. Halliday and Hasan 1976; Gutwinski 1976). de Beaugrande and Dressler's (op.cit.) 'Procedural Approach' to the study of texts focuses on such elements of the text as connectivity and relations (i.e. cohesion and coherence), the attitudes of producers (i.e. intentionality) and receivers (i.e. acceptability), the communicative setting (i.e. situationality), information structure (i.e. informativity), and finally the ways in which the production and perception of a text depending on the participants' knowledge of other texts (i.e. intertextuality) (ibid:182).

The emphasis within the linguistic school of Prague is on a set of three issues related to the structuring of information in the sentence/utterance, which they call **'FUNCTIONAL SENTENCE PERSPECTIVE' (FSP)**. These issues are GIVEN/NEW, THEME/RHEME and COMMUNICATIVE DYNAMISM respectively. The THEME is that part of the utterance **"that is being spoken about in the sentence"** (Mathesius 1939, in Danes 1974:106) and the RHEME is **"what the speaker says about this theme"** (ibid). So in the following example, the pronoun **[I]** in the first unit (see CHAPTER THREE for the definition of 'unit') is the Theme and the rest of the unit is the Rheme. Studies on this distinction are covered under the rubric 'THEMATIC STRUCTURE' or sometimes 'THEMATIC PROGRESSION' (cf. Danes op.cit.; Nwogu 1989; Nwogu and Bloor 1989).

1.5 (CON VII)

GM. 86. 1 because.. e::m **[I]** was living with.. **a girl** called **Joe**
2 and **a lad** called **Chris**..
3 and **{Joe}** got **/Bronchitis/**..

COMMUNICATIVE DYNAMISM (CD) as developed by Firbas (1964:270) means **"the extent to which the sentence element contributes to the development of the communication"**. Firbas (ibid) discusses this issue in

relation to 'Given' and 'New' information within sentences. His conclusion is that certain elements within the sentence carry more CD than others. Consequently, the suggestion is that New elements within the information structure seem to have more CD than Given information. CD, in this sense, seems to be similar to de Beaugrande and Dressler's notion of 'INFORMATIVITY' (de Beaugrande and Dressler 1981:138) discussed in CHAPTER SIX (Section 6.1.1) below.

The Given/New distinction within FSP mentioned above, is the most relevant of all to the purposes of this study. Therefore, it is discussed in some detail in relation to the analysis of information structure in CHAPTER SIX. The other two issues (i.e. THEMATIC STRUCTURE and COMMUNICATIVE DYNAMISM) will not be elaborated on further as they will not be included within the analysis of information structure of INT and CON.

1.2.3.3.3.2 The cognitively-based approaches

The approaches that can be grouped under the 'cognitive' categories are wider in range than those discussed above and include issues related to knowledge representation, discourse processing, cognitive interpretation and psycholinguistic analyses. Although these differ in their interpretive and methodological approaches, they tend to agree that the production and comprehension of information are part and parcel of a human's mental processing behaviour and capacities (cf. Minsky 1975; Schank and Abelson 1975; Rumelhart 1975, 1977, 1980, 1984; Rumelhart and Norman 1987; Rumelhart and Ortony 1977; de Beaugrande 1980, 1981, 1984; de Beaugrande and Dressler 1981; van Dijk 1977, 1984, van Dijk and Kintsch 1983, among many others). From the point of view of the way they handle the issue of information analysis and comprehension, two main lines of these approaches can be distinguished. The first line is represented by the work of psycholinguists whose main interest is to account for how

communicators convey, interpret and comprehend information from both a linguistic and psychological perspective (see, for example, the work of Chafe 1970, 1972, 1974, 1976, 1977, 1980, 1982, 1984; Clark 1975, 1978; Clark and Clark 1977; Clark and Haviland 1974, 1977; Haviland and Clark 1974). Such studies normally deal with the structure of information within the lower units of discourse (e.g. 'Idea Units' and 'Intonational Sentences' in Chafe's work and 'Sentences' in Clark's group).

The other line within the psychological school of thought whose scholars are preoccupied with finding models of comprehension under the rubrics 'Text/ Discourse Processing' and 'Artificial Intelligence' is more concerned with what they refer to as **'PROPOSITIONS'** (cf. van Dijk 1977; van Dijk and Kintsch 1983, **'SCHEMATA'** (cf. Rumelhart 1977; Norman and Rumelhart 1975); **'FRAMES'** (cf. Minsky 1975); 'Scripts' (cf. Schank 1975; Abelson 1976; Schank and Abelson 1977); **'SCENARIOS'** (Sanford and Garrod (1981). It should be clear that these notions are not synonymous (cf. Nwogu 1989:106), but they are related in the sense that they all function to account for holistic or higher-order knowledge representation (ibid). It is recognised here that these studies are of great importance in the field of information processing and have yielded rich insights for the study of informational structure in linguistic and discoursal terms. However, since the analysis undertaken in this thesis is not based on issues related to information processing and psychology in general, they are not going to be elaborated on except in as much as they contribute to the main arguments within the subsequent chapters.

1.2.3.3.3.3 The discoursally-based approaches

The third and final collection of approaches to the study of information within discourse is represented by the work of those scholars who discuss the structure of discourse in terms of the relations that exist between elements of 'a text' and the 'discourse model' which proponents of this

school hypothesise to be created by speakers and hearers in communicative situations (cf. Prince 1981:235). Also included under this rubric is the work of those scholars who are interested in developing a theory of 'reference' through 'textual relations' and 'discoursal relations' (cf. e.g. Stenning 1978; Hawkins 1978; Webber 1979, 1981; Prince 1981). Within another vein of 'discourse analysis' research in the structure and organisation of information, the work of Sinclair and Coulthard (1975) and their hierarchical approach to classroom interaction is outstanding. This will be discussed within the salient approaches that have dealt with spoken discourse analysis proper in the following few sections.

The analysis undertaken in CHAPTER SIX concerning the analysis of information structure will be mainly based on a modified version of Prince's (1981) taxonomy to information status. It is believed that, in spite of some of its shortcomings (see Section 6.3.3 below), the taxonomy is rich and comprehensive and has taken into consideration most of *the issues* and principles which relate to the status of information in discourse. It offers the discourse analyst a linguistic model of information structure. Besides, it is seen as the best convenient taxonomy for the analysis of SES whose syntax and overall organisation are not very easy to describe and analyse using the other traditional taxonomies.

In the next few sections, the main approaches available in the literature to the analysis and description of spoken discourse (especially SES) will be briefly commented on and assessed. This is done for the purpose of obtaining more insights into the analysis of such speech systems both on the linguistic and discoursal levels. It will also help us in finding a suitable approach whose techniques and methodologies can be exploited for applying the Model of linguistic complexity analysis proposed above.

1.3 Approaches to the analysis of Speech Exchange Systems

The approaches that have dealt with the analysis of spoken discourse in general and that of SES in particular are too many to be described in a short review of the present type. However, it seems useful at this stage to give a brief outline of the most salient approaches in the literature so as to be able to pick the one which will be the most suitable for the purpose of the analysis of the type of data that has been recorded for the present study.

For the purposes of this study, the approaches that have dealt with the description and analysis of spoken interactional discourse can be grouped under the very general rubrics of **'QUALITATIVE'** and **'QUANTITATIVE'** approaches respectively. Although it is realised here that this is a rather crude way of classification since some of the approaches can fall under either of them equally well and overlap is really likely, the classification will help in pinpointing what is thought by the present researcher to be the right approach for handling the data of INT and CON in its present format. It is also realised that some of these approaches have origins other than linguistic, e.g. sociology or philosophy, but the main emphasis will be on those approaches that have direct relevance to the linguistic tradition either through the general veins of socio-linguistic and psycholinguistic research or through their direct contact with the analysis of language within their own disciplines.

1.3.1 Qualitative approaches

Three main approaches can be distinguished within the qualitative type. These are represented by : The **sociologically** oriented approaches of the ethnomethodologists and ethnographers (see for e.g. Sacks 1970; Schegloff 1968; Turner 1974; Gumperz and Hymes 1972); the **philosophically** oriented work on speech acts and conversational implicature (e.g. Austin 1962; Searle 1969, 1972, 1975; Grice 1957 and 1975;

Kempson 1977) and the **linguistically** focused work done under the general rubrics of 'discourse analysis' and 'text linguistics'. (See for e.g. the work of the 'Birmingham English Language Research' group represented by such research work as Sinclair and Coulthard (1975), Coulthard (1975,1977), Coulthard and Montgomery (1981) and also the work of what has come to be known as 'European Text Linguistic Research' represented by such works as de Beaugrande (1980), van Dijk (1985) among many others). This classification is adopted by Burton (1981a) in her analysis of conversational discourse.

Levinson (1983) recognises two main types of approaches corresponding to two of Burton's categories above, namely the **sociologically** and the **linguistically oriented categories**, which he calls the **'conversation analysis'** (CA) and the **'discourse analysis'** (DA) approaches respectively. Burton's **philosophically** oriented approaches i.e. the **Speech Act Theory** and **Conversational Implicature**, Levinson discusses under two separate chapters. This classification seems to be more practical for the purposes here and and less confusing in that it offers less overlap of the approaches recognised.

Levinson distinguishes two basic categories within the DA approaches: the text grammarians and the speech act (or interactional) theorists (see discussion above). The former view discourses as sentences strung together in much the same way as clauses within sentences can be conjoined with connectives of various kinds (ibid:288). This Levinson dismisses as inappropriate for the analysis of conversation since **"..the links between speakers cannot be paraphrased as sentential connectives"** (ibid). Speech act theorists, on the other hand, view discourse not in terms of linguistic expressions but in terms of 'speech acts' and 'interactional moves' that are produced within the utterances of speech (ibid) (see also Austin 1962; Searle 1965, 1969, 1975 for details of this approach). Space does not allow a detailed discussion of the approaches mentioned above. But a

brief critical assessment of these approaches in relation to the purposes of this study seems to be necessary at this point. This will be done in the following few sections.

1.3.1.1 The philosophically oriented approaches

The Speech Act Theory (SAT) has been drawn upon by many analysts especially in the specification of speech acts and the rules that tie them together, in their analyses of different speech styles (see for example Labov 1972a; Labov and Fanshel 1977; Widdowson 1975; Sinclair and Coulthard 1975; Coulthard 1975, 1977; Franck 1979; also see the collection of papers in Coulthard and Montgomery 1981). However, in spite of the popularity the theory has gained in its application to spoken discourse, it is doubtful in its ability to handle conversational discourse in general. Many criticisms have been put forward in the literature against the approach as inadequate and inapplicable to certain speech styles such as natural conversation (Levinson 1980:20) and for its lack of a principled basis and unsystematicity in its dependence on felicity conditions (Levinson 1983:240), for its inability to provide a basis for classification of discourse (Johns-Lewis 1986:207); for missing how speaking is organised as a social activity (Streeck 1980:147); for its incompatibility with the CA approach and its insufficiency as a theory of interaction (Franck 1981:226), and for the lack of match between levels based on linguistic expressions and actual purposes in human interaction and for its ignoring both naturally occurring data, and also connected sequences of speech acts (Stubbs 1983:148).

The **'CONVERSATIONAL IMPLICATURE'** approach as developed by Grice (1975), though essentially a theory of how people talk and exchange talk, has also been found inapplicable to analyse large bodies of conversational discourse. The theory is based on Grice's work on 'communication' and 'meaning' presented earlier (1957). This work has

been developed into a more elaborative theory normally referred to as the "COOPERATIVE PRINCIPLE" (1975), in which Grice puts forward the view that communication is governed by a number of standards and maxims. These maxims are those of **"QUANTITY, QUALITY, RELATION and MANNER"**(8). The theory has given very good insights into the whole idea of how people interact and put their language into use (cf. Levinson 1983:101) and how they co-operate when communicating ideas with each other. This is clear in Grice's statement quoted below :

Aston University

Content has been removed for copyright reasons

Levinson (op.cit.) believes that Grice's "COOPERATIVE PRINCIPLE" (see Grice op.cit.:43) seems to govern aspects of non-linguistic behaviour to describe **"rational means for conducting cooperative exchanges"** (Levinson op.cit.:103). However, the reason for the linguistic interest in the maxims is that they generate inferences beyond the semantic content of the sentences uttered.

The theory, is believed to be too broad for a linguistic theory (Levinson op.cit.:118). He states that if any use is to be made of it, much has to be done to tighten up the concepts employed and to work out exactly how they apply to particular cases. Sperber and Wilson (1986:32) think that, though influential in its descriptive power of the process of communication as a whole, the theory seems to be weak in its explanatory power. These authors have suggested an alternative approach, which is essentially an extension of one of the four maxims of the original theory, which is that of **"RELEVANCE"**. In Grice's (1975) theory this maxim reads as follows **"BE relevant"**. This in Sperber and Wilson's (op.cit.:46) belief is

a cognitive property **"which makes information worth processing for a human being"**. Greater relevance is achieved in communication when communicators interconnect new and old information in such a way that further new information can be derived (ibid:48).

1.3.1.2 The sociologically oriented approaches

The sociologically oriented approaches represented mainly by the work of the 'ethnomethodologists', namely Garfinkel (1967, 1972); Goffman (1981); Sacks (1972, Sacks et al. 1974, Sacks and Schegloff (1979), Schegloff (1972, 1979) Jefferson (1972, 1973), have looked at conversation from the point of view of its being able to accommodate a wide range of situations within which persons or groups of persons are operating. The attention of the group has focused on conversational materials because :

Aston University

Content has been removed for copyright reasons

This justifies the name attached to the theory as the 'Conversation Analysis' theory. Garfinkel's preoccupation with the language of talk comes from his concern with the interpretive processes underlying communicative acts whether verbal or nonverbal.

The work of the conversation analysts concentrates on many aspects of conversational properties which any analyst (including the present one) finds helpful and illuminating, such as **"turn-taking"** (Sacks et al. op.cit.), **"adjacency pairs"** (ibid), **"conversational sequencing"** (Schegloff 1968),**"openings and closings"** (Schegloff and Sacks 1973), **"side sequences"** (Jefferson 1972), **"indexical expressions"** (Garfinkel and Sacks 1970:348).

Space does not allow a detailed description and discussion of the interesting work of the ethnomethodologists but the reader is referred to the literature specified above and to Levinson's (1983) detailed critical assessment of their work.

The ethnomethodological approach has also come under direct criticism for many reasons, most important of which are related to the notion of 'adjacency pairs', which, in Levinson's (op.cit.) view, is too strong a requirement for conversational exchanges. Coulthard (1975:91) criticises the approach, as he thinks **"the descriptive categories are in the main not defined"** and that **"there is no attempt to relate form to function"**. Garfinkel's approach has been criticised for its "unsystematicity in its intuitive way of collecting and handling data" (Pearce 1973:77), for its phenomenological bias and for its failure to indicate clearly the nature of the documentary method and the source of meanings and definitions (Coleman 1958:128).

As it stands, the ethnomethodological approach has certainly yielded, together with the related substantial work of Sacks and his school of thought, a number of insights which have added to the understanding of conversational organisation, sequencing and consequently conversational interpretation. For example, the emphasis on analysing recorded transcripts of naturally occurring conversations is highly useful for linguistic description. Besides, the theory has added to the understanding and treatment of indexical expressions which constitute a major part of linguistic expressions used in everyday speech.

1.3.1.3 The Discourse Analysis Approach(es)(9)

Under this rubric one can include those approaches that deal with the speakers' utterances in terms of the 'functions', 'speech acts' and

'propositions' which these utterances bear. Most of these approaches represent a shift from those traditional linguistic ones that have the grammatical notion of "sentence" as the main unit of analysis. This shift has brought about some conflicting views about the characterisation of such terms as 'discourse', 'text', 'discourse functions', 'discourse units' and 'units above the sentence' and the like. Consequently, any characterisation of DA according to these terms would appear to be rather crude and too general especially considering the wide range of uses with which these terminologies were associated and also considering the confusion their multi-facetted use in the literature has caused. However, Levinson's (1983:288) classification of these approaches into 'text grammar' and 'speech act' (or interactional) theories mentioned above seems to be the most reasonable one since his criteria for the classification are well-defined in contrast to those of the CA approach mentioned above.

Clearly, the DA approaches draw on many insights from SAT in its reliance on the specification of discourse in terms of 'speech acts' and on traditional linguistics with a major difference of the emphasis being placed more on the function and meaning of utterances than on their form (cf. Sinclair 1966; Sinclair and Coulthard 1975). The emphasis is put on the notion that there is a finite set of units which relate to each other in a hierarchical fashion and which combine in ways which are predictable by rule.

The DA approaches have been applied to a variety of spoken and written discourse registers such as "radio broadcast interviews" (Pearce 1973; Lang 1984), "Classroom interaction" (Sinclair and Coulthard op.cit.), "article introductions" (Swales 1981), "dramatic dialogues" (Burton 1980), "doctor-patient interaction" (Coulthard and Ashby 1976), "dentist-patient interaction and casualty consultations" (Candlin et al. 1974-1980), "therapeutic interviews" (Labov and Fanshel 1976)(10) and "face-to-face and telephone conversations" (Stenström 1984a). The main difficulty with

the DA approaches lies in its application to those discourse modes which include complex conversational features such as natural conversations and interview. This is also acknowledged by proponents of the theory themselves (cf. Sinclair and Coulthard op.cit.:4). There have been some attempts (cf. Burton 1981b; Francis and Hunston 1987) to facilitate its application to natural conversation and conversation-like activities but these attempts have not been fully successful (see discussion below, however), since the language of conversation is too variable and complicated to be accounted for in terms of the underlying acts that its utterances involve.

The DA approaches seem to be more easily applicable to certain highly structured types of discourse where there are strong norm expectations. It applies more readily to monologues, some kinds of classroom interaction and certain types of interviews with a simple exchange structure such as those discussed above. The approach does not apply easily to debates, discussions and conversations, since in these, speech acts are almost always difficult to identify, and in any case often multifunctional. It is of course obvious that they also contain many features of spontaneity, syntactic/semantic reformulation, repetition, slips, pauses, pragmatic particles, interruptions and overlapping speech.

However, one successful application of the approach to the language of naturally occurring conversation (including telephone conversations)(11) is found in the substantial work of Stenström (1984a) in which questions and responses in English conversation were thoroughly investigated using the exchange as the basic interactional unit, realised minimally as question, response and an optional follow-up move. This application shows clearly that the DA approach can be successfully exploited in analysing just certain features of SES like questions/responses, commands-requests/refusal- acceptance, greeting/greeting, and the like. It has proved up to now doubtful whether the approach can be applied to

longer stretches of SES, which involve a complex turn-taking structure, interruptions, simultaneous speech and so on. Work which can substantiate these remarks is needed.

The main conclusion that we can draw from reviewing the 'qualitative' theories of discourse analysis above is that they are more easily applicable to theoretical investigations of specific types of spoken discourse. Although various types of SES have been researched within the two trends of CA and DA, the only insights that one can get for the linguistic analysis of SES and more specifically, the language of INT and CON are more related to the organisational features of spoken discourse, turn-taking structure, functional descriptions of utterances on the level of discourse rather than its complexity on the lexical and syntactic levels of comparison. Even on the level of linguistic variation between types of discourse, the above studies seem to give us very little indeed.

1.3.2 Quantitative approaches

Three quantitative descriptive approaches are generally distinguished in the literature of variationist studies of discourse (cf. Johns-Lewis and Skelton 1987). They are a) the model which relies on the notion of variable within sociolinguistic variation (cf.Labov 1972); b) the model which derives ultimately from Speech Act Theory and c) the 'relative frequency model adopted by the research group at Aston University. The first model i.e. the variable and the variants model, which has come to be known as the 'quantitative paradigm' (Cameron 1988), is clearly **"more suited to morphological or morpho-syntactic variation"** (Johns-Lewis and Skelton op.cit:5). Adherents of this model believe that the analytic concept of the linguistic variable allows sociolinguists to **"quantify their data and to make correlations between linguistic and social structures"** (Cameron op.cit.:3). The second one (SAT), which has been criticised widely (see section 1.3.1.1 above) can hardly be applied within a quantitative

framework. The third model compares the relative frequency of a linguistic expression type in contrasted text types. It is thought that this model yields some fascinating results (ibid:8), which can contribute to the understanding of language variations (See also Farag 1986; Asaad 1987; Al-Jabr 1988; Rimmer 1988 as examples of applications of this model).

1.3.3 Choosing an approach for the analysis of Speech Exchange Systems

Having looked at the various trends that have dealt with the structure and description of spoken discourse and more narrowly SES, it seems appropriate at this stage to start looking for a suitable approach for applying the model proposed above (see Fig. 1.1) to the linguistic analysis of the data of INT and CON recorded for the purposes of this study. This necessitates a deeper look at three main issues; first the actual type of data collected for the study and the method of collection; secondly, the unit for the analysis of SES language which will satisfy the conditions necessary for the analysis of lexical, syntactical and informational criteria in the language of conversational variations; thirdly the criteria over which the above discussed model of linguistic complexity is based. This third factor is going to be discussed now leaving the other two issues to be tackled in considerable detail in CHAPTERS TWO and THREE below.

As has already been mentioned in the foregoing discussion, the study of language variations in English SES is necessarily a study of the specialised styles the participants in a conversation are employing at the time and place of the conversational encounter. As such, the selected approach for the analysis will have to draw on the type of stylistic approaches available in the literature. One can opt for the qualitative types discussed above. This will necessitate the introduction and employment of a more theoretical descriptive framework for the analysis and a lengthy treatment of the ways people produce talk, organise it and

exchange it in meaningful settings. This approach will not necessarily need a huge amount of data for its application. A few typically short specimens of conversations will suffice to provide the researcher with the necessary data for his qualitative investigation akin to the type of research adopted by the ethnomethodologists. The other point which should be mentioned about the qualitative types of approaches is that they fit specific types of language variations rather than others, such as those of 'literature' and 'humour' (cf. Crystal and Davy 1969:78) as those represent varieties which are **"fundamentally unspecifiable linguistically, and thus stylistically"** (ibid).

Another choice for the variational study of SES is provided by the quantitative approaches which will take into consideration all the linguistic features along the lexical and syntactical dimensions of analysis and which will give a clearer picture of the type of data handled in the analysis, its complexity and degree of variation from one particular conversational situation to the other. It will also pinpoint when a particular feature within the linguistic structure of discourse becomes a "**STYLE MARKER**" (cf. Enkvist 1978b:174), that is, a linguistic feature whose density in a text is significantly different from its density in another text (normally the norm). This can only be done by quantifying the occurrences of such features in selected speech texts. As such, a quantitative study of the linguistic features within two or more texts for the purpose of comparison will show the level and degree of variation among participants themselves and will show to a considerably greater accuracy the degree of complexity or simplicity of the language used by particular individuals with specific age and educational groups. Crystal and Davy (op.cit.:124) suggest that such an approach can be applied to conversational variations of discourse. They think that one can compare statistically any 'variety' of conversational English to that of natural conversation and according to the degree of linguistic similarity or difference one can label the variety under analysis as conversation or

non-conversation, so to speak.

A quantitative technique to the analysis of SES variations will conditionally require an examination of a large amount of data , large enough to make valid generalisations that might eventually help in the application of qualitative approaches as well. The larger the amount of data the more valid one's generalisations will be and the more realistic the comparisons will become (cf. Crystal and Davy op.cit.:62, Farrugia 1988:20).

Some arguments have been put forward against the validity of quantitative research methods in the literature (See e.g. Hatim 1981, Swales 1981, Widdowson 1979). However, many other counter-arguments have been introduced by other researchers to the importance of quantitative analysis, a method which is commonly used in stylistic and sociolinguistic research. Enkvist (op.cit.) emphasises the importance of quantitative analysis to the study of stylistics. Johns-Lewis and Skelton (1987:2) point out that :

Aston University

Content has been removed for copyright reasons

The technique which is going to be adopted for this work will be mainly the 'relative frequency approach' outlined above, but that will not prevent us from adopting, at certain stages of the analysis, some procedures employed by the other quantitative or qualitative approaches. Descriptions will certainly be provided of the data used for the analysis, its overall organisation, the participants who have produced it and its most important features and characteristics. The main generalisations, however, will be drawn from the statistical analysis of the features selected and counted for making the comparison. There will be more elaboration on the technique and methodology in the next chapter.

1.4 The present investigation

1.4.1 Aims and purposes

We are now in a position to summarise the main purpose of the analysis carried out for the present study, which is to investigate what has been referred to above as 'linguistic complexity' in two samples of spoken discourse, conversation and interview. The definition of complexity adopted includes lexical density as well as other more traditional measures of syntactic complexity : subordination versus coordination; appositional clauses, comment clauses, ellipticals, etc. Other dimensions of the linguistic complexity of SES which will be looked at include such features as referential relations and pronominalisation, the structure and organisation of information within the units of information that constitute the whole body of discourse. General descriptive features of conversational organisation such as turn-taking mechanism, (cf. Sacks et al 1974), interruptions (cf. Ferguson 1975), simultaneous talk (cf. Coates 1986) and the like will be commented on in the data under investigation and will be shown in the transcription of the data for future research. Other features common to SES types of discourse, and which are thought to influence its linguistic complexity and structure such as question tags, back channels and discourse markers will also be considered (see, however, Stenström 1984b).

For the purpose of achieving the aims above two sets of data of INT and CON are analysed for this research, both produced by the same speakers. The first comprises recordings of interviews, relatively formal and pre-thought about and planned, in which the eight pairs of informants responded to questions put to them by an interviewer (a university chaplain in his capacity as spiritual mentor and guide). Each pair of interviewees had a long-standing friendship, and each pair was

interviewed together. The second set of data consists of the talk which followed the interview sessions between the two friends. More about the collection of the data will be provided in CHAPTER TWO below.

A first motive for the research at hand is provided by what is thought to be a gap in the linguistic research on English SES. While something is known of the surface characteristics of SES (ellipsis, left and right dislocations, fragmentation, word order inversion, parenthetic comments and 'softeners' (Crystal and Davy 1975), less is known about the extent to which the term 'conversation' subsumes a variety of distinguishable classes of communication. The decision to integrate the structural characteristics of conversation and formal interview arises out of the belief that complexity in the spoken domain, is still not fully understood. This is hardly surprising since relatively less is known about variations within it. As Stubbs (1983:37) points out **"Much linguistic description neglects intuitively important aspects of linguistic competence, because it neglects a close study of naturally occurring conversational language".**

To understand the structural dimensions along which conversation varies, it is important to consider syntax, lexis and discourse function **together**, and not in isolation from each other. There will thus be a better chance of discovering interrelationships and dependencies between these linguistic levels. More than fifty years ago, Firth (1935:32) noted that :

Aston University

Content has been removed for copyright reasons

This indicates that the understanding of **'what language is'** and **'how it works'** is impoverished because of the lack of understanding of conversation. It is only recently that analysts have realised the importance of the message implied in Firth's (and before that Malinowski's) notion of

"context of situation". A systematic study of conversation has been undertaken by ethnographers and ethnomethodologists for the last three decades, and to them we owe current understanding of turn-taking regularities, conversational ellipsis, the significance of pause and overlap, internal structures (such as adjacency pairs), participant perspectives and participant knowledge. What we know far less about is the characteristics of conversation at the structural level. Studies have been carried out of the syntax of oral as opposed to written varieties (e.g. O'Donnel 1974; Poole and Field 1976; Kroll 1977; Akinnaso 1982; Chafe 1982; Beaman 1984; Bilton 1987), but there are practically no studies of the syntactic dimensions of conversation as opposed to other spoken modes, nor of those dimensions along which conversation varies. Specifically, what this study provides which has hitherto not been available is information of a quantitative nature on the usage of syntactic characteristics of conversation and interview.

Finally, the study of INT and CON, will in general terms provide insight into the structural characteristics of spoken interaction. These insights may eventually influence language teaching materials and methods of teaching English conversation to speakers of other languages. This point will be elaborated on in the concluding chapter below.

1.4.2 Field of research

As has already been mentioned above most of the research work on linguistic complexity and variability is usually carried out under the general heading of 'sociolinguistics', since the latter represents a free discipline under which most studies on language interaction can be classified (cf. Hatim 1981, Yule 1981, Brown and Yule 1983a, Orestrom 1983, Stubbs 1983, Stenström 1984a, Thavenius 1984) and since most of these studies employ the methodologies which are normally advocated by the sociolinguists proper, such as 'quantification', 'variation', 'description',

'empirical examination' and so on. It seems, then, that it is quite plausible to include the present investigation under the general heading of 'sociolinguistics' as well, since most of the methods and practices employed here are the same as those above in one way or another. One branch of sociolinguistics within which the present investigation fits properly is the one referred to as **"secular linguistics"** (cf. Trudgill 1978:12), a term which was originally developed by Labov (1966) to refer to those sociolinguistic studies which are concerned with learning more about language, and to investigate topics such as the nature of linguistic variability, the structure of linguistic systems and the mechanisms of linguistic change (cf. Trudgill op.cit.). This will become more clearly apparent in the course of the analysis of the data and results of the investigations obtained hereinwith.

1.5 Conclusion

Throughout this chapter, a lot of emphasis has been placed on the notion of linguistic complexity especially as far as SES is concerned, the reason being the great gap within (socio)linguistic research in this important field of knowledge. This is why this research sets out to investigate this phenomenon in two types of English SES, INT and CON produced by the same speakers under relatively different conditions. A hypothesis is being put forward, supported by some research investigations on spoken versus written discourse variations, that an SES displays a certain level of linguistic complexity on the syntactic level (Halliday 1979, 1985b, Beaman 1984), depending on the actual use of lexical and syntactic structures to express information. The main task in this thesis is to test the validity of this hypothesis against the data of INTs and CONs and also to see the differences between the two sets of data in this respect. A statistical analysis on the three levels of lexis, syntax and information structure will be carried out on the corpus for checking the hypothesis mentioned above.

The review of the literature surveyed above has shown contradictory points of view concerning the issue of linguistic complexity. It has shown that there is a lack of research on SES and hence a need for a linguistic theory which takes the language of SES into consideration.

Studies in both CA and DA are rich in their theoretical insights as far as the organisation, sequencing, speech acts of discourse are concerned. However, they offer the linguist very little indeed about issues of linguistic nature such as complexity, density, and variability of structural items and constituents.

The next two chapters will be devoted for some methodological considerations, which include the collection of the data (CHAPTER TWO) and its segmentation into units of discourse for micro analysis (CHAPTER THREE). CHAPTERS FOUR, FIVE and SIX will be devoted for presenting and discussing the results of the statistical analysis in terms of structural complexity on the levels of lexis, syntax and information structure.

FOOTNOTES TO CHAPTER ONE

1. Notice, however, the controversy associated with the terms 'variety' and 'variation' in the literature. Some have used the two terms interchangeably (e.g. Hatim 1981:1), while others use them confusingly to cover a wide range of senses (cf. Gregory and Carroll 1976 for a review). In this research we distinguish between the senses of the terms drawing mainly upon the general sociolinguistic characterisation of the two terms (cf. Hudson 1980: 24).

2. See also the work of other sociologists and sociolinguists (e.g. Maccoby and Maccoby 1956; Denzin 1970; Allen and Guy 1974; Wolfson 1976; Milroy 1980), psychologists (e.g. Mahl 1954; Goldman-Eisler 1954a, 1954b, 1958; Maclay and Osgood 1958; Boomer and Dittman 1964; Wayne 1971) and social psychologists such as Brenner 1978,1981; Brenner et al. 1978; the collection of papers in St. Clair and Giles 1980).

3. See, however, Zora (1986) for a brief treatment of the relevance of one prosodic parameter, pausing, to the complexity issue where a pilot experiment gives evidence that in a more comprehensive study prosody ought to be considered.

4. See, for example, Labov and Fanshel (1976), who studied the language of therapeutic discourse, Farrugia (1988) who studied the language of newspaper editorials and Cheshire (1978) who studied the present tense verbs in Reading spoken English.

5. 'Spoken discourse' is used here in its most general sense to refer to all types of monologic and dialogic types of verbal interaction.

6. This is originally taken from the Sun Newspaper of its issue of 5th May 1988.

7. This is originally taken from I. S. Rombauer and M. R. Becker (1931), "The Joy of Cooking" (P. 408)

8. See Grice (1975) and Levinson (1983) for a full definition of and characterisation of these maxims.

9. Levinson (1983) puts all these approaches under one heading "Discourse Analysis Approach" as opposed to to the "Conversation Analysis Approach". In this research, they will be referred to collectively as the "DA Approaches" as they represent a collection of approaches and not a one single approach.

10. Labov and Fanshel (1977) seem to be using the term 'conversation' in its widest sense to cover types of spoken interaction which involve the exchange of talk and turns i.e. 'conversational activities' in Levinson's (1983) terminology.

11. Stenström's data consist of 25 conversations from the London-Lund corpus of Spoken English, which vary between spontaneous face- to-face and telephone conversations (See Stenstrom 1984a for more details).

CHAPTER TWO

The DATA : COLLECTION, DESCRIPTION AND ANALYSIS

2.0 Introduction

The aim of this chapter is to explain the methodologies used for obtaining the two sets of data to be used for analysis and comparison of one set against the other. This includes an elaboration on the choice of participants i.e. interviewer and interviewees, the setting, topics and so on. Other information about the data including transcription methodology and conventions will be touched upon as well.

The methods used for collecting natural(1) conversational data have been of some concern to both sociolinguists and discourse analysts (cf. Gregory 1967; Labov 1972c; Crystal 1980; Brown and Yule 1983a; Stubbs 1983). The difficulty faced by all analysts of conversational data is summed up in Labov's observer's paradox :

> "To obtain the data most important for linguistic theory, we have to observe how people speak when they are not being observed." (Labov 1972c:112)

The main problem lies in the question of how to go about getting natural data from subjects without letting them be aware that they are being observed and recorded. If this is done surreptitiously, one runs the risk of having behaved unethically in eliciting the desired natural type of talk for analytical purposes, even if subsequent remedial apologies are presented and permission sought to use the data in research.

Since the major aim of this study is to compare the language of talk used by the same interlocutors in INT and CON, certain methodological considerations were important. The problems were (1) to find suitable

subjects who would be able to act both as interviewees in formal planned interviews, and as participants unconsciously engaged in natural informal conversation for some time. This necessitated a careful selection of (2) an experienced interviewer who would conduct interviews skillfully, and then engage in informal conversation in the best possible way; and (3) a suitable setting to collect the spoken data, one in which all situational and technical factors such as place, time and recording facilities could be held constant across the two speaking tasks.

2.1 Collecting the corpus of data

2.1.1 Some preliminary considerations

As has already been outlined in CHAPTER ONE, most studies of text type differentiation do not take into account the need to control context of utterance, within which all variables are controlled. Topic, mode of delivery, participants and setting are the main factors to be controlled. The controversial issue of topic is to be discussed in Section 2.1.5 below. There is no problem concerning mode of delivery in our data since it is wholly of the spoken conversational type and no attempt has been made to get written discourse from our subjects for comparison with their spoken output. Both the setting and the participants were identical in both the interview and the conversation encounters, the same participants performing the two tasks to be compared, as will be seen in the following sections.

It was essential that the interviewer should know the interviewees sufficiently well to engage in informal conversation. To this end, the interviewer selected was (JA), a highly skilled and popular chaplain at Aston University (see Section 2.1.2 below). With each pair of students, the format is the same : JA elicits information from them about their personal life at the university, courses, social and religious activities, the interview

comes to a conclusion, and the interaction slips into chat. The interviewer, JA, becomes deliberately more and more unobtrusive as the chat progresses. He appears to wish to allow students more freedom. The interviews have a real purpose for the interviewer : they are not a vacuous exercise. He uses these interviews to enrich his personal contact with the students, and to gain a fuller understanding of the difficulties and problems they are likely to face in their academic and social life. In this sense, the interviews conducted can be described as authentic in that they will enable him to fulfil his role as chaplain. The authenticity of the interviews is of course crucial to the distinction between interview and conversation.

Full freedom was given to the interviewer to choose his respondents and questions. The interviewer and the interviewees are all native speakers of English. At first the interviewees were 18 in number (later they were reduced to 16 for reasons of poor quality recording), two of whom were of Scottish origin, two Welsh natives and the remaining twelve were English students coming from different parts of England. Most of them were religious students belonging to different denominations and religious societies. The chaplain's relationship to these students comes from his contact with them through these societies or through the Guild of Students.

The permission of all parties was obtained for recording interviews. Little attention was paid to the tape recorder after the first task (the interview) was over, i.e. during the conversation. This is the reason why the interview task was carried out before the conversation. Eight encounters were taped, each with two students interviewed jointly. In all eight encounters, by the time the interview ended and the conversation started, the participants were looking fully relaxed, thus producing, as far as could be judged, a fairly natural sort of conversation, with the inhibitors of communication(2) (cf. Gorden 1969:88) eliminated as far as possible.

2.1.2 The interviewer

More comments on the choice of the interviewer may cast additional light on the data. First of all, his position as a chaplain at the university and his caring pastoral relations with the interviewees (through his religious services and gatherings) provided an atmosphere of trust within which to conduct the interviews. Secondly, he is a popular figure. Thirdly and perhaps most importantly, he is an experienced elicitor of personal information. He keeps a stock of good recording facilities in his office. His experience as a radio broadcaster(3) undoubtedly contributes to his interviewing skills.

It will be clear to the reader that a specific type of interview material was collected. It is possible that it broadly represents "counselling interviews" - or at least interviews which are essentially non-combative, supportive, and involve self-evaluation. This is not to say that the interview data collected adequately represent all or any class of interview. The task of proving or disproving comparability of the interviews examined as against other categories would go well beyond the scope of this thesis.

2.1.3 The interviewees

The following table gives a summary of the social and educational status of the participants in the eight sessions of INT and CON recorded for this study :

PAIRS	INITIALS	AGE	SEX	OCCUPATION	REGIONAL ORIGIN
I	HC	23	M	Postgrad.	Edinburgh (Scotland)
	KW	28	F	Postgrad.	Lampeter (Wales)(4)
II	JH	23	F	Postgrad.	London
	BG	23	F	Postgrad.	London (5)
III	RH	24	M	Postgrad.	Midlands
	DD	24	M	Postgrad.	Aberdeen(Scotland)(6)
IV	RF	21	M	Undergrad.	London
	PM	20	M	Undergrad.	London(7)
V	HK	21	F	Undergrad.	Northamptonshire
	KSH	20	F	Undergrad.	Northamptonshire (8)
VI	AM	22	M	Undergrad	Bournemouth
	CB	23	F	Undergrad	Runcorn (Cheshire)(9)
VII	JC	20	M	Undergrad	Devonshire
	GM	20	F	Undergrad	Berkshire(10)
VIII	AB	21	M	Undergrad	Berkshire
	KS	20	F	Undergrad.	Bangor (Wales)(11)

Table 2.1 The Participants in the INTs and CONs

The criteria for pairing the students as interviewees were set in advance and they were as follows :

a) roughly the same age (early twenties)

b) roughly the same educational level (undergraduate, postgraduate)

c) similar interests (political, religious)

d) pre-existing friendship between interviewees

e) balanced number of males to females (four pairs mixed sexes, two

groups just males and two groups just females)

It was felt that regional accent could not be controlled in any consistent way without constraining criterion (d) : since the study was not designed to focus on factors which were phonetic (as opposed to syntactic), that the regional accents of participants were not uniform was not regarded as a problem. It is assumed, in the absence of evidence to the contrary, that the major characteristics of conversation are shared by all speakers of the language. This assumption may turn out to be wrong ; but it is a necessary baseline at this stage.

The decision to interview two people at a time was taken in the hope of getting more varied and more natural data containing as many features of dyadic conversation as possible. Another reason for getting couples of interviewees to participate jointly in interview was to minimise participant anxieties, and provide a mutual support situation for the casual conversation which was to follow the interview encounters as has already been explained above.

2.1.4 The setting

The session recordings took place at the Martin Luther King Centre(12) in the interviewer's office which is a quiet place, suitable for recording purposes. An efficient UHER 4000 REPORT recorder was used in the recording, which was done using a speed of 3.3/4 (i.p.s.). The sessions took place over a period of six months. They were approximately half an hour each; the first ten minutes or so were in the form of planned interview and the rest of the time in the form of chat(13).

The diagram below shows roughly the sitting posture of the interviewer and the interviewees during the interview encounters with the microphone on the ground directed towards the interviewees, at a

distance of approximately 2-3 feet. At the end of each interview both the recorder and the microphone were pushed deliberately out of sight under the small table between the two interviewees, and the interviewer adjusted his seat in a way so as to show the interviewees that their major task was over and what was left was time for chat and a coffee.

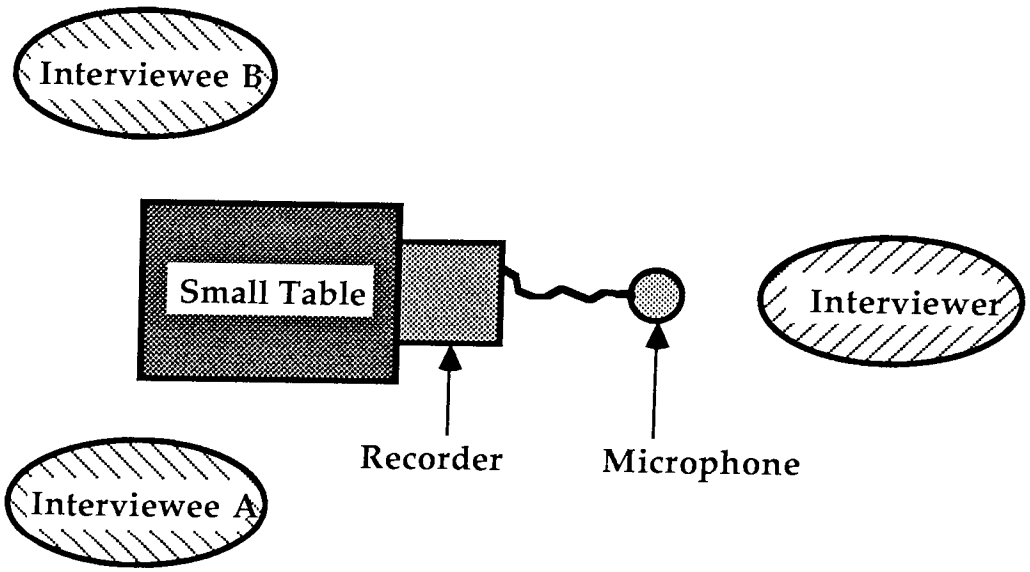


Figure 2.1 Diagram showing seating positions of participants during interview sessions.

2.1.5 The topics

It is practically impossible to have full control over topics discussed in naturally occurring conversations, but that is much easier during interview sessions due to the role of the interviewer as the organiser and controller of what is discussed. In most interview encounters, the dominant role of the interviewer as a presenter, expander and terminator of topics is very clear. An interviewer can easily shift the direction of talk if he/she wants to, can introduce new topics or can relate topics to each other in one way or another. This, indeed, can be one of the major skills of an interviewer.

During some of the conversations, the participants continued discussing topics raised in the interview sessions, but the continuation of interview topics into conversation exchanges was not problematic since the interview, by definition, rests mainly on the participants' role and the relationship which holds between them. And since the interviewees are friends with similar interests and are usually accustomed to going out together and chatting about various sorts of topics, there was no difficulty in obtaining samples of natural talk.

Generally speaking, the interviewer's questions, although not identical were comparable in content across the eight interview sessions, and mostly related to the respondents' life at the university. They touched on previous school experiences, religious or political views, social life, making friends and some other general topics. No such control over topic during the casual conversation encounters was possible.

The strategy employed by the interviewer during the INT encounters was to start the session by questions put to either of the two respondents who would then give their answers in accordance with the questions put. Then a sequence of questions and comments would follow as they arose from the topics discussed. Like Denzin's (1970) non-standardised interviews, no prespecified set of questions has been employed by the interviewer nor have the questions been asked in a prespecified order. Some of the questions are similar to those described by Labov (1972b:354), the original question calls for only one or two words, which is normally a yes-no question. Then the respondent becomes involved in a more detailed account of what he/she has to say.

During the casual conversations which followed the interviews, the participants were intentionally left on their own, chatting freely the way they liked. It was thought that pursuing predetermined topics would make

the participants feel that the present conversational situation was planned in advance. So in some of the encounters we had the interviewer reentering the office and exchanging free chat with the other participants. This we thought did not affect the spontaneity of the talk since the role he was now achieving was not that of a formal interviewer but that of a friend. Coffee, sweets and (on one occasion) dinner added to the 'naturalness' of the conversation and to the 'relaxation' of the atmosphere.

2.1.6 The transcription

The recordings were transcribed by the analyst and later checked by two native speakers of English, one of whom was the interviewer himself. Then all the informants were contacted again and a final check of the transcripts against the recordings were made by them in quiet Audio/Visual laboratories(14) to ensure accuracy of the recorded data and the transcriptions. However, it should be made clear that accuracy is unlikely to be absolute even when a talker is checking his own output. There were some bits of talk in the recordings (especially during the conversation encounters) which were not clear even to their speakers themselves. These bits were left out of the transcripts and were not included in the calculations.

The transcription conventions were adapted from a number of resources most important of which were : Levinson (1983); Brown and Yule (1983a) and Tannen (1984). Punctuation marks were used for notational purposes and not to mark syntactic boundaries. A list of these conventions including the special tags for specifying referential entities for CHAPTER SIX is presented at the beginning of this thesis.

2.2 The analysis

The crudest level of analysis involved starting with counting all the words produced by each speaker, including the interviewer in the two modes under investigation (See Tables 3.1 and 3.2 below). Then an overall segmentation of the transcribed texts into units -Major and Minor Units- for linguistic analysis was made depending on semantic, syntactic and intonational cues (See CHAPTER THREE below for details of these units and the method and criteria for the unitisation of data). The next step was to relate these units to speaker turns (See Tables 3.3, 3.4 and 3.6 below). Then a statistical analysis was done of the general linguistic features of the two speech types under investigation. The counts include the ratio of lexical to grammatical words (CHAPTER FOUR) ; syntactic clause types and clause complexes according to their grammatical and semantic functions (see Quirk et al. 1985:Chapter 15) (CHAPTER FIVE) and units in referential entities for the study of information structure and organisation (CHAPTER SIX)(15).

Computer programs were used to handle most of the counts, drawing tables and graphs, and applying tests of significance, as will be shown in various parts of the thesis. The major test of statistical significance used was Wilcoxon's Signed Rank Test (Henceforth WSRT)(16) (see Medis 1975; Butler 1985, for details of this test and its main applications).

The standard deviation (SD) has been found for all the tables in the analytical chapters. It has been adopted as a measure of variability or what statisticians normally refer to as **'dispersion'** (cf. Harper 1965). This measure has been chosen as it represents **"the most widely used measure of variability for variables with interval or ratio level of measurement"** (Butler 1985:37).

A more reliable test of variability which can fit the comparison made between INT and CON results is the one normally referred to in statistical research as **'coefficient of variation' (or V)** (Harper op.cit.). This test yields results which can show the amount of variability between the two sets of data (if at all present). The value of V is usually arrived at by applying the following statistical formula :

Coefficient of Variation (V) = 100 X SD/MEAN

In order to check the accuracy of the MEAN values obtained for the 16 speakers in INT and CON , the '95% Confidence Interval' (see Garrett and Woodworth 1926; Harper 1965 for statistical formula and details) has been worked out using an EXCEL computer program. This interval is produced in such a way that there is a probability of 95% that the actual MEAN lies inside it. It produces two boundaries (lower boundary 'LB' and upper boundary 'UB', within which the actual MEAN falls. These represent the limits within which the true population MEAN lies in a probability of 95%; the other 5% would not. In other words, the upper and lower bounds of the 95% confidence interval are the 95% CONFIDENCE LIMITS for the MEAN.

The LB and UB limits for the MEAN values in the data have been calculated and included in the data tables. As such the statistical tables in the analytical chapters contain the frequencies (as OPTWs or percentages) of the features investigated, the MEAN, SD, LB, UB and V. Interpretation of the results follows each table as bound necessary.

2.3 Conclusion

The criteria for choosing the subjects for this analysis and the methodology for collecting, transcribing and analysing the data were reviewed in this chapter. Elaboration on the setting in which the recordings took place, the participants and the interviewer was made. The role of the interviewer, it must be said, was critical in collecting this corpus of data. He also helped to transcribe some unclear recordings, which, otherwise, we would not have been able to do since the analyst was not present at the recording scene. The participants also took part in solving most of the transcription problems which were not very clear to the other transcribers because of poor recording quality especially during the casual conversation encounters during which the microphone had to be concealed out of sight.

A brief overview of the method of analysis was touched upon in this chapter. Related to this issue is the unitisation of spoken discourse into units for linguistic analysis. In the following chapter we discuss some of the major approaches to the segmentation of spoken discourse, outline the approach adopted and offer a justification for its adoption in the present research.

Footnotes to CHAPTER TWO

1. The term 'natural' is used here in the sense of Stubbs (1983) to refer to that type of talk which is ***"spontaneous, unplanned and casual"*** (P.33) as opposed to that type which is ***"invented, introspective, intuitive and hypothetical"*** (ibid) (*Italics* in original).

2. Gorden (1969 : 90) defines *'inhibitors of communication'* as **"any social-psychological barrier which impedes the flow of relevant information by making the respondent unable or unwilling to give it to the interviewer at the moment"**. He distinguishes eight of these inhibitors : **competing time demands, ego threat, etiquette, trauma, forgetting, chronological confusion, inferential confusion, and unconscious behaviour**, respectively (p.104).

3. The Interviewer, (JA), has also got links with BRMB Broadcasting stations for the Midlands as a presenter of four-hourly weekly program (Open Air) which is broadcasted live every Sunday evening from 10.00 p.m. to 2.00 a.m. Live telephone interviews, meetings, comments and music are presented during the program.

4. HC and KW are very close friends, members of the Catholic Society. HC is continuing a postgraduate course whereas KW has temporarily interrupted her studies to join a two-year job in Europe.

5. JH and BG are very close friends who are both postgraduate students on sabbatical. JH is President of the Guild of Students at Aston University. BG is her Vice President Administration. They are both highly politically educated.

6. DD and RH are very active members of the CND *(Campaign for Nuclear Disarmament)* and are, accordingly, highly politically orientated and educated. They are close friends. They are both pursuing a Ph.D course.

7. RF and PM are close friends and members of the Methodist Society at Aston University. They are very religious people and join in a great number of religious meetings and activities.

8. HK and KSH are two religious, close friends of the Methodist Society who are actively engaged in the religious activities of their group.

9. AM and CB are undergraduate students on sabbatical for a year as Vice President Services and Welfare at the Guild of Students. They are close friends.

10. JC and GM are friends who have joined the University recently. They are both members of the Methodist Society.

11. AB and KS are very close friends, members of the Christian Union at Aston University. They are very religious people and join in religious activities with their group.

12. The Martin Luther King Centre is a small building on Aston University Campus, situated near the halls of Residence. It is frequently used by student groups and societies for meetings, gatherings, prayers and so on. The interviewer (JA) occupies the main office in the building and he is responsible for its booking and running.

13. The term 'chat' is used here to mean the same thing as informal conversation.

14. These are modern fully computerised laboratories, recently installed at the Department of Modern languages, Aston University.

15. An early pilot experiment on part of the data was done on the relationship between syntactic complexity and pausing in terms of their frequency, functions and distribution within the syntactic units of discourse (see Zora 1986).

16. Wilcoxon's Signed Rank Test (WSRT) has been used to test all the statistical results in the analysis chapters. A Macintosh computer statistical program called '*STATWORKS*' has been applied to get the significance of the results.

CHAPTER THREE

THE UNITISATION OF SPOKEN DISCOURSE

3.0 Introduction

The segmentation of spoken discourse into units (henceforth 'unitisation'), unlike that of written discourse, for the purpose of facilitating its analysis for linguistic research is problematic and has proved to be controversial especially on the syntactic and semantic levels. Prosodic units of discourse do not always coincide with syntactic units; the sentence as a main linguistic unit does not work as properly as it does with written, well-formed discourse. In this chapter a review of some of the attempts that have been made in this concern is undertaken and assessed in the light of segmenting SES, as opposed to the more syntactically organised types of spoken discourse such as prepared monologues, dramatic dialogues and the like. An eclectic approach is adopted to segment the language of INT and CON with some justificatory arguments put forward concerning its application in this study.

It should be made clear at this point that the unit of analysis proposed in this research has been adopted to facilitate the process of analysing spoken discourse and it is by no means equivalent to any one particular syntactic unit such as the clause. It can be equivalent to a word, a phrase, a clause or a clause complex. The frequency count of the linguistic units (words, phrases, clauses and clause complexes) is done within these discourse units and is compared to their general frequencies. The aim is to find out how the language of INT differs from that of CON concerning the total organisation and length of these units and also concerning their internal structure in as far as lexical, syntactic and informational frequencies are concerned.

3.1 The unitisation of spoken discourse

3.1.1 Problems and prospectives

Defining a unit of linguistic analysis for spoken discourse has proved to be controversial in current research. Practically, the only concept to obtain unanimous acceptance is the inapplicability of **the sentence** as a unit for analysing spoken interactive discourse. This notion apart, there is a distinct lack of agreement as to an appropriate analytical unit for spoken discourse. However, approaches seem to favour a prosodic unit mainly recognised by an intonation contour and a prominent syllable carrying most or all the information **"weight"** (Crystal and Davy 1969; Crystal 1975; Halliday 1967,1985a; Brazil et al. 1980; Lierberman 1980). Other approaches (O'Donnell 1974; Kroll 1977; Beaman 1984; Farag 1986) have used syntactic and/or semantic criteria for defining a unit which would be an essential component of every internally cohesive and coherent piece of discourse. Some have approached the issue from a cognitive perspective, and proposed **"propositional units"** (eg.van Dijk and Kintsch 1983); **"temporal cycles"** (see Beattie 1983); **"focuses of consciousness"** realised linguistically as **"idea units"** (the combination of which would produce a **"center of interest"** (Chafe (1980). The present study does not, of course, resolve the issue. While the issue of an appropriate unit of analysis is clearly important, it is seen as subservient to the question of general linguistic differences between INT and CON. The rationale for choice of unit of analysis is presented in Section 3.1.2 below.

3.1.2 Types of spoken discourse modes and the unit of analysis

It is axiomatic that the term 'spoken discourse' is too general covering too wide a variety of verbal activity types. We have already explained in CHAPTER ONE above how spoken discourse has been treated as falling

under one of two broad and overlapping categories **'monologic'** versus **'dialogic'** discourse (Gregory 1967). The monologic types are more readily segmentable into clear-cut units (see examples below), and are similar in this respect to written discourse types. The dialogic types (i.e. SES 'see Section 1.1 above'), on the other hand, are in general more difficult to segment. Various researchers (e.g. Chafe 1982) have commented on the fragmentary nature of spoken discourse in SES, on which may be superimposed a tendency by speakers to alternate between cycles of greater or lesser fragmentation - i.e. fragmentation is neither uniform nor consistent. A further source of variation cannot be ignored, namely the fact that SES types are themselves not a single 'variety' or 'category' but rather a collection of types (Sack's et al's (1974:7) and varieties of interactional discourse within which variation in syntactic tidiness is (clearly) identifiable. The following extracts from published data show how SES can vary between the two extremes, with syntactic chaos at one end, and near-correlation with conformity to sentence-like units (characteristic of the written mode) at the other. Notice that the latter types are more noticeably amenable to easy and well-organised units which are often coherent from the points of view of syntax and semantics and prosody (when, for example, read aloud) :

(1) <u>Teacher led classroom discussion</u> (From Cheshire 1982:19)(1) :

Aston University

Content has been removed for copyright reasons

Aston University

Content has been removed for copyright reasons

(2) <u>An interview with a regional informant</u> (caretaker from Walsall, with a very clearly recognisable distinctive West Midlands accent) (from Hughes and Trudgill (1979:55).

Aston University

Content has been removed for copyright reasons

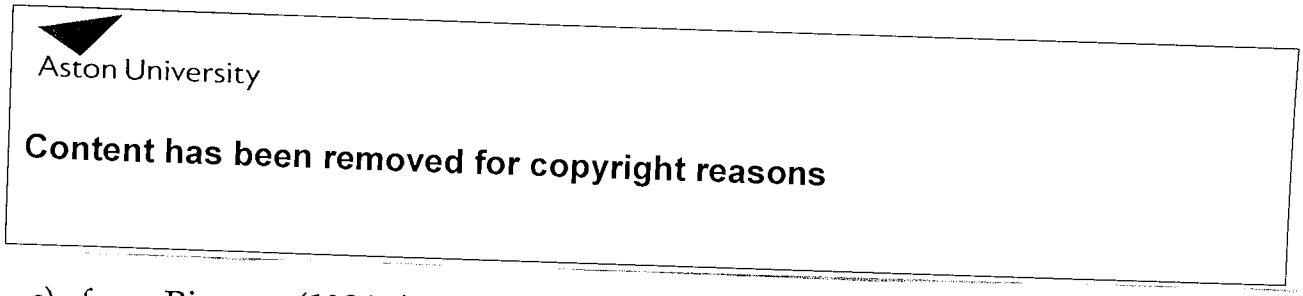
(3) <u>Oral Narratives</u>

a) from Beaman (1984:59)(2)

Aston University

Content has been removed for copyright reasons

b) from Chafe (1980:19)



c) from Rimmer (1984:5)

(4) <u>Casual conversation</u> : The excerpt that follows is taken from Sacks et al. (1974:25-6)

Aston University

Content has been removed for copyright reasons

Content has been removed for copyright reasons

(5) <u>T.V. political interview</u> : In order to see how interview fits into the typology of spontaneous spoken discourse, an example of T.V. political interview now follows (3) :

Interviewer :

> How much e:::m trust.. how much confidence can the public have in the good faith of your.. Department.. in the week in which one of your colleagues Lord Caithness e:::m attacked Brussels as a pest.. for seeking to extend the environmental protection as our rural countryside.. countryside..

MP :

well.. we want to look at the measures that are necessary on their merits.. and it is undoubtedly true that some of the measures that have been coming out from Brussels or which threaten to come out from Brussels would not in our view actually have the effect of improving our environment but as far [as the water Bill is concerned..

Interviewer :

[pest.. is that a right term.. pest..

MP [let me.. let me deal with your question.. but as far as the water Bill is concerned we are not actually asking people to take anything on faith.. the point about the water Bill.. is that it will all be there for people to see.. what we will achieve will be there for people t o see.. the independent bodies who would insure that those standards are met will be set up under the legislation.. we are after all the government that first made it necessary for.. the water authorities to publish information as a result of their monitoring of drinking water standards so that the public will see exactly what is happening.. you don't have to take it on trust.. it will all be open for everyone to see exactly what's happening..

Interviewer :

but again you can protect the environment without privatising it can't you.. you simply have this raft of regulations and insist that the land and the environment is protected.. you don't need to privatise it first to protect it..

MP :

Oh you don't have to privatise to enforce planning system and planning regulations I quite agree.. the point is if you want to achieve higher environmental standards in terms of water quality.. drinking water quality.. and all the rest of it.. you do need to put more money into it.. and the way in which to do that is to give industry access to private capital.

The striking factor about this sample of interview is its syntactic orthodoxy, and the fact that sentence-like units are distinguishable. Since this feature is shared with Cheshire's teacher led discussion, and to a lesser extent with the oral narratives, a hierarchy of fragmentation suggests itself with the most informal type of SES and certain types of narratives (e.g. Sacks et al's casual conversation and Rimmer's oral narrative) at the top and those types of well organised monologic-like interviews and discussions (e.g. Cheshire's teacher-led discussion, Chafe' and Beaman's oral narratives and the BBC broadcast interview), at the bottom. The types at the bottom are more easily segmentable into sentence-like units, whereas those at or near the top of the hierarchy show that sentence-like units are impossible to establish, spoken output in these cases being extremely fragmentary, and involving many reorientations and incomplete fragments in mid-stream. It is significant that transcripts of some kinds of data (the oral narrative quoted for example) are only partially interpretable without prosody, a term which subsumes all prosodic systems in Crystal and Davy's (1969) sense : i.e. pitch, tempo, loudness, voice quality, pause and other parameters of delivery. Prosody plays a highly significant role in segmenting during auditory processing.

It is worth pointing out that 'oral narrative' is too all- inclusive a category. The syntactic tidiness of the data from Beaman and Chafe (albeit repetitive of coordinative clause relations) is strikingly different from the syntactic chaos at the beginning of Rimmer's oral narrative extract. The significant factor is that Beaman's and Chafe's subjects perform for a single unfamiliar listener, whose role is simply to listen, whereas Rimmer's subject is clearly interacting with a peer group audience of several people, who share professional interests.

Little progress will be made in the analysis of spoken discourse if analysts ignore the presence of variation in the dimension of fragmentation. As the excerpts above show, the differences are too clear to

be ignored. There is a good deal of variation between the data of INT and CON in this study concerning the issue of fragmentation corresponding to that between the two extremes shown by the examples cited above. Generally speaking, the results of the analysis in this thesis (see especially those of CHAPTER FIVE) show 'fragmentariness', interruptive and overlapping speech, elliptical units, word disorder variation, and so on to occur roughly twice as much in CON than in INT (see Table 3.6 and 3.7 below). This, in turn, shows the language of CON to fit those types of casual talk which are at the fragmentary end of the continuum akin to Sack's et al's (op.cit.) casual conversation and the language of INT to be structurally similar to those at the non-fragmentary end akin to the teacher led discussion, the political interview and Chafe's 'Pear Narratives'. This is not to mean, however, that the language of INT is easily segmentable into sentence-like units and the language of CON is not. It just shows the degree of syntactic tidiness and order of those types of spoken discourse (including INTs) which fall at the non-fragmentary end of the continuum, which, as will be seen later on, is an important factor in facilitating the process of segmentation of any piece of spoken discourse into units for linguistic analysis.

3.1.3 Analytical approaches and trends

In the literature on spoken discourse, there seems to be some consensus that approaches to spoken discourse unitisation fall into three broad categories : prosodic(4), syntactic and cognitive. Various analysts have used one or more of these categories to identify a unit of discourse which can be used for the analysis of discourse. The following is a discussion of some of the approaches which are of direct relationship to the purposes of this study.

3.1.3.1 Approaches based on prosodic cues

Prosodic features such as intonation, pitch and pausing are characteristically significant in the specification and identification of prosodic units. Phoneticians recognise what they normally refer to as **'Breath Group'** (cf. Lieberman 1980) as the basic natural unit of speech. This has been adopted but modified to suit linguistic analysis by various analysts. The most widely used prosodic approaches are Crystal's Tone Unit (TU) and Halliday's Tone Group (TG). Crystal's TU is based on the belief that any analysis of connected speech should necessarily be based on prosodic cues particularly pitch and duration. Crystal and Davy (1969) state that the TU may be recognised by a combination of prosodic features which occur at its boundary and also by its internal characteristic structure. These boundary features are :

(1) a marked shift in pitch;
(2) modification to the final phonetic segments in the unit
(3) various types of pause.

The internal structure of the TU typically consists of :

Pre-head / Onset Nuclear Syllable / Tail
Head

3.1 (from Crystal and Davy ibid:27)

Content has been removed for copyright reasons

The Head consists of the stretch of utterance from the first stressed syllable (Onset) of the TU up to but not including the Nuclear Syllable. The Pre-head comprises **all the unstressed syllables before the Onset. The Tail** refers to the syllables after the Nuclear Syllable to the end of the TU.

Except in emphasis or contrast, the Nuclear Syllable usually occurs towards the end of the TU.

Brazil (1980, 1985) adopts the term Tone Unit but uses a different terminology to refer to the parts of the (TU) which, in his terms, consists of three parts the first and last of which are optional :

(Proclitic Segment) Tonic Segment (Enclitic Segment)

3.2 (from Brazil et al. 1980:40)

he was GOing to GO

that's a VERY TALL STO ry

it was a WED nesday

In Brazil's view prominence is only associated with the Tonic Segment which means, by definition, that there are no prominent syllables in the Proclitic or Enclitic Segments. However, he states that making a word prominent constitutes a meaningful choice. Therefore, it is up to the speaker to put prominence on any syllable of the TU for emphasis or contrast.

Halliday refers to the unit of intonation as the **'Tone Group'** (TG) (See Halliday 1967, 1970, 1985a). This is characteristically similar to Crystal's TU as far as the prosodic features are concerned. However, Halliday, too, uses different terminology to refer to the parts he distinguishes within the internal structure of the TG. In his terms the Nuclear Syllable is the Tonic (T). He does not distinguish the Tail terminologically but instead treats the Nucleus and the Tail together as the Tonic. Also he does not

distinguish between the Head and Pre-head but treats them together as Pretonic (P). So the TG consists typically of :

(P) T

which means obligatory Tonic T , optionally preceded by Pretonic. There is no separate post tonic element.

3.3 (From Halliday 1985a:283)

Aston University

Content has been removed for copyright reasons

3.1.3.2 Approaches based on syntactic cues

Linguistic units are recognised by syntactic and/or semantic cues. These include the use of the traditional notion of 'sentence', (Bowman 1964 among other traditional linguists), Idea Unit (Kroll 1977, Farag 1986), and T-Unit (Hunt 1971, O'Donnell 1974, Hasan 1988).

Attempts to use the traditional notion of 'sentence' as a unit for analysing spoken discourse are very few in the literature due to the problems and difficulties of identifying sentence boundaries in spoken discourse types, especially so in SES. As has already been argued in Section 3.1.2, SES types, especially the ones which are at the top of the 'fragmentary continuum', are characterised by the use of a large number of fragmentary units which cannot be accounted for using the traditional definitions of

'sentence'. Other difficulties, include the frequent use of certain expressions, lexical items, discourse markers and conversational cliches, all of which cannot fall within any definition of sentence whether traditional or otherwise. Crystal (1980:155) thinks that there are three main reasons why the use of the notion 'sentence' is problematic in conversational discourse **"indeterminate connectivity, indeterminate ellipsis and intercalation of structures"**. Conversational discourse, especially the naturally occurring and spontaneous types, contain, more often than not, a large number of weakly coordinated clauses and phrases, which are difficult to handle in sentential terms. Notice, for instance, the following example taken from Pawley and Syder (1975:11) :

3.4

Aston University

Content has been removed for copyright reasons

and she wanted to bring him back to Australia,..

This example contains probably 17 clauses, two of which are clearly cases of subordination (the ones introduced by *'that'* and *'because'*) while the others are either loosely connected by *'and'* or by a *'zero'* connective. Whether the entire example is one grammatical sentence in grammatical terms is debatable, as Pawley and Syder themselves seem to confirm as well.

Kroll (1977:85) summarises the view of the non- applicability of the

sentence to speech as follows :

Aston University

Content has been removed for copyright reasons

Bowman (1966) has discussed and actually used the 'sentence', in conjunction with intonation and other prosodic features, to analyse a corpus of English conversation. She distinguishes between what she terms **"major sentences"** and **"minor sentences"**. A major sentence in her terms consists of a subject and a predicate whereas all other types are taken to be of the minor type including fragmentary sentences, elliptical sentences, subjectless sentences and so on. The approach seems to stand on weak grounds as there are many features of conversational material which are not adequately handled by her segmentation.

Kroll (1977) uses the term **"Idea Unit"** (IU) to chunk her data of spoken and written narratives into units based on purely syntactic constituents. She states that an IU represents:

Aston University

Content has been removed for copyright reasons

The IUs Kroll uses in her analysis reflect clausal structures marked by words indicating subordination or coordination because she regards those words as showing points at which choices are made to create some kind of relationship between ideas. Kroll's criteria for identifying her IUs can be summarised as follows :

1. A subject and verb together with, when present :
   a. a direct object
   b. prepositional phrase
   c. adverbial element
   d. mark of subordination
2. Full relative clauses when the relative pronoun is present.
3. Phrases which occur in sentence initial position followed by a comma or which were set off from the sentence with a comma.
4. Verbs whose structure requires or allows a verbal element as object.
5. Reduced clauses in which a subordinator is followed by a non-finite verb element.
6. Post-nominal ing-phrases used as modifiers.
7. Other types of elements include :
   a. appositives
   b. absolutes
   c. verbals

Kroll applies these criteria to the analysis of units in both spoken and written narratives equally well, discarding any type of feature which characterises one and not the other. For instance, she discarded prosodic features altogether when analysing her spoken narratives, even when need arises to use one or more particular prosodic cues to make decisions about the plausibility of a certain structure. For example, Kroll's third criterion cited above is only relevant to written discourse. In spoken discourse, however, only prosodic cues such as pausing or intonation falls and rises would definitely tell the correct relationship of such phrases (see Kroll's third criterion) with the preceding or following constituents. This is why Farag (1986:57), in her application of the IU had to modify Kroll's third criterion to read as :

> "3. Prepositional phrases when in head position and when intonationally marked in speech or set off by a comma in writing."

Farag (op.cit) adopted the IU with certain modifications to suit her analysis of spoken and written narratives as she believes the IU is :

> "mainly a syntactic unit that can be objectively identified. In addition, it has the advantage of capturing the amount of grammatical manipulation 'work' needed to create the units of discourse" (ibid:53).

Farag, following Kroll (op.cit.), considers the IU beginning with a coordinator followed by a subordinator as one IU but lists it under both coordination and subordination.

Another modification added by Farag to Kroll's IUs is that she tabulates separately those IUs beginning with an adverbial element like 'obviously', 'presumably' and so on, which are also used as connectives. She also has a separate tabulation for IUs beginning with what she terms **"pragmatic particles"** such as 'well', 'anyway'...etc.

O'Donnell (1974) and more recently, Hasan (1988), following Hunt (1965), use the (T-Unit) as a basic unit for analysing spoken discourse. The T-Unit consists of one independent clause and the dependent clauses (if any) syntactically related to it, which can be equivalent to a simple or a complex sentence in traditional terms. The compound sentence would, therefore, contain more than one T-Unit. However, both Kroll (1977) and Farag (1986) show that the applicability of the T-Unit to spoken discourse is doubtful since spoken discourse is produced in quite a different way from written discourse and since, like the 'sentence', it has proved to be difficult and problematic to specify clear-cut boundaries for such units. It seems intuitively easier to apply the T-Unit approach to samples of spoken or written discourse such as those used by O'Donnell (broadcast journalistic interview and samples of newspaper columns) or by Hasan (1988) (native speaker/non-native speaker classroom interaction) where the language is highly controlled and organised and disfluency features are rarely used.

3.1.3.3 Approaches based on cognitive cues

Units of discourse which are thought to have cognitive interpretation (Chafe 1980) are mainly those that are recognised by a combination of prosodic and syntactic/semantic criteria (e.g. Chafe's Idea Unit 1980), hesitational features such as pausing, fillers and so on (Maclay and Osgood, 1965 ; Beattie, 1983) and propositional contents (van Dijk and Kintsch 1983). Only Chafe's 'Idea Unit' will be elaborated on here as it is the most relevant one to the purposes of the present analysis.

Chafe (1980) adopts Kroll's terminology of IU to describe a cognitively realised unit which is a linguistic expression of **"focuses of consciousness"**, based mainly on the following criteria:

1. a coherent intonation contour
2. bounded by pause of some sort
3. usually exhibiting a small set of syntactic structures

Although IUs have a tendency to be set off by the three of the above factors, Chafe mentions that all three are not necessarily always present, nor does the presence of one of them necessarily signal the boundary of IUs. These IUs have a mean length (including hesitations) of approximately 2 seconds or about 6 words each (ibid:14).

More recently, Chafe (1984) uses the term **"intonational unit"** to mean exactly the same as his IU without any mention of why the change has been made. Apparently this 'intonational unit' is defined identically to the IU and the change seems to be entirely terminological with no theoretical motivation.

Chafe believes that his IUs represent spurts of **"focuses of consciousness"** or short-time memories of events. Since the human

capacity for recall is limited, the function of each 'focus of consciousness' is to package short-term memories in manageable quantities. Chafe suggests the presence of a larger unit than the IU, (presumably cognitively linked to focuses of consciousness, which he confusingly calls the **"Intonational Sentence"** (See Chafe, 1980:25). This he defines as a sequence of idea units bounded by sentence final intonation and syntactic closure, expressing a single **"centre of interest"** and typically having a mean of three idea units each. A speaker organises his focuses of consciousness into the expression of centres of interest and judging that he or she has achieved adequate communication of an image he/she would use that sentence-final intonation and syntactic closure (ibid:27).

Chafe points out that his intonational sentence could be equivalent to one linguistic sentence :

> "spontaneous spoken language, then, suggests the existence of some sort of cognitive entity which I am calling a "centre of interest, and which corresponds roughly to what is expressed in a linguistic sentence"(ibid:29).

But again, it must be pointed out that the relationship between intonational sentence and linguistic sentence is not necessarily one-to-one as pointed out in Chafe's summary (ibid:48) :

> "I then showed various ways in which centres of interest fail to be expressed in intonationally and syntactically defined sentences in any clear-cut, one-to-one fashion" (ibid:48).

3.1.4 Discrepancies between prosodic and syntactic units

With the exception of those units based on syntactic principles such as Kroll's IU and O'Donnell's T-Unit all other units (especially those based on prosodic cues) display some incompatibility with syntactic groups. All

proponents of prosodic units in discourse such as Crystal, Brazil, Halliday(5), and Grimes mention that their units do not correspond on a one-to-one basis with any single syntactic unit. However, they all tend to believe that the clause is the nearest possible syntactic unit to their prosodic units (see for example Halliday's (1967, 1970, 1985a) characterisation of his 'Information Units' and Grimes (1975) for similar suggestions concerning his 'Information Blocks')(6). But there are many cases in spontaneous speech where one prosodic unit stretches across more than one clause. On the other hand, a prosodic unit can be as small as one lexical item or a back-channel item such as 'mmhmm' or 'O.K.'.

Chafe (1980) also relates his IU to the clause as the nearest possible corresponding syntactic unit. He states that, from the syntactic point of view :

> "there is a tendency for idea units to consist of a single clause, one verb with whatever accompanying noun phrases associated with it."(ibid:14)

Other researchers have talked about the general disagreement between prosodic and grammatical units. (See Farag,1986:55, for a review of these notes and also see the collection of papers in Johns-Lewis (1986) for a discussion of the relationship between syntax and prosody).

3.1.5 Uses of units in the literature

All the units that have been reviewed in the previous sections have been used by researchers in analysing and describing types of spoken discourse at various levels. Crystal and Davy (1969) use the TU in their linguistic analysis and description of different types of English spoken discourse such as casual conversation, legal and religious types of English, the language of ceremonial and sports commentaries and sports and so on. Their research and other research done in the field imply that TUs help to

distinguish one discourse mode from the other(s). For example, Crystal and Davy (1969:159) state that the language of religion and sermons contain short and level tone units. Speakers in casual conversation have a strong tendency to break up potentially lengthy TUs whenever possible. Casual conversation is characterised by a relatively high proportion of incomplete TUs, unlike other spoken discourse modes (ibid:106). The language of commentary, on the other hand, involves the use of longer tone units (ibid:132).

Many researchers have adopted the TU in their investigations of spoken discourse on different linguistic levels (See for example Quirk et al. 1972, 1985, Brazil et al., 1980, Brazil, 1985). The TU has also been used as the basic unit to analyse large amounts of various types of spoken discourse in two computer-based projects; The Survey of Spoken English, University of Lund and its sister project, The Survey of English Language at University College London (See Svartvik and Quirk 1980 and Svartvik et al. 1982), and other research work which is based on data extracted from these projects (see for example, Oreström 1983; Stenström 1984).

Halliday (1980, 1985a) uses the term **"Information Unit"** for his Tone Group when talking about the informational structure of units of discourse. The Tonic element in his TG is said to be the carrier of the information focus. He believes that the Information Unit is used as a process of interaction between what is already known or predictable and what is unknown and unpredictable. Thus the Information Unit is a structure made up of two functions, the New and the Given. The New is obligatory and the Given is optional. The Given typically precedes the New.

3.5 **// now silver needs to have LOVE //**

Given---- ----------New

Typically, the Tonic element will be the last functional element of the unit. Hence the typical order

Given followed by New

But whereas the end of the New element is marked by tonic prominence, there is nothing to mark where it begins, which sometimes causes some indeterminacy in the structure. This will be discussed in some detail in CHAPTER SIX below.

The Given/New functions displayed by tone units have been expounded by Brazil et al (1980) into what they call **'referring'** and **'proclaiming'** tones respectively. The referring tones express the 'Given' function in the tone unit and the 'proclaiming' tones express the 'New', as apparent from the following quote:

Aston University

Content has been removed for copyright reasons

Besides using the Information Unit, Halliday uses the 'clause' as a unit to discuss thematic structure, following the Prague School of linguistics. He divides the clause into two major parts, the **Theme** and the **Rheme**, which when mapped onto each other, make up the clause as a message. The **Theme** is what the speaker chooses to talk about as his or her point of departure of the message. The remainder of the message, the part in which the Theme is developed is the **Rheme**. Halliday (1985a) explains that there is close semantic relationship between the Given/New dichotomy of the Information Unit and the Theme/Rheme structure of the clause. A speaker will choose the Theme from what is Given and locate the focus, the climax of the new somewhere within the Rheme.

3.6

Now silver	needs to have	LOVE
Theme	Rheme	
GIVEN		NEW

This shows that Theme and Information together constitute the internal resources for structuring the clause as a message. And in order for a sequence of units, whether taken as Information Units or Clauses, to constitute a text it is necessary to do more than give an appropriate internal structure to each. It is necessary to study the relationship between one unit and another and between one clause and clause complex and another. This is what Halliday refers to as **"Cohesion"**. On nearly similar grounds Grimes (1975) uses his **"Information Blocks"** as a point of departure to study cohesion in discourse.

Another major area of research in which the above-discussed units have been used is in the description and quantification of text type differentiation such as the differences between spoken and written discourse and between mode variations. For example, Kroll (1977) and Farag (1986) use the IU to investigate the syntactic structure of spoken and written narratives. O'Donnell (1974), uses the T-Unit to investigate one individual's written and spoken output. His investigation is mainly made to validate Hunt's T-Unit as the most useful measure. Hasan (1988) uses the T-Unit to measure the differences between native speakers and nonnative speaker's classroom interaction. The Berkeley group represented by Chafe, Tannen, and others use the IU in their investigations of the narratives in the **'Pear Stories'**. Chafe's **'intonational sentence'** has been used to investigate the syntactic differences between spoken and written narratives (e.g. Beaman 1984). It has also been used in studying pausing and syntactic complexity in conversation and interview (see Zora 1986).

3.2 Choosing a unit of analysis for Speech Exchange Systems

Virtually none of the approaches discussed above satisfactorily applies to SES if the aim of the analysis is to investigate in depth the type of complexities of discourse units on the lexical, syntactic and informational levels of analysis. Obviously all spoken discourse can be chunked into prosodic units having prominent syllables as carriers of informational weight which speakers want to convey to their listeners. However, in addition to the fact that prosodic unit boundaries in spoken discourse, particularly SES, are sometimes very difficult to identify (See Halliday 1970; Brazil et al. 1980; Brown et al 1980), there is a tendency for these prosodic units, as has been shown earlier, to be incompatible with the syntactic units of speech, which would make it difficult to use them as units of analysis for syntactic purposes.

As has already been argued in Section 3.1.3.2, using the grammatical notion of 'sentence' as a unit of analysis for SES is practically impossible (cf. Svartvik 1982). In addition, Kroll's (1977) IU and O'Donnell's (1974) (T-Unit), though mainly based on syntactic criteria, have also got their own limitations when applied to such spoken discourse types as SES. They, first of all, fail to account for a great number of conversational utterances such as discourse markers, tag questions, fragmentary and truncated units and so on. So any approach based on these units would have to discard these important conversational features from all counts and descriptive accounts, the result of which would be to treat SES as any piece of written text. Secondly, they can be more suitably applied to written and spoken discourse types which have well formed consecutive clauses in which clause boundaries can easily be depicted. There are too many complexities in SES to be handled by analyses using one particular grammatical unit such as the clause, the IU or the T-unit.

Chafe's IU, although not without limitations, probably offers the most

suitable basis for analysis into organisational units. What is practically appealing about his IU is 1) the criteria are a combination of syntactic, semantic and functional levels; 2) IUs can combine together to form larger 'intonational sentences' (see Section 3.1.3.3 above) which are in his terms, equivalent to a grammatical sentence. But, although prosody can be helpful in determining the end of these sentences, it is not regarded by Chafe as crucial; and in the present study, prosody is seen as important but it will not be used as a major criterion for the unitisation simply because of the discrepancies between syntactic and prosodic units referred to in Section 3.1.4 above.

Notice also that although Chafe's units are based on linguistic criteria, they may perhaps be especially suited for analyses involving cognitive rather than linguistic research work. This is confirmed by Chafe's (1980) summary of the functions of his IUs, which reads as follows :

1. Personal interaction between the speaker and his or her audience.
2. Processes of recall as such.
3. The recall of narratives as a series of introductions of characters and their engagement in states and events.
4. Evaluative comments.

It seems that the approach to segmenting spoken discourse data into units for linguistic research, one which can handle all types of spoken discourse including SES, should be based on a combination of criteria involving syntactic/semantic as well as functional criteria. Such a segmentation would then have the capacity of accounting for all the units which are recurrent in SES. It would also account for syntactic units as small as a word and as extended as a clause complex consisting of an independent clause and any other dependent clause(s) attached to it. It would give accounts of the greatly recurrent loosely coordinated units, expansion units such as exemplifications, elaborations, extensions, and the

like. All these factors, and indeed many others as well, have led us to adopt a syntactic/semantic-functional approach to the unitisation of spoken discourse, akin but not exactly similar to Chafe's 'Intonational Sentence' (see Section 3.2.1.1 above), which is going to be outlined below.

3.2.1 Free-Standing Syntactic Unit

The need to chunk SES of various types into well-defined units is inevitable for achieving linguistic analysis on the levels outlined above; and in spite of the difficulties encountered by the analyst in this concern, we notice that the language of SES can be amenable to unitisation once the criteria on which one depends and the problems related to them are well defined in advance. In this research, a unit of information for the analysis of data is distinguished, which, for simplicity of reference, is going to be called Free-Standing Syntactic Unit (FSSU). This is defined as **a block of syntactically independent spoken information, semantically coherent and serving a particular function within the stream of running speech.** It can be regarded as chunking at the syntactic level, representing utterances as small as *'mmhmm'*, *'yes'*, *'no'* etc. or as long and complex as combinations of phrases, clauses and clause complexes. From the semantic point of view each FSSU is a coherent block of information but one which might have phoric(7) relations with preceding or following units. Prosodically, an FSSU does not represent any typical prosodic unit such as the ones discussed in 3.1.3.1 above. An FSSU can sometimes be equivalent to a TU but a long FSSU involving a combination of clauses and phrases can have more than one TU. It could prove an interesting research work to study the FSSU in relation to prosody and TU segmentation.

As an example of FSSU segmentation and of how to distinguish units in the manner of classification adopted in the analysis the following exchange of turns between the interviewer (JA) and two interviewees, KW and HC, from INT I is presented. The main numbers refer to an FSSU

each. As such, the example contains 10 major FSSUs. The subnumbers refer to what is going to be referred to as 'Minor Units' (see discussion below) :

3.7 (INT I)

JA. 1. 1 right..
2 --..--..
2. 1 you've.. you've both lived on campus for a year now..
2 e:::.. w.. w.. what would you see as the main difficulties..
3 for postgraduates..
4 of of living.. on a campus like this..

KW. 3. I think it's the fact they have no privacy really..
4. 1 they have people wandering in and out their room..
2 cleaners for instance..
5. 1 and when you are asked to move
2 you're given.. absolutely no notice..
6. 1 it's as if.. the housekeeper has no idea
2 that when you're doing a research project
3 that you do actually..
4 'cos.. alter your days..
5 that you fit everything in around it..
7. 1 only when you're given a forty-eight hour notice to move
2 it's difficult to rearrange everything..

JA. 8. mmhmm...

HC. 9. 1 yes
2 I'd.. I'd agree with that..
10.1 I mean I'm coming at it-- ..
2 from a different angle than Kate's..
3 and I didn't live on campus on my first degree..
4 so I can't compare it with.. with the way ** she can..

The bulk of the analysis in the following chapters will concentrate on the study of FSSUs from the point of view of lexical, syntactic and referential structures. However, since a relatively great number of these units may be represented by long stretches of phrases, clauses, discourse markers, interruptions, dangling expressions and the like, it seems more convenient both for analysis and exemplification to talk about two types of

FSSUs : Major FSSUs and Minor FSSUs which can be considered as sub-units within the major units. The minor FSSUs are not of great significance for the analytical part of the research but they have been found greatly helpful in characterising and identifying units especially for the purpose of exemplification. Moreover, they can be used as a basis for a microanalysis of SES in future research. Following is an explanation of the two types of FSSUs and the criteria over which they are based.

3.2.1.1 Major Units

This is used to refer to those examples of FSSUs which stand as complete blocks having both syntactic and semantic independency. Semantically, an FSSU represents a block of information having a single discourse function, e.g. informative or inquisitive functions, which could be as short as one single word like 'yes' and 'no' or a long stretch of clause complexes. The following exchange of turns contains three FSSUs represented sequentially by a question (invitation to dinner), a short response (acceptance) and a comment (elaboration) :

3.8 (CON III)

JA. 17.1 can I offer you lunch..
2 just next door..

DD. 18. oh yes.. *(he he)*
19.1 I brought some sandwiches
2 but there hasn't been time..-- --

As it stands, these 'Major Units' seem to be similar in many ways to Chafe's (1980) 'Intonational Sentence' discussed in the previous sections in that they end, too, with a syntactic and intonational closure and are semantically coherent. However, they are different in many ways. First of all, unlike Chafe's 'Intonational Sentences' the FSSUs can be as small as

one word as in **UNIT 18 of e.g. 3.8** above. Secondly, Chafe's (op.cit.) main interest is not in those 'intonational sentences' as such but in the **"focuses of consciousness or Idea Units"** which these sentences may consist of. In the present analysis it is just the other way round. The main interest is in the full Major FSSU, which may consist of just one unit (e.g. **Unit 18 of e.g. 3.8** above), or may consist of a number of minor units (e.g. **Unit 19 of the same example**). Thirdly, and perhaps most importantly, Chafe is not very clear about those units of SES which are joined by a loose coordinator such as *'and'* or *'but'*. His intonational sentence can have up to about 9 IUs sometimes loosely joined by such coordinators. In the present classification, (see discussion below), a distinction is made between the use of these coordinators in their 'strong' form of coordination (see Quirk et al 1985 and also discussion in CHAPTER FIVE below) and between their 'weak' form of coordination. Accordingly, the former occurs within the FSSUs whereas the latter always introduce these FSSUs, as will be exemplified below.

Syntactically, 'Major Units' have been found to be represented by one or more of the following criteria :

1. **Full finite clauses** (i.e. Subject + Verb) together with any other simple group(s) (nominal groups, prepositional groups, adverbial groups and so on) when present (e.g. **Unit 2. e.g. 3.9** and also **Unit 16 of e.g. 3.9** below).

2. **Full non-finite clauses** together with any other group(s) when present. (e.g. **Unit 4. of e.g. 3.9** below).

3. **Clause complexes**(8) involving :

a. Full independent clause(s) with one or more dependent (subordinate) clauses or phrases(i.e. Clauses in hypotactic relations in Halliday's (1985a) terms. (e.g. **Units 1., 11, 13, 17 and 19 of e.g. 3.9** below).

b. Two or more independent clauses coordinated by *'and', 'but', 'or'* etc. displaying strong cases of coordination (See Quirk et al. 1985). (having paratactic relations in Halliday's (1985a) terms). Omission of these coordinators would cause some sort of semantic or syntactic anomaly. (Compare with 4. below) (e.g. **Unit 19.** of **e.g. 3.8** above).

4. **Loosely coordinated units**(9) beginning with *'and', 'but', 'so'* etc. but mainly functioning as weak coordinators used only for continuation, initiation, summation and so on. Pragmatic 'and' (see Farag 1986) and the use of these coordinators as fillers would fall under this category. (See e.g. 3.10 below).

The difference between this category and 3.b above is that omission of these weak coordinators would not cause any semantic or syntactic anomaly. However, their presence adds to the coherence of connected speech. They are normally prolonged and they often occur with filled or unfilled pauses.

5. **Elliptical units** (phrases, clauses, clause complexes etc. **(e.g. Unit 3. of e.g. 3.9** below).

6. **Combination of sub-units** (Minor Units) (See B. below) which are semantically related, are taken to be one full FSSU(10). (e.g. **Unit 13. of e.g. 3.9** below).

The following example from CON IV contains most of the categories mentioned above. The numbers followed by a dot represent a major unit which when containing subunits, the latter will be represented by the subnumbers without a dot.

3.9 (CON IV)

JA. 1. I am not quite sure what the research is..

RF. 2. 1 mmhmm..
2 so you're being researched as well *(he he he)*...

JA. 3. 1 well..
2 I'd suppose..

RF. 4. put like that.. *(he he)*..

JA. 5. 1 yes..
2 yes...
6. 1 e:::m.. you're in the Monday groups
2 aren't you..

PH. 7. yes..

RF. 8.1 yes..
2 yes..
9. 1 you must be on the other group..
2 I'm with Teddy..

PH. 10. mmhmm..

JA. 11. 1 yes..
2 I've just been doing the:: e:::m thing for [tonight..
RF. 12. [mmhmm

3 which is looking on the evil and suffering..
13. 1 it seems to me that that's.. e::.. the thing
2 which we get asked..
3 well..
4 we're.. we're liked to get asked next term
5 to.. sort of push the relevance of Christianity..
6 to.. to.. the real world
7 and real problems
8 and so on..

PH. 14. how did your group.. group go

JA. 15. 1 I was very pleased really...
2 so e:::m-- --..
16. we haven't got a very wide spectrum
17. 1 I think in both the other groups
2 we had a wide spectrum of.. Catholics and

3 well..
4 non-Catholics and [..
PH. 18.1 [and well
2 doubtfuls [..
JA. 5 [doubtfuls [..
RF. 19.1 [*(he he he)*.. well
2 I think in the group that Teddy's leading
3 there's definitely..
4 e::m I mean there's an even balance
5 between Catholics.. and non-Catholics..
6 if I may phrase it that way
7 because.. there's some.. Anglicans..
8 there's some Church of England
9 and.. e::
10 or one or two of each
11 and there's some I think who's.. undecided [..
JA. 20. [mmhmm
12 and there's some -- like myself..
13 I've been to many different churches
14 so::..

3.2.1.2 Minor Units

They are those small recurrent units of discourse which are essentially bound : semantically and/or functionally bound to adjacent discourse without which they are uninterpretable. They are mainly constituents of Major FSSUs or peripherally attached to Major units on semantic and functional grounds or can combine to form one FSSU (in which case they are considered as sub-units and are thus sub-numbered accordingly).
Minor Units can be one of the following categories :

1. <u>Discourse Markers</u> such as : Oh, well, now, now then, alright, O.K. etc. (as in **Unit 2.1 of e.g. 3.10 and Unit 1.1 of e.g. 3.11** below).

2. <u>Back-channels</u> such as mmhmm, yeah, yes , no, yes and no, and other **'listener responses'** (as in **Unit 12.1 of e.g. 3.10 and Unit 15.1** below).

3. <u>Vocatives</u> (as in **Unit 5.1 of e.g. 3.10 below)**

4. Tag questions (as in **Unit 6.2 of e.g. 3.9 above)**

5. Elliptical units (as in **Unit 2.2 of e.g. 3.10 below).**

6. Units (phrases, clauses etc.) which occur attached to FSSUs, expressing one of the following functions:

a. Expansion (elaboration, enhancement, extension, exemplification, particularisation and so on (as in **Unit 7.2 and Unit 8.2 of e.g. 3.10** below).

b. Summation. (as in **Unit 18.4 of e.g. 3.10** below)

c. Structural revision, reformulation and self-correction (as in **Unit 11.2 of e.g. 3.10** below).

3.2.1.3 Examples from the data

The following example from INT VIII contains most of the categories above :

3.10 (INT VIII)

JA. 1. e:::m.. what's your impression of your first year at Aston..

AB. 2. 1 oh..
2 not bad..
3. e:::m.. it's hard to think of everything ** I've said before..
4. 1 e::m:.. e:::.. no
2 I'm stuck *(he he he)*..

JA. 5. 1 Kath..
2 has.. has it been a good first year..

KS. 6. e:::m... yes and no..
7. 1 I've.. made friends a lot quicker than I thought I

would..
2 I've made some very close friends..
8. 1 a:nd the work has been very very different..
2 e::m a different way of doing it..
9. 1 a::nd I think I found that hard to cope with
2 to a certain extent
3 to make myself do the work..

JA. 10. so.. the work has been different in what way..

KS. 11.1 e:::m.. very much in school there were lots of essays to..
2 I mean even though we weren't pushed necessarily to a specific date for them..
3 there were always..essays to be handed in..
4 whereas I've only.. had
5 I've had to do very little actual coursework this year..

JA. 12.1 mmhmm..
2 is that your experience as well
3 Andy..

AB. 13.1 yeah
2 the::..
3 hardly having anything to hand in at all
4 or anything to do for a specific date [..
JA. 14. [mmhmm

15. e::m..it's been very different to what I've been.. used to..
16.1 although I had extreme flexibility while I was at school
2 but it still had to be done at some point [..
JA. 17. [mmhmm

18. 1 e:::m.. and there is not.. that.. pressure there
2 to.. to get it done
3 other than.. a dozen things that you had to hand in
4 and that was it..

3.11 (CON VI)

JA. 1. 1 e::m.. yes..
2 I'm sorry that e::m I haven't got a.. glass or something
3 for all that alcohol.. [*(he he he)*
AM. & CB. [*(he he he)*-- --

2. 1 I didn't organise that well enough..
2 I'm sorry about that..

CB. 3. 1 well
2 let you off *(he he he)*

AM. 4. just this once *(he he)*

JA. 5. 1 oh..
2 are you fairly idle at the moment..
6. 1 well
2 no no
3 I know you're not really idle..

CB. 7. no..

AM. 8. never.. --
9. 1 I was last week..
2 we've started now..
3 sort of every..
4 all the staff are back..
5 so it's time.. time to work..

CB. 10. yea::h..

AM. 11. 1 so instead of arguing with the students
2 we argue with the staff as well [..
CB. 12. [*(he he)* yes..

AM. 3 a great situation [to be in..
JA. 13.1 [is that the truth
2 or is it what you're presenting yourself out to be..

AM. 14.1 well
2 I'm not in the right..
3 they just can't all understand it [*(he he)*..
CB. [*(he he he)*

AM. 15. 1 yeah..
2 I don't know what I'm doing wrong
3 but everyone seems to.. to have totally different ideas to me..
4 about.. what to put on the bars..
5 and.. what our aims should be..

3.3 Quantification of data : Words , Turns and Units

Out of the nine interview and conversation sessions, eight were chosen for the present investigation and for conducting an overall statistical counts of the lexical, syntactical and referential features of the linguistic turns exchanged in these sessions. The ninth session was omitted because the conversation part was very short and therefore unsuitable for comparison with the interview part of the same speakers. In the general calculations the interviewer input was included except when otherwise stated. The following table shows the total number of words for all eight pairs of subjects (including the interviewer's) in INT and CON :

Sessions	INT	CON
I	1323	3979
II	1662	2143
III	1812	1883
IV	1777	2506
V	2052	1046
VI	2897	2501
VII	1769	2246
VIII	1862	2897
TOTAL	**15154**	**19201**

Table 3.1 Total number of words in INT and CON

The figures above represent the total number of words for the group of speakers within each session of INT and CON including the interviewer's speech. The following table gives details of the number of words for each of the sixteen subjects and the interviewer (JA) as well.

SESSION	SUBJECT	INT	CON
I	JA	311	842
	HC	437	1009
	KW	575	2128
II	JA	421	461
	JH	557	958
	BG	684	724
III	JA	208	646
	RH	636	675
	DD	968	562
IV	JA	449	908
	RF	492	1402
	PM	836	196
V	JA	221	457
	HK	760	296
	KSH	1071	293
VI	JA	413	100
	AM	1026	1525
	CB	1458	876
VII	JA	375	9
	JC	571	1007
	GM	823	1230
VIII	JA	324	127
	AB	742	1279
	KS	796	1491
	TOTAL	**15154**	**19201**

Table 3.2 Number of words for each subject in INT and CON

Two points about the above figures should be borne in mind. First, all the interviewer's figures have been excluded from the calculations in the analysis chapters first of all because he did not participate actively in the conversation encounters and most of the time he was not present during the conversations and also because it is the language of the interviewees as opposed to their chat in conversation that is being compared here. Secondly, Since the number of words in the INTs and CONs is not uniform throughout the sessions, and drawing on Beaman (1984:53), a simple frequency index will be used to measure the particular features of interest in the two variations under study. This will be referred to as the **'Occurrence Per Thousand Words'**, (henceforth OPTWs), which is worked out by dividing the total occurrences of the feature (e.g. word, structure,

turn, unit, etc.) by the total number of words in INT and CON and then multiplying the outcome by 1000. In certain calculations, and for the sake of comparison with other researchers' results, a percentage for those features was favoured.

The following table shows the number of turns exchanged by the Subjects in INT and CON. The figures represent the total number of turns for all speakers and also their OPTWs :

	RAW FIGURES		OPTWs	
SESSION	INT	CON	INT	CON
I	70	205	4.6	10.7
II	49	113	3.2	5.9
III	52	84	3.4	4.4
IV	65	142	4.3	7.4
V	51	37	2.4	1.9
VI	110	144	7.3	7.5
VII	43	150	2.8	7.8
VIII	67	240	4.4	12.5
TOTAL	**507**	**1115**	**33.5**	**58.1**

Table 3.3 Number of turns and their OPTWs in INT and CON

The figures above show that there are significantly more turns in CON than in INT. On average the calculations show that there are 58.1 OPTWs of turns in CON whereas there are just 33.5 OPTWs in INT. Table 3.4 below shows the number of Major Units and Minor Units in INT and CON and Table 3.5 their OPTWs :

SESSIONS	INT		CON	
	MAJOR	MINOR	MAJOR	MINOR
I	107	134	316	772
II	100	279	159	426
III	100	281	124	395
IV	102	312	205	555
V	112	351	64	185
VI	181	538	208	498
VII	84	303	188	435
VIII	127	313	280	595
TOTAL	**913**	**2511**	**1544**	**3861**

Table 3.4 Major and Minor Units in INT and CON (Raw Figures)

SESSIONS	INT		CON	
	MAJOR	MINOR	MAJOR	MINOR
I	7.1	8.8	16.5	40.2
II	6.6	18.4	8.3	22.2
III	6.6	18.5	6.5	20.6
IV	6.7	20.6	10.7	28.9
V	7.4	23.2	3.3	9.6
VI	11.9	35.5	10.8	25.9
VII	5.5	20.0	9.8	22.7
VIII	8.4	20.7	14.6	31.0
TOTAL	**60.2**	**165.7**	**80.4**	**201.1**

Table 3.5 OPTWs of Major and Minor Units in INT and CON

The figures above show clearly that OPTWs of Major and Minor Units within the data is significantly more in CON than in INT. These results, along with the results that there are more words and turns in CON than INT (Tables 3.2 and 3.3) show that Major and Minor units in CON are shorter than those in INT and are less lexically dense (see CHAPTER FOUR below).

In order to find out the complexity of the Major Units in the data of INT and CON in terms of the size of these units, another type of

calculation has been developed, in which the distribution of these units was counted in the eight encounters of INT and CON as one-word units, one-clause units, two-clause units, multi-clause units and incomplete (elliptical units). The results are displayed in OPTWs in Table 3.6 and 3.7 below :

	1-word Units	1-clause Units	2-clause Units	Multi-clause Units	Incomplete Units
I	0.1	5.6	11.2	25.3	4.2
II	21.2	1.2	5.4	32.1	0.6
III	15.8	4.4	4.9	27.2	1.1
IV	23.5	3.9	1.7	26.9	0.6
V	15.7	2.9	6.4	28.9	1.0
VI	20.2	5.2	5.6	25.0	3.8
VII	5.7	6.8	10.2	25.4	1.1
VIII	25.4	7.0	6.5	28.1	1.1
MEAN	**19.7**	**4.6**	**6.5**	**27.2**	**1.7**

Table 3.6 Distribution of Major Units in terms of words, clauses, clause complexes and ellipticals in INT

	1-word Units	1-clause Units	2-clause Units	Multi-clause Units	Incomplete Units
I	20.3	13.3	9.5	24.1	11.1
II	26.2	10.3	5.6	23.4	11.7
III	13.8	5.4	8.5	25.0	10.1
IV	25.8	7.9	7.9	21.5	9.1
V	14.8	3.0	3.0	27.6	6.9
VI	17.6	10.2	9.8	27.5	12.2
VII	14.2	16.9	15.6	23.2	15.1
VIII	19.3	11.3	9.2	24.6	11.8
MEAN	**19.0**	**11.3**	**9.2**	**24.6**	**11.8**

Table 3.7 Distribution of Major Units in terms of words, clauses, clause complexes and ellipticals in CON

These two tables give the first clues about the syntactic differences between INT and CON in terms of the nature and amount of talk. They also point to the presence of syntactic complexity in the two sets of data due to the relative high occurrence of multi-clausal units, incomplete

units and so on. However, there seems to be a considerable amount of variation between the interactants in the level of this complexity. This will be the major concern of the chapters that follow.

The figures and tables above represent the general calculus to which reference will often be made in the following three chapters as necessity arises. CHAPTER FOUR will deal with a statistical analysis of lexical features and lexical density and variability across speakers in INT and CON. Syntactic density and complexity will be the main concern of CHAPTER FIVE below. The final chapter of the statistical analysis, CHAPTER SIX will look at the linguistic behaviour of participants as far as discourse referentials are concerned relative to information density and complexity. The findings of the three chapters will be compared to those of other researchers in the field of spoken discourse analysis.

3.4 Conclusion

The various attempts at unitising spoken discourse in the literature were reviewed and assessed in the light of SES nature and behaviour. Because of the unique character of SES and the many features it entails in running discourse, it is believed that virtually none of the above approaches towards its unitisation is totally satisfactory. Each one puts emphasis on certain features of discourse but neglects others with the 'sentence' being the most unacceptable as a unit for the analysis of SES. A method of unitisation is suggested in this chapter, which takes almost all relevant characteristics of SES into consideration, based on both formal and functional criteria. The unit distinguished according to this approach has been termed **'Free-Standing Syntactic Unit'** since, in syntactic terms it can be equivalent to any grammatical unit (word, phrase, clause, clause complex) provided it is informationally and prosodically coherent. For further facilitating the process of analysis, longer FSSUs were divided into 'minor' units. Here I am taking into consideration such units of discourse

as 'markers', 'vocatives', 'ellipticals', backchannels' and so on. This method of unitisation is thought to be satisfactory for the analysis of SES since it takes into consideration syntactic, semantic and prosodic cues and such features of SES as discourse markers, backchannels, elliptical units and repetitive segments of discourse.

A calculus of the general features of the data of INT and CON, such as number of words, turns, FSSUs and so on has been provided in the chapter for reference. The initial results show a higher number of words used in CON than in INT (Table 3.1 and Table 3.2), a significantly higher occurrence of turns in CON than INT (Table 3.3), and also a considerably higher number of Major and Minor units in CON over INT (Tables 3.4 and 3.5 respectively). Tables 3.6 and 3.7 show the Multi-clause units to occur significantly more in INT than CON, but the 1-clause units to occur more in the conversation encounters. Incomplete units (ellipticals) occur significantly more in CON than in INT.

The results show that a good deal of variation between one speaker and another is possible in terms of the differences in their output in INT as opposed to that in CON. This will be investigated in depth on the levels of lexis, syntax and information structure in the following three chapters FOUR, FIVE and SIX respectively.

Footnotes to CHAPTER THREE

1. The transcript represents a highly organised interview-like conversation. It does not offer explicit information regarding the presence or absence of overlap and interruption. However, the layout leads the reader to interpret the interaction as being markedly "sequential".

2. Both narratives (a) and (b) are from Chafe's (1980) 'Pear Narratives'. Beaman (1984) uses the same data for her investigation of subordination and coordination in spoken and written narratives.

3. BBC TV interviewer Jonathan Dimbleby interviews MP Michael Howard, Minister of Water and Planning. The broadcast was videotaped and transcribed by the present author in October (1988) for the present investigation.

4. The term 'prosodic' is used here in Crystal and Davy's (1969) sense to refer to such features as intonation, pitch, stress, pausing, tempo, relative variation, voice quality and so on (see ibid : 24-37).

5. Halliday (1967, 1970, 1985a) takes as the unmarked case the occurrence of one **'Information Unit'** per clause with the Tonic falling on the last lexical item. The term 'Information Unit' Halliday uses to denote a unit of discourse which is prosodically realised as a **'Tone Group'**. Grimes (1975:277) makes similar suggestions concerning his 'Information Blocks' (see Footnote 6 below).

6. Grimes (1975) uses the term **"Information Blocks"** to a unit of discourse prosodically realised by almost the same cues as Halliday's **"Tone Group"**.

7. See Halliday and Hasan (1976) for a detailed explanation of these terms.

8. The term **"Clause Complex"** is used all throughout the thesis in the sense of Halliday (1985a) , to refer to **"a head clause together with other clauses that modify it"** (p. 192)

9. The term **"loosely coordinated"** is borrowed from Crystal and Davy (1969).

10. These units could well be any of the other five categories 1., 2., 3.a and 3.b , 5 and 6 , initiated by one of the conjunctions *and* , *but* , *so* etc., functioning as weak coordinators (cf. 3.b above).

CHAPTER FOUR

LEXICAL DENSITY IN INTERVIEW AND CONVERSATION

4.0 Introduction

As stated in CHAPTER ONE above, linguistic complexity in SES can be investigated on different levels, most important of which for the purposes of this research are those of LEXIS, SYNTAX and INFORMATION STRUCTURE. Syntax and Information structure will be dealt with in CHAPTERS FIVE and SIX respectively. The present chapter will deal with the issue of complexity in SES on the lexical level. Of course, the notion of lexical complexity is too general and can cover all sorts of matters related to both syntagmatic and paradigmatic relations (see definitional comments below). From within this multiplicity of features, the notion of Lexical Density (LD) has been chosen as the main focus for the present chapter as it has proven to be a suitable analytical tool for text-type differentiation, on the one hand, and because of its ability to touch upon such matters of linguistic complexity in discourse as lexicality versus grammaticality, information packaging and information processing and so on, on the other.

LD as a technique for investigating the frequency and type of information packaging in discourse involves the use of statistical packages to examine the frequency of lexical (L) as opposed to non-lexical (ie grammatical) (G) information in the data under investigation. In the present chapter, variation in LD is examined globally in the two types of spoken output, INT and CON. Secondly, it is then examined at the inter-individual level. Thirdly, variation in LD is examined in the context of differences between contrasting contiguous styles in spoken discourse. The results are discussed in relation to findings of other variationist studies of LD in the literature (cf. Ure 1971; Stubbs 1986).

A considerable amount of research has been done within the area of variability in LD within the last two decades or so (cf. Ure op.cit.; Stubbs op.cit.; Farrugia 1987; Hasan 1988). However, it has not hitherto been examined in relation to information structure, and little attention has been paid to inter-individual differences at the lexical level. Part of this chapter will be devoted to this issue as a first step to the more specific investigation of the structure of information in SES (see CHAPTER SIX below).

4.1 Lexical studies in the literature

4.1.1 Lexical complexity and lexical density

The term 'lexical complexity' is of course generally used to cover a multiplicity of issues which are mostly related to the ways users of language use their lexical items to express meaningful messages in communicative situations. Some of these issues include syntagmatic relations, collocability, lexical predictability, span, besides some other lexical matters which are beyond the present investigation (cf. Sinclair 1966). Another important issue of lexical complexity, which is more related to paradigmatic relations and form, can be the frequency of lexical item use, probability of occurrence of lexical items and also the choice of a particular item over other similar items. For various reasons and under different circumstances speakers might vary the use of lexical items when conveying information sometimes greatly in accordance with the particular situation they happen to be in, thus varying the degree of complexity of these items. This might well be style specific and can be the result of previous education, family background, personal interests and linguistic training. Complexity of this type can be statistically tested by such means as examining the items used within discourse and then separating those items that are thought to be lexical from those that are grammatical

(see definitional comments below). The ratio of L items to G items is known in linguistic research as **'lexical density', "the density with which the information is presented"** (Halliday 1985b:62). Many researchers (e.g. Ure 1971; Halliday op.cit.; Stubbs 1986) relate text complexity to the amount of LD present in a spoken or written text (see discussion below). Accordingly, LD has been adopted in this chapter as an analytical tool for the differentiation between INT and CON data, on the one hand, and on the inter-individual level, on the other. The major aim is to establish the level of lexical complexity in the two types of SES under examination and to see how far people differ in their employment of L items as opposed to G items in their different speech styles. The technique of LD, and the various problems associated with its application will be outlined below after examining in some detail the types and importance of lexical studies in the relevant literature.

The first assumption here is that there is a close relationship between the lexical complexity of a particular type of discourse and its LD both in relation to the whole body of discourse and to its smaller units of information which constitute that body. Intuitively, certain types of discourse have higher LD than others. This will be checked against the data of INT and CON as we proceed in this chapter. If it is going to be proved, it would be in itself an indication of the 'complexity' of the data under investigation on the lexical level. There is already some evidence in the literature to confirm these intuitive assumptions. Some researchers (e.g. Halliday 1979, Beaman 1984, Kroll 1977) have shown both theoretically and empirically that written discourse has more linguistic complexity than spoken discourse because it is more lexically dense. Lexically dense texts contain perhaps a multiplicity of complex verb phrases, noun phrases with pre- and/or postmodifications, qualifying attributive and predicative adjectives and modifying adverbs of different sorts. These categories constitute the substance of the complex phrases and clauses within a text or discourse, which in turn shows the close

relationship between lexical complexity and syntactic complexity (see CHAPTER FIVE below).

Halliday (1979:49) draws a clear distinction between lexical density and syntactic complexity when talking about these features in spoken and written discourse. He points out that, linguistically speaking :

> "spoken language is characterized by complex sentence structures with low lexical density (more clauses, but fewer high content words per clause) ; written language by simple sentence structures with high lexical density (more high content words per clause, but fewer clauses).

The part of Halliday's claims which is related to syntactic complexity in discourse will be discussed in CHAPTER FIVE below. In this chapter, however, his claims concerning the lexical complexity of spoken discourse will be checked against the data under investigation.

4.1.2 Importance of lexical studies

The importance of studying lexis lies in the fact that lexical items and lexical relations represent one of the interpenetrating ways, besides grammar, of looking at language form (cf. Sinclair 1966:411), which is in itself a crucial issue for the understanding of the other major function of language : **meaning**. Lexis involves general tendencies of use of items and their ability to collocate with habitual lexical settings to produce stretches of discourse.

Firth (1951) drew attention to the importance of studying lexis in linguistic research. He actually made a distinction between lexis and semantics showing that it is both possible and useful to make formal statements about lexical items and their relations. He distinguishes between **'meaning by collocation'**, which is an abstraction at the

syntagmatic level (see discussion below), which **"is not directly concerned with the conceptual or ideal approach to the meaning of words"** (Firth 1951:196), and **"contextual meaning"** which is **"the functional relation of the sentence to the processes of a context of situation in the context of culture"** (ibid:195)(1). Within the first of these distinctions, Firth introduced his notion of the **'collocational level'** within which all formal analyses of lexis and lexical studies fall (see, for example, the collection of papers in Bazell et al. 1966).

Crucial to the understanding of the role played by 'lexis' and 'lexical relations' in language interpretation is the important distinction first made by de Saussure in 1916 between **syntagmatic and paradigmatic relations**, the opposition between which gives a better understanding of the class of values each is able to generate (cf. ibid, 1959:122).

Syntagmatic relations are those which are supported by linearity in discourse, i.e. relations which specify the combination of elements into complex forms and units (e.g. sentences). Paradigmatic relations, (or what de Saussure (op.cit.:123) calls **"Associative Relations"** as they are **"associated in the memory"** (ibid), on the other hand, specify relations outside discourse, of units (e.g. words) which have something in common or which enter into contrastive relations i.e. they are substitutable in discourse. For example, all the words in the sentence **'John teaches French'** enter into syntagmatic relations with each other to determine the whole meaning of the sentence, but taken individually each of the three words enters into paradigmatic relations with other similar words (which are not in the same sentence), for example **'John'** to **'he', 'somebody', 'my friend'**, etc. The importance of these relations to lexical analysis (and indeed to any grammatical analysis as well) is beyond doubt great as they determine both the content and expression of the signs of language and their use and organisation in discourse. Halliday (1966:152) confirms this by pointing out that :

> "Clearly lexical patterns are referable in the first place to the two basic axes, the syntagmatic and the paradigmatic."

It is clear then that lexical items do not merely occur in sequences at random within discourse but they display highly organised syntagmatic relations within which semantic ties are coherently built up in a linear manner. This syntagmatic relation is referred to as **'collocation'** (cf. ibid:152). Each lexical item used within running discourse has a number of collocational items which occur before or after it and which give it its coherent character. The importance of any lexical item is determined by a number of factors most important of which are related to the degree and length of proximity between the 'collocated' items, collocational restrictions, the predictive power of lexical items and so on (cf. ibid:156).

Syntagmatic/paradigmatic relations are not just confined to lexis and lexical relations, but they also include grammatical relations between discourse constituents (cf. Halliday op.cit.; Bierwisch 1971). Notice the following quote from Halliday (op.cit.: 152) :

> "One way of handling grammatical relations on these two axes is by reference to the theoretical categories of 'structure' and 'system' with the 'class' definable as that which enters into the relations so defined. In lexis these concepts need to be modified , and distinct categories are needed for which therefore different terms are desirable."

In this chapter, the aim is not to go into the problem of semantics and semantic analysis, neither is it to give details about the issue of the syntagmatic and paradigmatic relations between constituents within units of discourse, important and interesting though these issues are in the analysis of discourse structure. Hopefully, they will be dealt with in subsequent research. As has already been mentioned, the focus of this chapter is to examine lexical variability on the simple level of LD in two

types of SES under investigation, INT and CON. The following sections involve an elaboration on the notion of LD, the main problems associated with its applications and the major studies in the literature that have used it for text type variation.

4.2 Lexical Density

4.2.1 Definition , technique and methodologies

The notion of LD has been used in the literature to refer to the proportion of lexical words expressed as a percentage of text. This entails an overall calculation of all the words in running texts and then calculating those words that are taken as lexical and/or those that are taken to be grammatical. Then a statistical formula, such as the one cited below (adopted from Stubbs 1986:33), is applied to get the final results :

If the total number of words in the text = T ,

and the number of lexical words = L ,

Then LD = 100L/T

Or else, one can apply the same sort of formula to find out the percentage of all G words from a total number of words in a text and then arriving at the percentage of LD by deducting the results from 100.

Two types of operating methodologies have been adopted by researchers to arrive at the ratio of G : L in the literature of spoken and written discourse. The first one is manual (cf. Ure 1971(2), Hassan 1988) and the second is mechanical and depends mainly on computer programs like the one written by Stubbs (1986), which has been designed to run on the London-Lund corpus of spoken English. The programs used by the

present researcher were written by Knowles (1988) to be used as utility programs for lexical and grammatical word calculations. The same programs were used by Farrugia (1988) for calculating LD in Newspaper Editorials. The manual approach has a greater degree of sensitivity to context since each problem is dealt with by the human linguist in its real context. However, the amount of text processable is presumably limited and humans are notoriously prone to non-systematic variation. Automatic analysis based on tailor-made software, though efficient and reliable to a great extent, especially in handling long types of texts, suffers from inability to take account of context in the wider sense - wider, that is than the term collocation implies. These problems will be discussed below.

4.2.2 Some theoretical and practical problems

4.2.2.1 Lexical items and words

Crucially important for any quantitative research on lexis and lexical frequencies is the distinction that has to be made between 'lexical items', and 'words', and the ambiguity arising from the use and interpretation of these terms. Obviously, lexical items are constituents of variable length because they may consist of more than one word (Halliday 1985b:63) : e.g. **stand up, take over,** which function as single lexical items. They cover a wide range of word classes, idioms, phrasal verbs, compound words etc. (See Section 4.2.2 below). A lexical item does not normally stand in a one-to-one relation with 'word' as used in linguistic literature. There is too much controversy over the use and definition of 'word' in linguistic literature. It is generally accepted that a word is a semantic unit which is capable of occurring alone (Bloomfield's (1933:178) **"minimum free form"**). But as Palmer (1976) and Lyons (1977a) have pointed out, there are **"semantic wholes"** which are nevertheless written as two words (e.g. **She passed away = died)**. Phrasal verbs are the most often cited examples. Bloomfield's definition of a word is of course flawed since ability to occur

alone fits awkwardly with certain grammatical forms **('a', 'the', 'to')**. The fact is there is no theoretically acceptable definition of 'word'. The analysis of LD is then forced to adopt the most primitive position namely to rely on conventions for written language in order to define a starting point. In company with others who have carried out analysis of lexical frequency/ textual density, the writer opts to treat the 'word' for practical purposes as a unit **"being written with a space on either side"** (Halliday 1985b:64). This definition is not theoretically satisfactory, though it is pragmatically expedient. As far as this analysis is concerned, the notion 'word' is used in this latter sense, i.e. in its simplest orthographic form. Many researchers have found this treatment of 'word' necessary for quantitative research. Stubbs (1986), for instance, confirms this from the computer's point of view :

> "the simplest definition is the orthographic text word form, since it is simple to get a computer to recognize a string of letters with a space or various punctuation marks on each side. However, this definition obviously has several limitations from a linguistic point of view.."

To avoid terminological confusion and for the purpose of conducting straightforward quantitative analyses which would be comparable to other studies in the literature, and since the concern here is not to study lexical items in relation to lexicality and collocability, it seems plausible to avoid the term 'item' in the sense generally used in the expression 'lexical item' altogether and use the term 'word' instead. The reason for this decision will become clearer as we proceed forward in the discussion.

4.2.2.2 Lexical Vs Grammatical Words

The second important distinction which needs further elaboration is that often made in English and indeed in many other languages between lexical (L) and grammatical (G) items. This distinction is directly related to

that made between 'open' and 'closed' word classes, an issue which has often been tackled in the seminal work of grammarians and semanticists. For Robins (1964), Lyons (1968) and Palmer (1971,1976) lexical items are the major content words which enter into open contrasts, therefore they belong to 'open word classes'. They refer to concepts, objects or persons and thus fall under the four grammatical categories: Nouns, Adjectives, Adverbs and Main Verbs. Grammatical items, on the other hand, are those that serve to express relations between content words and as such they are :

> "responsible for the syntactic coherence of phrases and sentences" (Opitz 1983:58)

These represent a number of 'closed class' items and include : Auxiliary Verbs, Modals , Pronouns, Prepositions, Determiners and Conjunctions. A major difference between open and closed class elements is their numbers. Open class elements are almost limitless in number, whereas closed class elements form a relatively small set.

The major difficulty with this sort of classification especially for a quantitative characterisation of L and G items in discourse is that, on specific words in a given context, the distinction is not always easy to make. Most researchers who have dealt with lexical quantification and LD in spoken and written discourse have expressed their concern about this problem. The general belief is that there is no clear line of demarcation between the two classes of words. Halliday (1985b:63) points out that :

> "there is a continuum from lexis into grammar : while many items are clearly of one kind or another, there are always likely to be intermediate cases."

In another place, Halliday (1966:155) actually postulates the existence of a scale on which items could be ranged from 'most grammatical' to 'most lexical', the former (i.e. most grammatical, e.g. **'a'**, **'of'**) being one :

"which is optimally specifiable grammatically. This can be thought of as 'reducible to one- member class for the minimum number of steps in delicacy."

One important point of distinction between L and G items is that of collocability which characterises lexical items in general but not grammatical items, the thing which gives them more predictive power over their occurrence within environments (cf. Halliday ibid:156). Another point of distinction is related to stress. L words always receive sentence stress (tonic placement) whereas G words do not receive such stress except when the speaker wants to achieve a certain **"meaningful choice"** (cf. Brazil et al. 1980:39). For example, in the following 'Tone Unit' the underlined words are all L items and are thus prominent by virtue of their **"function as a constituent of a tone unit"** (ibid) :

4.1 (from Brazil et al. (ibid))

Tom is the best boy in the class.

The rest of words in the Tone Unit are G words, hence unstressed. However, a speaker may choose to stress a certain G item, such as 'is' in the same example above to create a deliberate meaningful emphasis, as follows :

4.2 Tom is the best boy in the class.

4.2.2.3 Other more 'practical' problems

As has been mentioned above, the technique of LD conditionally involves sorting words into L as opposed to G, which is not in itself an entirely easy and straightforward task. One will, first of all, have to go through the tedious job of counting all the words of running texts and

then separating L from G words so as to arrive at the final percentage of LD out of that, which too has its own problems and complications. For example, many prepositions and certain classes of adverbs (e.g. always, far, others) fall on the borderline between 'lexicality' and 'grammaticality' in the senses discussed in Section 4.2.2.2 above, and unless these words are studied in the context in which they occur, it is virtually impossible to refer them to one set or another. Accordingly, ad hoc decisions will have to be made to solve some of these problems, some of which might not be analytically satisfactory.

Another example of problems that might be found with LD calculations arises with the so-called phrasal verbs in which the status of the preposition or particle element is sometimes difficult to determine. For example in :

4.3 **1. She made up her face**

2. She made up her story

3. They made up and kissed

4. She made up the hill at speed

(Example 1. is from Halliday 1966:153)

the grammatical object in 1 is optional, but not so in 2 (as in **She made up swiftly)**. The '**up**' in 1 would therefore appear to be more of an adverbial particle than a preposition, and is therefore directly comparable with the particle status of '**up**' in 3. '**up**' in 4 is of course prepositional. The point is that :

5) They made up

is ambiguous as between meaning 1 and meaning 3. The implication is that the grammatical status of '**up**' can only be resolved by reference to the lexical context.

Stubbs (1986) proposes that for the purpose of computer processing of texts **'up'** in 1 above should simply be regarded as (a) non-lexical (i.e. grammatical) and as (b) a separate item. This position is analytically unsatisfactory, even if it is appealing in practical terms and can handle some of the problems mentioned above. In the present writer's view, any automatic analysis of L and G items must be either post-edited or manually prepared. However, for the sake of comparison and cross-checking of results, both Stubbs (ibid) and Ure (op.cit.) have been drawn on in the present research, in considering such phrasal verbs as separate items.

In addition to the problem of phrasal verbs, other types of problems can also arise, some of which have been reported by Stubbs (op.cit.), which no algorithm can resolve. One such problem is that some of the modal verbs such as **'can'** and **'will'** can also occur as main verbs or nouns in certain contexts. Auxiliary verbs such as **'be', 'have'** and **'do'** can also be G or L according to the grammatical contexts in which they are used. Stubbs solved such problems in his program by building into it a routine to deal with potentially ambiguous words which are categorised according to their context in running text.

There is, of course, a more general problem in word classification. What one researcher counts as lexical, another will classify as grammatical. Stubbs, for example, lists **'be'** as lexical or grammatical. Ure, on the other hand, counts it as grammatical even when it has a more lexical function as in **"If you don't be good"**.. (Personal Communication)(3). She also states that when there is indeterminacy in certain fuzzy terms, she has opted for grammar rather than lexis. For instance, she recognises two **'wells**, one with discourse function which she takes as G and one L related to **'good'**. She has also tended to consider as G those items which are **'collocationally uninteresting'** such as numerals.

Plausible as these points of view might be, they are not typically the same as what other researchers take 'grammatical' and 'lexical' items to mean when making frequency calculations of running spoken and written texts. It is believed that in any analysis of this sort, one has to have a distinct and clear-cut line as to which items should be taken as G or L, in accordance with the grammatical or functional contexts in which they are used. Stubbs has provided an overall list of what he takes as G words with the rest of the words in English being L, meanwhile keeping in mind those ambiguous words and phrases to be dealt with separately and leaving a minority of others to be allowed **"to pass as being statistically insignificant"** (ibid:42). This statement might well work with long texts but it can be dubious when applied to comparatively shorter texts such as the ones used in this study. It should be mentioned at this point that for the calculations made in this study, Stubbs's distinction of G and L words and his list of G words in English were adopted(4).

4.2.3 Previous research on LD in types of discourse

Previous research of LD has shown that the ratio of L to G items within a text can allow texts to be ranked in relation to each other. In very general terms, it shows how 'informationally loaded' one text is as compared with another.

A detailed study of LD was first carried out by Ure (1971) as a measure of differentiating between spoken and written discourse for pedagogical purposes. The aim was to find out the major characteristics of language-in-action register, which is a variety :

> "that is needed by all foreign learners who are exposed to and require to use, language for everyday practical purposes." (ibid).

Ure used 'manual' calculations to find out LD in 34 spoken texts and 30 written texts comprising a total number of 21.000 words each. Her results show that the former texts, all except two, have an LD of **less than** 40% , whereas the written texts, all except two, have an LD of **greater than** 40%. Although these results are suggestive, they are not conclusive, since different subjects produced the spoken and the written data. This is an important source of variation as has been outlined in CHAPTER ONE above, (cf. also Beaman 1984 and Farag 1986).

Ure relates her findings to a variety of factors some of which are connected with the physical setting accompanying the actual production of spoken as opposed to written mediums and to other "extra-linguistic situational factors that

> "can affect the overall patterning of a text in terms of the selection of items and the frequency of their occurrence." (ibid:447)

Stubbs (1986), commenting on Ure's results, relates the reasons for these results to the fact that written texts are generally shorter, less redundant with fewer repetitions than spoken texts, and then they are permanent, highly redrafted and rehearsed with a greater degree of forethought and planning which spoken discourse generally lacks. Most of these reasons will be discussed in relation to the results of the present study and those of Ure's in the following sections.

Similarly, Stubbs (ibid), adopting the computer approach mentioned above, in analysing six spoken sub-texts of the London-Lund corpus, which represents recordings of highly educated informants, revealed a significantly higher LD than those reported by Ure (1971). Stubbs's calculations show the LD to be ranging between 44% and 56% as in the table below :

Text Number		Description	Density
1	10-5	radio commentary : state funeral	56%
2	10-1	radio commentary : cricket	54%
3	11-4	house of commons	48%
4	12-1	church sermons	47%
5	07-1	phone conversations : friends	45%
6	08-1	phone conversations : business	44%

Table 4.1 LD of six sub-types of spoken English texts (from Stubbs 1986)

Stubbs relates the difference of his results and those of Ure to the different methods used in the calculations and the nature of the corpora studied. He also mentions the level of education of his subjects as an important factor in this respect, which , as will be seen later, has been confirmed by the results of this investigation as well.

Hasan (1988:127-132) compares LD in native and non-native speaker speech in five types of formal and informal types of spoken discourse. The results are shown in the following table :

Data	Native Speakers	Non-Native Speakers
Formal Interview	47.02%	33.67%
Formal Classroom Interaction	38.25%	38.96%
Informal Classroom Interaction	41.15%	40.64%
Informal Classroom Discussion	37.97%	43.69%
Informal Conversation	42.48%	41.60%

Table 4.2 LD in five types of spoken discourse as reported by Hasan (1988 : Unpublished data)

Particularly interesting in the above percentages are those of native speaker formal interview 47.02% and native speaker informal conversation 42.48% as they show great similarity to the results of the present investigation, as will be seen below.

Farrugia (1988) in a recent computational investigation of 45 newspaper editorials from nine different newspapers : eight British dailies and a Maltese daily in English, has found that the LD of the newspaper editorials ranged between 45.80% for the Mirror daily newspaper and 50.40% for the Times of Malta. The motivation for including the Maltese newspaper editorials was an interest in Maltese English as a variety of world English, the aim being to see whether there is any marked difference from the British editorials in the context of his research. Whereas the editorials examined vary considerably in terms of style from one newspaper to another, the LD percentages are relatively uniform throughout.

All of these studies have aimed to rank text-types in terms of LD. In the following sections LD ranking of INT and CON will be discussed but attention will also be paid to why such ranking is problematic. The analysis falls in two main parts. In the first part variability in LD across all INTs and CONs (Inter-text lexical variability in LD is examined). The aim is to examine the variability in LD first between the same speaker's output in INT and CON and secondly between one speaker and another. In the second part we report the results of LD in the smaller various functional units of the speakers' output in the two variations under consideration (i.e. Intra-text lexical variability in LD).

4.3 Inter-text lexical variability in LD

4.3.1 Method of analysis

Drawing on insights from all the investigations reviewed above, two methods were used for calculating the L : G ratio in the overall corpus of INT and CON as a first step towards checking whether there is any difference in the speech of the same speakers in the two different settings. The first method used was done manually and the second one computationally. In the second method two simple computer programs (see Section 4.2.1 above) were used to identify all L and G words. Whereas the computer method was used to analyse as much data as possible, the manual one was adopted to check the results and handle such ambiguous cases as the ones mentioned above. Thus the final calculations represent an adjustment to the computer programs so as to take account of ambiguous classification, in all of which context was the basis of the decision. The results of the two computer programs are identical as they both present frequencies of L and G but in two different ways.

The results are going to be presented in tables, and graphs. Where graphs are presented, the raw figures will be presented in Appendix II, VOL.II of the thesis for further consultation. Statistical tests of significance(5) have been made where necessary to check the validity and significance of the obtained results.

4.3.2 The results

4.3.2.1 Overall LD in INT and CON

Before presenting LD results for INT and CON, two factors must be mentioned which could affect the results : Repetition (REP) and interviewer input. In order to differentiate between the output of subjects and that of the interviewer, Table 4.3 includes interviewer output and repetitions, while Table 4.5 excludes them altogether. The results of these two tables are presented collectively (i.e. for each pair of speakers in INT and CON). The reason for this type of presentation is to observe the influence of interviewer exchanges on the overall results and that of repeated L and G items on these results as well, which would not have been obtained from having the results of each subject presented individually. This would also help in observing the difference between the output of postgraduate speakers (PGs) and the undergraduate speakers (UGs). Each pair of subjects is identified as A, B, C etc. A, B and C pairs are PGs, pairs D to H being UGs.

Pairs	LD in INT	LD in CON
A	48.2%	46.9%
B	47.2%	44.4%
C	50.4%	47.4%
D	46.9%	44.3%
E	47.4%	47.6%
F	45.3%	47.3%
G	48.7%	47.6%
H	43.7%	46.3%
MEAN	**47.2%**	**46.5%**
SD	**02.05**	**01.38**
LB	**45.95%**	**45.66%**
UB	**48.45%**	**47.34%**
V	**04.34%**	**02.97%**

Table 4.3 Overall Lexical Density in INT and CON (including repetition and interviewer speech)

Table 4.3 above shows a slightly higher mean percentage of L words in the interview situation. However, the difference is not statistically significant (WSRT $p < 0.14$). The SD and V figures show a very slight difference between INT and CON with the former having slightly greater spread of LD values than in CON.

4.3.2.2 Repetition of Lexical and Grammatical words in INT and CON

Before examining the influence of REP on the overall LD of both INT and CON, it is interesting to have a look first at the percentages of repeated G and L items and see how far their recurrence in the data affects the overall results. For this purpose all repeated G and L items in the data were calculated as percentages. The results strikingly show REP to occur highly significantly more with G items than with L items as shown in Figure 4.1 below :

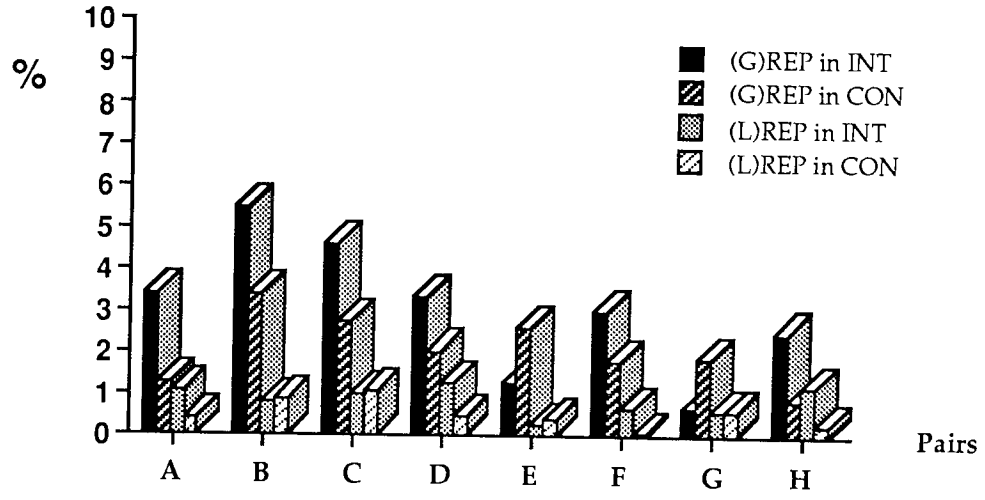


Fig.4.1 Percentage of Repeated G and L Words in INT and CON

The figure above shows the difference in the occurrence of REP in G and L in both INT and CON. The results clearly show that repetition occurs significantly more with G words than with L words (WSRT $p < 0.01$). The

results also show that REP of G items is more recurrent in INT than in CON with a statistically significant result (WSRT $p < 0.05$ level). REP of L items, on the other hand, though relatively rare in occurrence, occurs more in INT than in CON but the result is not statistically significant (WSRT $p < 0.10$ level).

4.3.2.3 LD in INT and CON excluding Repetition and interviewer speech

In order to assess the influence of repetitions and interviewer input, the LD percentages were recalculated excluding these. Table 4.4 presents corrected figures.

Pairs	INT	CON
A	48.7%	47.5%
B	48.5%	45.3%
C	51.2%	47.7%
D	48.1%	44.8%
E	48.%	48.1%
F	45.8%	47.6%
G	48.5%	47.8%
H	44.0%	46.5%
MEAN	**47.9%**	**46.9%**
SD	**02.14**	**01.25**
LB	**46.59%**	**46.14%**
UB	**49.21%**	**47.66%**
V	**04.47%**	**02.67%**

Table 4.4 Lexical Density in INT and CON excluding repetitions and interviewer speech

Comparing Table 4.3 with Table 4.4 above, it is noticed that there is no significant difference between the two which, in turn, suggests that REP has practically very slight influence on the ratio of G : L in both INT and CON. The results also show that, excluding interviewer speech and REP from the calculations, again LD is somewhat higher in INT than CON but the difference is still statistically insignificant (WSRT $p < 0.081$.) Again, it can be noticed that the SD and V figures of the interview data reflect the

fact that there is a greater spread of LD values. In other words, it is a less internally consistent set of figures than the CON set.

4.3.2.4 Postgraduate/Undergraduate variation in LD

A comparison of the MEAN percentage values for PGs as opposed to UGs is revealing.

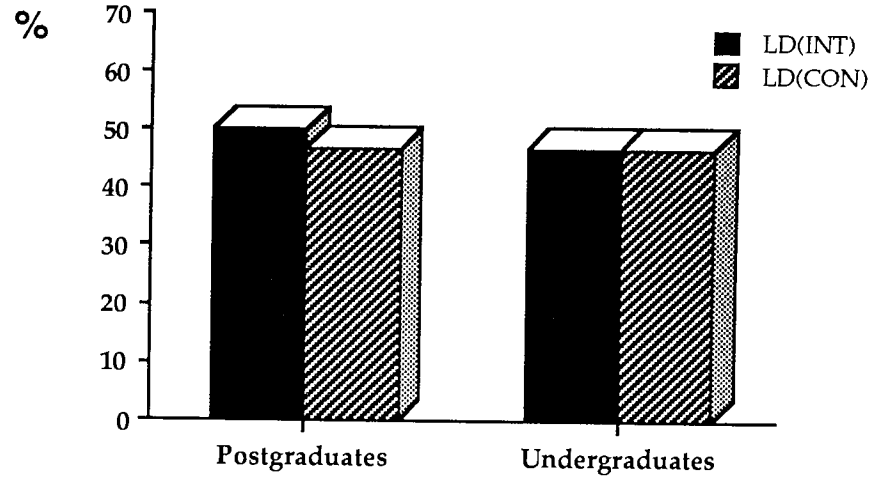


Fig.4.2 Mean Percentage Values of Lexical Density of Postgraduate and Undergraduate Students in INT and CON

Although the numbers are too small for valid statistical testing, there would appear to be some evidence that:-

1) UGs do not differ from PGs in terms of LD, in the conversational setting.

2) UGs, who are less mature and have received a shorter period of higher education, do not increase LD in the formal interview, whereas PGs do. This possibility requires further work with a bigger corpus of conversational data.

4.3.2.5 LD in the individuals' output

One aspect of the results above so far not discussed is that, as Ure (1971) found, the absolute LD value varies from one participant to another, in the same speaking task. First of all let us look at the variability in the Interviewer input in INT and CON (Table 4.5).

Sessions	INT	CON
A	43.1%	44.5%
B	43.5%	39.7%
C	45.7%	47.1%
D	41.4%	43.4%
E	41.2%	47.3%
F	43.3%	45.0%
G	49.6%	22.2%
H	42.6%	40.2%
MEAN	**43.9%**	**41.2%**
SD	**02.97**	**08.16**
LB	**42.08%**	**36.21%**
UB	**45.72%**	**46.19%**
V	**06.77%**	**19.81%**

Table 4.5 Lexical Density in Interviewer input in INT and CON

There seems to be quite a significant but inconsistent variation in the interviewer speech, which suggests that he constantly modifies his use of lexical items according to the type of situation and participants he is interviewing. The values of V show the variability to be greater in the CON data than in the INT data. This result might be due to the fact that the interviewer's participation in the conversation encounters was not consistent and he actually did not attend most of them. However, Table 4.6 below shows that there is less variability among participants' speech :

Participants	INT	CON
HC	48.1%	47.0%
KW	50.4%	47.8%
JH	48.7%	46.8%
BG	48.4%	44.2%
RH	52.0%	46.4%
DD	50.4%	49.1%
RF	50.0%	45.4%
PM	48.0%	41.3%
HK	47.4%	47.3%
KSH	48.7%	48.5%
AM	44.8%	45.9%
CB	44.8%	45.9%
JC	50.6%	47.9%
GM	46.5%	47.5%
AB	44.1%	46.9%
KS	43.8%	46.3%
MEAN	**48.0%**	**46.9%**
SD	**02.24**	**01.13**
LB	**46.77%**	**46.28%**
UB	**49.23%**	**47.52%**
V	**04.67%**	**02.41%**

Table 4.6 Lexical Density in subjects' output in INT and CON

The differences between the output of individual speakers in INT and CON are more apparent looking at the results of the table above, slightly more than the results of the previous tables. The results of Table 4.6 above are significant at well below the 0.05 level, which, though not too highly statistically significant, are more significant than the results of the same subjects in groups with the interviewer speech included, (Tables 4.3) or with the subjects in pairs (Table 4.4). It is also worth pointing out that the direction of difference is not consistent. There are 4 individuals for whom CON has a higher LD than INT (CB, GM, AM and KS); there are 3 individuals who produce the same or virtually the same LD in INT and CON (DD, HK and KSH) and there are 9 individuals for whom there is a clear step up in LD in INT compared with CON (HC, KW, JH, BG, RH, RF, PM, AM, and JC). Thus, almost half the speakers manifest a trend which is not in agreement with the trend established by averaging across the whole population. If LD is affected by maturity and educational level, further

work paying attention to the output of individuals will be required.

The relatively low degree of variability in the results as shown by the SD and V values are similar to that of Table 4.3 and 4.4 respectively. In spite of this variability, it can be noticed that :

(i) more G words than L words occur in the two conversational types of discourse examined , and

(ii) higher LD in the participants' speech in INT than in CON, which confirms the results reported earlier.

4.3.3 General discussion of results

Looking back at the results reported above, two obvious findings need to be looked at in some detail. The first one is the apparent difference between these results and those reported by other researchers on LD in various types of spoken and written discourse and the second one is the amount of inter-text/subject variability in LD. These two findings will be discussed under two separate headings.

4.3.3.1 Differences from other findings

Comparing the results above with those of other researchers, there seems to be a considerable discrepancy in percentages reported by each researcher. The aim of this section is to discuss those in relation to the findings reported above.

Obviously, Ure's (1971) results, taken as a whole, offer the greatest amount of variability as her data contained a great variety of spoken and written text types and they were produced by different subjects. It is also clear that the results of the above tables are quite different from hers in a

noticeable way. The LD percentages are considerably higher than those reported by her spoken data and are generally closer to those reported by Stubbs (1986) and Hasan (1988). Even the lowest percentage, in the present study, is higher than the highest in Ure's spoken data where percentages range from 23.9% *(assembling Angel Chimes)* to 43.2% *(radio sports commentary)*. In Stubbs (op.cit.), the range is from 44% *(business telephone conversations)* to 56% *(radio state funeral commentary)*.

However, before the reasons for the difference in results are investigated, it has to be pointed out that when particular text types are examined individually and compared to other counterpart text types, the degree of variation drops. For example, Farrugia (1988) reports that his LD results of newspaper editorials are analogous to six of Ure's written texts which are newspaper reports. Stubbs (1986) also reports that in spite of the variation between his own results and Ure's **"the tendency of the findings is clearly in the same direction.."** (ibid:41). For example, his texts 1, 2 and 4 are monologues in which one would expect a higher density; texts 3, 5 and 6 are dialogues and have relatively lower LDs (See Table 4.1).

Also if the mean values of Stubbs's three dialogic texts (45.7%) (Table 4.1) are compared with the mean values of the results of overall LD in CON (46.5%) (Table 4.3) and those of Hasan's native speakers' *'Informal Conversations'* (42.48%) (Table 4.2), a striking similarity between the three results can be found. Notice also the great similarity between Hasan's Formal Interview results of LD (47.02%) (Table 4.2) to the mean values of the INT results (47.2%) (Table 4.3).

It seems, then, that similar text-types have comparable lexical densities and it also appears that most types of texts, except some of Ure's language-in-action texts, for reason which will be mentioned below, seem to have an LD generally ranging between 45% and 55%. This is applicable to both spoken and written types of discourse. However, there are many

researchers in the field of lexical studies who take it for granted that spoken discourse types have lower LD than written text types. Halliday (1985b:62), for instance, believes that:

> "The difference between written and spoken language is one of density : the density with which the information is presented. Relative to each other, written language is dense, spoken language is sparse"

This claim has not been supported by the general tendency of the results reported above nor has it been supported by other research works. It is of course highly relevant that Halliday's notion of lexical-grammatical is not identical with what other researchers (including the present writer) take to mean. For example, Halliday points out that although the following items are lexical, they are actually on the borderline of grammar :

thing, people, way, do, make, get, have, go, good, many,

He then goes on to say that :

> "they often perform functions that are really grammatical-for example **thing** as a general noun (almost a pronoun) as in **that's a thing I could well do without; make** as a general verb , as in **you make me tired, it makes no difference**. They therefore contribute very little to lexical density" .

It would appear that semantic slipperiness (or multiplicity of reference) enters into his willingness to see 'thing' etc.. as grammatical. It should be noted that Ure's low LD results may be at least partly due to her adoption of Halliday's interpretations of 'Lexical' and 'grammatical' items exemplified above (Personal Communication). As far as can be judged, most other researchers (e.g. Stubbs 1986, Farrugia 1988, Hasan 1988) consistently take these "semantically slippery referentials" as lexical rather than 'grammatical' for reasons of consistency and to facilitate application

of computational techniques to the analysis of large corpora of data.

The results suggest, as do those of Stubbs's and Hasan's mentioned earlier, that spoken discourse modes, even those that might be assumed to be of the sparsest type, such as casual conversation or telephone conversation can have a high degree of LD and hence a high degree of informativity. Stubbs's results confirm this but he attributes it to the high level of education of his informants, **"whose spoken language is heavily affected by written language"** (ibid:41). The data from the postgraduate students examined here, support Stubbs's point of view, as they have been noticed to have higher LDs than the undergraduates' (See Fig. 4.2). So it seems that there must be factors other than educational level per se which contribute to the relatively high lexical densities of some spoken discourse. The density of information and its distribution in the discourse and also the types of informativity units within discourse might well have a role to play in this respect. The specific roles of the social functions of texts (Ure 1971) cannot be properly commented on unless the smaller units and chunks of discourse have been examined on a microanalytical level especially with the type of data examined in this study. As will be seen in the next few sections there is certainly a close relationship between LD and the discourse functions of these units in terms of 'informative', 'narrative', 'elicitive, or 'argumentative' force.

4.3.3.2 Inter-Text/Subject variability in LD

Concerning the level of LD per se in the INTs and CONs, the results seem to suggest that the language of INT has relatively higher LD than that of CON under all the different circumstances examined, though the results vary in the degree of statistical significance between insignificant to relatively significant. The reason for the relative difference in the speaker's output in INT and CON might well be due to the greater degree of self-monitoring the speaker experiences over his speech in INT than in

CON. A further explanation, one which is also mentioned by both Ure and Stubbs (ibid) when talking about the higher LD of written discourse over spoken discourse, concerns the 'planning' time the writer (or interviewee in this research) has available to him when producing the text and the degree of prethought he/she devotes in this production. Certainly, during the INT sessions the speakers experience more self-monitoring, prethinking and are more aware of what they are saying than in CON (cf. Zora 1986) simply because during INT they are constrained by the interview environment which includes the interviewer, the questions which they have to respond to and the controlled topic under discussion, all of which might not be so constraining in the CON sessions. This is certainly confirmed by the consistency of the results displayed in the tables above, which have shown that under all the circumstances examined the results have been consistent.

The results that different subjects produce consistently higher LD than others suggest that LD in SES is perhaps a style constant of the individual in much the same way that Farag (1986) found syntactic style to be stable across spoken and written performance of narrative discourse. This requires further controlled experimentation.

Another factor which can be noticed when examining each participant's results and which might have some influence on the results as far as LD is concerned is the degree of education and sophistication and also the skill and training a speaker has over another. All the results confirm that, in general terms, those speakers who have a higher degree of education produce units which have relatively higher LD than others. The impressionistic judgement of this analyst is that some of the subjects have a greater degree of education and are more sophisticated politically or religiously, and these are the ones who have a higher LD as shown in the tables above. Although the population examined is relatively small, and

possibly not representative of the conversational style and behaviour of the larger population in everyday normal interaction, the results seem to suggest that education maturity, previous political and religious skill and training can have a lot to do with the production of speech in general and with LD in particular. This confirms Stubbs's suggestion that the high LD he obtained from the analysis of the sub-texts of the London-Lund corpus is due to the high level of education the speakers have. Ure (1971) has also expressed a similar view when talking about the influence of previous education, training and skill on the performance of her subjects such as the results obtained from the sports commentator in her spoken data.

The sensitivity of the interviewer to his interviewees as seen in the variability in his LD from one interview session to another and also the inter-subject variability in LD suggest that the comparison of one text type against another may be misleading. It may obscure individual variation which can work in a direction opposite to the global one. Therefore, work on interactive sensitivity is needed, taking into account individual speakers and then in pairs and groups of speakers preferably in larger samples and under different circumstances. Only through this can one make valid conclusions and generalisations about the occurrence of G and L in SES.

4.4 Intra-Text lexical variability in LD

4.4.1 Some preliminary points

All the work that has been discussed above concerns studies of LD in whole running texts. No work that is known to the present analyst that has dealt with the problem of measuring LD in units smaller than whole texts except that of Halliday (1985b), which is principally theoretical in nature and is more concerned with written discourse than with spoken discourse or with spoken discourse but in relation to written discourse.

Halliday (ibid:66) rightly draws attention to the idea that since words are not packed inside other words but inside larger grammatical units, then

> "it is this packaging into larger grammatical structures that really determines the informational density of a passage of a text".

So his suggestion is to examine LD in clauses and clause complexes as this unit can fit both spoken and written discourse. In the present analysis, this approach is applied to the units of information(6) which the data was originally divided into since having the data of SES tidily segmented into clauses is faced with many difficulties as it has been outlined in CHAPTER THREE above.

Applying the LD approach to individual clauses with clear-cut boundaries is of course easier than applying it to units of the sort obtained by the segmentation adopted in this research simply because there is no one-to-one relationship between those units and any individual syntactic unit. As has already been observed in CHAPTER THREE above, the FSSUs can be a unit equivalent to a clause or may be a phrase, clause complex and so on. This makes the specification and counting of lexical/grammatical words even more difficult. This brought about the idea of observing each unit individually from the point of view of the '**social function**' (cf. Halliday 1973; Ure 1971) it is performing in speech and also from the point of view of the level of 'informativity' each unit is providing. The aims were, first of all, to see which type of units have high LDs. This would certainly lead to the second aim which is to find out what type of units of SES carries more information load than others as expressed by their higher amount of LD.

Ure (1971) adopted a similar, but more general approach in that she

specified the social functions of each of her spoken and written texts as a whole, i.e. as registers (ibid:450-451), and then she classified them according to the amount of lexical densities they have. Broadly, two types of texts appeared in her data, the **Language-in-action** type and the **Narrative** type. She divided the Language-in-action type into the following subtypes, **Immediate Doing, Consulting, Non-Immediate Describing, Directions, and Discussion**. The Narrative type, on the other hand, she divided into **Entertainment, Information and Exposition**. Her findings indicate that the Language-in- Action texts (all spoken) had the lowest LDs whereas the Narrative type, especially the written types had the highest LDs.

The problem with Ure's classification mentioned above is that she takes the whole text as one unit and assigns it a particular function. This is rather a crude type of classification in that many of these texts, especially the longer ones, can have more than one of the functions specified by Ure. For example, a text can be 'Informative' and 'Entertaining' at the same time. It is clear from Ure's Table 3 (P. 451) that she has not taken such overlapping into consideration and has treated each text as one individual unit with one particular social function. If this was to be applied to the type of data analysed in this study, the data would have just two types, 'INT' and 'CON' and it is totally unclear where they would fit in Ure's subclassification because each of them can have a multiplicity of functions of the sort distinguished by Ure.

4.4.2 A functional analysis

The solution, then, has to be to study each of the units of discourse the data of INT and CON consist of. The point of departure for the study of these functions is the distinction drawn by Brown and Yule (1983:1) between the **'TRANSACTIONAL'** and **'INTERACTIONAL'** functions of language (**IDEATIONAL/INTERPERSONAL** in Halliday's (1973:37) terms).

Since the language of SES in general is of the interactive type whose purpose is solely to negotiate meaning in social surroundings, it seems appropriate to include its functions under the 'interactional' type although one may find some exchanges and turns that may be classified under the 'transactional' type whose function is merely to convey information (Brown and Yule op.cit.). Halliday (op.cit.:41) expresses this relationship in a different way, as follows :

> "These two macro-functions, the ideational and the interpersonal together determine a large part of the meaning potential that is incorporated in the grammar of every language"

It is just right to mention here that Halliday (ibid) distinguishes a third function of language which he calls the **"textual"** function, which refers to the requirement that language should be operationally relevant- **"that it should have a texture, in real contexts of situation, that distinguishes a living message from a mere entry in a grammar or a dictionary."** (ibid). This is also interrelated with the other two functions to provide the 'meaning potential' of discourse. This interrelationship of functions is exactly why one needs to look in more detail at the function(s) each of the smaller units (than a whole text) of SES has within any piece of conversational interaction.

4.4.3 Kinds of functional units

When we examine the data of INT and CON on a less micro-analytical level of analysis the FSSUs of both types fall into functional categories according to the type and amount of information these units convey in running discourse. The labels used to describe these functions are adopted from various sources (e.g. Ure 1971) and are chosen on the basis of their suitability to the social function of the unit. They are as follows:

1. Exposition (EXP)
2. Information (personal) (IP)
3. Information (general) (IG)
4. Elicitation (ELI)
5. Argumentation (ARG)
6. Narration (NAR)
7. Description (DESC)

Following is a categorisation of each of these functions with examples from the data of INT and CON.

1. Expositional Units

Expositional Units are typically represented by those units which involve some form of expounding or explanation, commentary or interpretation. They are quite informative and are generally used as elaborative units after the first 'initiative' unit by a speaker who is normally responding to a question put to him/her by another speaker; unit 10 of the following excerpt, for example, is taken here to be of the expository type:

4.4 (INT II)

JA. 8. 1 when you say that it's male dominated
2 are you talking ab.. about.. just numbers..
3 or are you talking about sort of [structures
JH. 9. [attitudes..

4 and so on..

JH. 10. 1 attitudes and structures and.. absolutely everything..
2 university hierarchy..
3 the Guilders..
4 equally to blame really..

2. Information (Personal)

They are the units that contain personal information and points of view. They are informative of the speaker's personal experiences; thus they are normally characterised by the frequent use of first and second person pronouns; units 79 and 80 of the following example are representative of these units :

4.5 (INT II)

JA. 76. 1 so what would you try and do..
2 I mean you've only got twelve months
3 haven't you [..
BG. 77. [mmhmm

JA. 78. it's.. *(he he)*..

BG. 79. 1 well
2 I don't think we can d'..
3 we can change anything overnight like that
80.1 the only thing ** we can do is..
2 that we are women..
3 we are going on to those meetings..

3. Information (General)

As the name suggests, these units involve the provision of general information and points of view about topics discussed and hence their high degree of informativity; for example :

4.6 (INT IV)

DD. 16. mmhmm..
17. 1 I think it was::.. higher than.. what turned out to be.. the reality..
2 because I'd been at Sussex University before..
18. 1 I know I know Sussex has a reputation anyway..
2 of being of fairly left sort of radical.. university..

4. Elicitive Units

These are represented by inquisitive units usually in the form of direct or indirect questions put forward for the purpose of eliciting information. They normally entail the use of grammatical words, negation, hesitation and repetition. In the data, these units were highly recurrent in the Interviewer's speech during INT. Unit 2 in the following extract is an example :

4.7 (INT V)

JA. 1. 1 O.K..
2 e::m.. Karen.. Helen..
3 welcome..
2. 1 e::m.. can I ask first..
2 be.. fo::re you ever came to university..
3 e:::m.. perhaps why you chose Aston in the first place..
4 maybe you didn't choose Aston I don't know..
5 why did you choose Aston..

HK. 3. 1 well
2 I applied because I'm doing pharmacy..

5. Argumentational Units

These are units that involve discussion and argumentation about any of the topics discussed during the conversational encounters. They can be accompanied by interruptions and simultaneous speech. Such units are highly recurrent in political debates and interviews. They are relatively informative ; For example :

4.8 (INT IV)

DD. 62. 1 yes
2 if Aston would become more middle class
3 we ought to have instead [more support for CND..
RH. 63. [I don't know if it would

prove that simple..
64. 1 I don't know..
2 I think it's.. not that simple [..
DD. 65. [why..

RH. 3 because of the age differences [..
DD. 66. [mmhmm

6. Narrative Units

These are characterised by a succession of units involving the telling of stories, past experiences, jokes and episodes. Notice the following example from CON II in which the speaker (JH) starts a funny story. Most of the units it contains are of the Narrative type:

4.9 (CON II)

JH. 20. 1 another thing that was funny was
2 e:::m.. at the May Ball..
3 you know we have a line-out
4 with [.. the V.C. and.. Ch.. the Chancellor
JA. 21. [mmhmm

and.. the outgoing President and the incoming President..
22. 1 well
2 this year..
3 usually e::m.. you know you get tickets take a guest..
4 and usually the guests are ushered off upstairs into the bar..
5 and they wait for.. you to finish on the line-out
6 and you go upstairs..
23. 1 but this year.. Veronica Warner said..
2 no no
3 come along
4 all the guests can come along as well you know..
24. 1 so.. e:::m.. Don comes on the other side with his girl-friend..
2 a::nd I was on the other side
3 with.. a bloke who.. I wasn't going out with at the time
4 but we were just good friends..

25. 1 and e::m.. I was standing next to the VC and his wife you see
2 so everybody was coming up and sort of shaking hands and that..
26. and e::m.. the VC would introduce a few of them to me you know..
27. 1 he was quite good actually
2 the most important..
3 well
4 some of the important ones
28. 1 he said..
2 he turned to me..
3 gestured to me and that
4 oh
5 that sort of thing
6 and this is Jo Herterich
7 the next year's president sort of thing..
29. 1 a::nd all these old men were walking up and turning to Dave
2 the bloke with me and saying..
3 oh
4 how do you do..
5 pleased to meet you..
6 you know because..
7 you know with the name being Jo as well..

7. Descriptive Units

This type of unit involves the use of description especially the one that reflects vivid imagery, metaphor and sometimes even imagination. Such units are highly informative and involve the use of adjectives and adjectival phrases. The following excerpt contains some descriptive units.

4.10 (CON IV)

JA. 94. 1 there seems to be a lot of interest
2 and a lot of enthusiasm..
95. 1 e:::m.. I think e:::m.. there's quite a lot of fear about..
2 fear of each other..
3 and.. fear of e:::m whether.. people will be able to.. to do anything..
4 or be able to..
5 people know enough.. you know..

6 I don't know to be able to.. to invite anybody..
7 or I don't know enough to be able to answer questions people might ask of me
8 and then sort of fear or distrust of each other..
9 you know.. university Catholics..
10 or those Baptists or those Christian Union people or whatever [..

RF. 96. [mmhmm

97. 1 e:::m but I..
2 there's really quite a high level of a commitment
3 I think [..

RF. 98. [mmhmm

4 among a wider group of people than I had expected..

It has been noticed that SES, whether naturally occurring or otherwise, can be divided into such units, the degree of informativity and hence the use of the relevant lexical items is dependant on the amount of information provided within these units. It is also found that these units differ in the use of lexical and grammatical units, each in conjunction with the level of informativity and expressibility. The following analysis of data into these units and an overall examination of these units in terms of their lexical variability of LD makes that clear to a great extent.

4.4.4 The analysis

As has already been seen in CHAPTER THREE, the total number of Major Units in the data consists of 913 in INT and 1544 in CON (See Table 3.4). But for the purposes of this Chapter and for examining LD in these units all units that are ten words or more each have been isolated. The total number of Major Units which have more than ten words each and which have been investigated accordingly is 639 in INT and 773 in CON. These units were examined on a microanalytical level in terms of the functions each of them has in as far as the amount of information is concerned. First, the units were labelled accordingly on intuitive bases, and then a statistical analysis was done of the amount of LD each unit has.

A classification of these units has been duly arrived at. The following is the first of the classificatory figures showing the frequency of occurrence of each type of the above-described units in percentages in INT and CON.

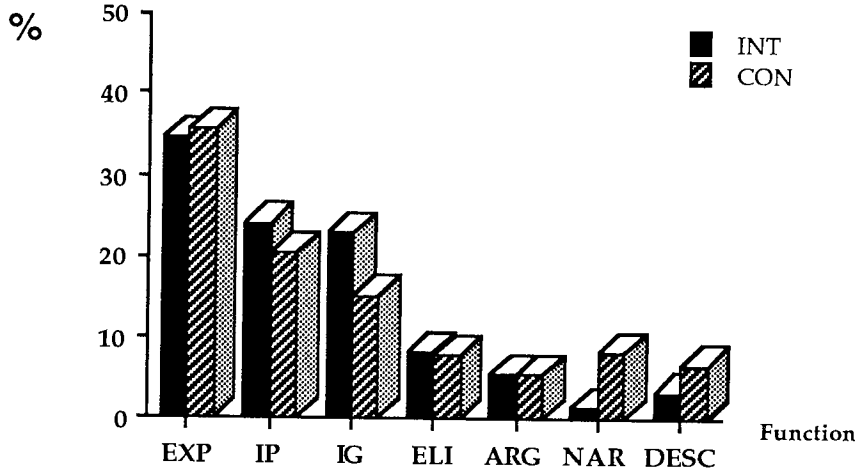


Fig.4.3 Percentage of Frequency of Occurrence of Major Units According to Function

This figure shows that, in the data, EXP units are the highest in occurrence and NAR and DES units the lowest. EXP units seem to occur slightly more in CON than in INT. IP, IG, ELI, ARG units occur more in INT than in CON. NAR and DESC units, though the lowest in the hierarchy of occurrence, occur significantly more in CON than in INT. Now in order to confirm these results, a thorough investigation of LD in these units was made drawing on the proposals of Halliday (1985b) stated earlier, by counting the G : L ratio in each unit. This yields another classification of three broad categories of LD units : under the first category all those units that have less than 40% LD were included; this will be referred to as **'LOW LD UNITS'**. Under the second the units that have an LD ranging between 40% to 50% were included. These are termed **'MEDIUM LD UNITS'**. The third category, which is called **'HIGH LD UNITS'**, covers all units that contain more than 50% LD. The following figure shows the percentage of each of the three categories LOW, MEDIUM and HIGH in both INT and CON.

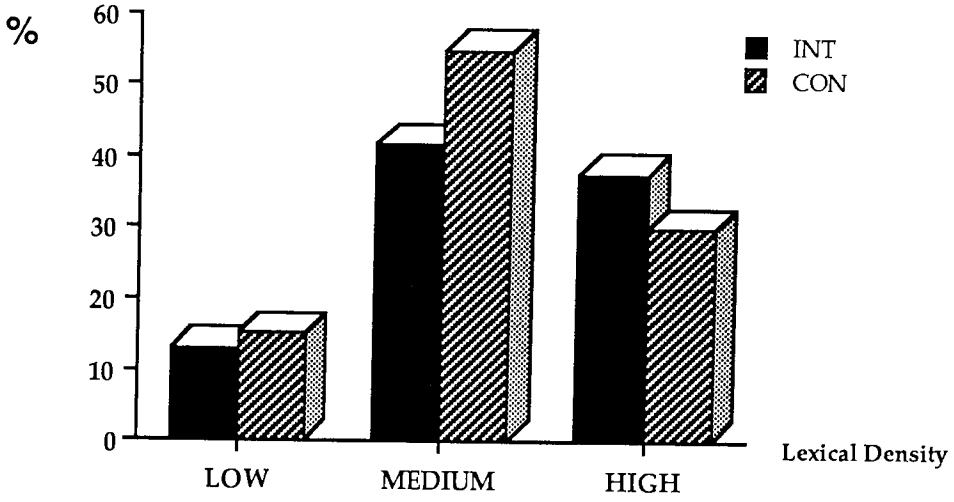


Fig.4.4 Percentage of LD Units in INT and CON

These simple statistics give us a clear idea of the distribution of LD within the units of discourse in the interviews and casual conversations analysed for this study and they also confirm, to a great extent, the results reported in section 4.3.2 above. First of all it can be noticed that, in both INT and CON, the majority of units (48.4% of the total number) have an LD between 40% to 50%. 33.8% have an LD higher than 50% but just 14.2% have an LD of lower than 40%, which is a striking result compared to the percentages reported by other researchers (cf. Ure 1971, Stubbs 1986 and Halliday's theoretically-based remarks) who all tend to imply that such percentages could only be obtained from written discourse or from 'very highly educated' near-written spoken discourse.

Next some attention is paid to how the 'functional informational units' (EXP, IP, GP...etc) are distributed within these categories and how they are organised again according to frequency of occurrence. The aims are first to examine the special properties of the data under investigation relative to the organisation and distribution of information density and its variability across INT as opposed to CON; secondly, to see whether one can establish a hierarchy of conversational units in terms of their density of

information, by measuring which of these units have a higher LD than others and which types can occur in which situations within the conversational organisation; thirdly to see how the presence or absence or low level of information density in these units is related to the density of information in the whole of the conversational body on the one hand, and to the way information may be packaged in other types of discourse especially written discourse as reported by other researchers, on the other.

4.4.5 Results of the investigation

Fig. 4.5 below shows the percentage of Major Units which have an LD of lower than 40%, hierarchically organised from top to bottom (from left to right in the figure). Remember that these units represent only 13.1% and 15.3% of the total number of units in INT and CON respectively.

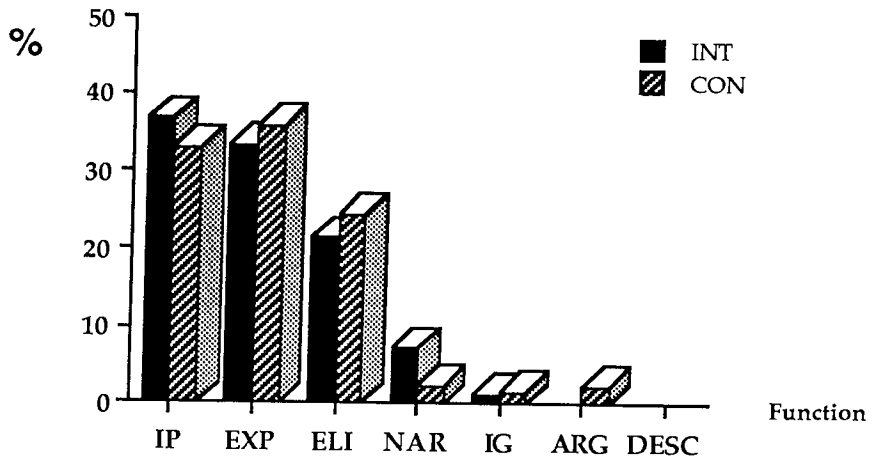


Fig.4.5 Percentage of LOW LD Units in INT and CON

Following the same procedure, Fig. 4.6 below shows the percentage of Major Units which have an LD ranging between 40% to 50% . In both INT and CON they represent the highest percentage of the total number of Major Units (49.1% in INT and 54.9% in CON).

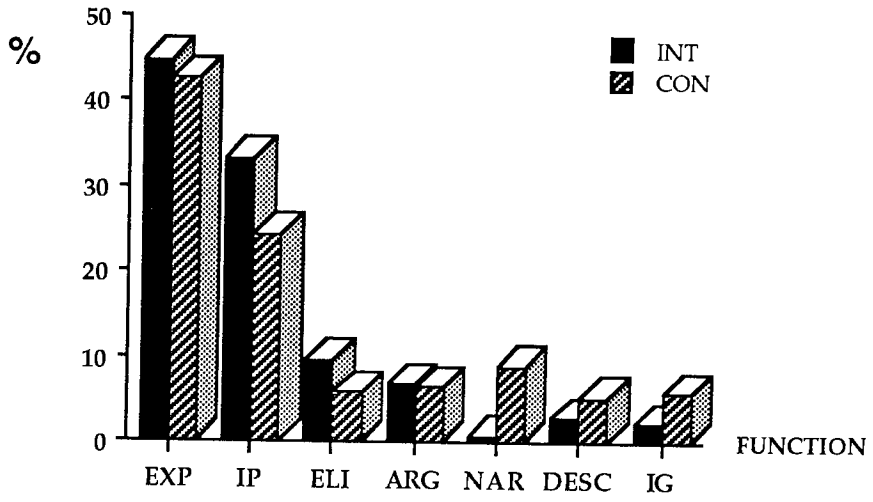


Fig. 4.6 Percentage of MEDIUM LD Units in INT and CON

The following figure shows the percentage of Major Units which have an LD of more than 50% . These represent 37.7% and 29.9% of the total number of Major Units in INT and CON respectively :

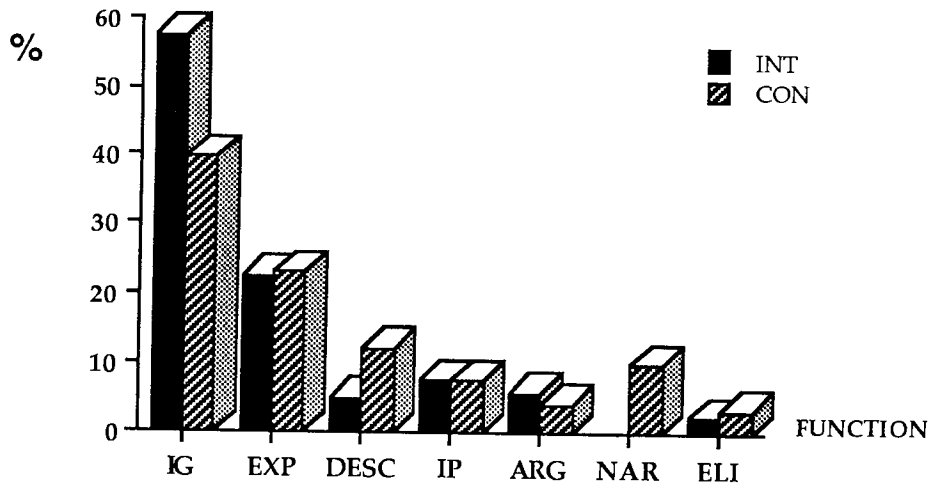


Fig. 4.7 Percentage of HIGH LD Units in INT and CON

The percentages in the three figures above show that, generally speaking, the results are uniform throughout, i.e. the hierarchical organisation in INT and CON is quite similar, which in turn shows that there is a lot of consistency in using lexical items in the various types of SES. In spite of this hierarchical conformity of results, there seems to be some variability between INT and CON in LD across the three categories under

investigation. However, due to the limited size of the population in this study, the results need to be checked against a larger body of conversational data in order to make valid generalisations.

The results in the figures above show clearly that EXP units have a high degree of occurrence within the three categories. They occupy the highest level of percentage within the second category (MEDIUM LD Unit), especially the ones which generally contain rather more general than personal information. There is also a high frequency of IP units within this category, significantly more in INT than in CON. They tend to occur within the LOW and Medium categories more than within the HIGH category. ELI units almost always tend to have a low LD for reasons which will be discussed below but they also occur within the MEDIUM category. In the HIGH LD category a considerably higher amount of IG units can be noticed, but greatly fewer units containing IP. EXP units also occur within this range but comparatively less so than with those falling within the two first categories. DESC units occur more within the HIGH LD than within the MEDIUM category, but notice that none of them occurred within the LOW LD category.

4.4.6 General discussion

The analysis of the small units of information in terms of LD has shown that the majority of LOW LD units (i.e. those that have an LD of less than 40%), generally speaking, are characterised by a high frequency of loosely connected units containing personal information and exposition. They tend to be mostly elliptical and have a great tendency to display such grammatical features as negation, passivisation, interrogation and so on. Notice these features in the following exchanges:

4.11 (CON IV)

JA. 122.1 well...
2 we had to make some sort of [..
PM. 123. [mmhmm

3 e:::m I mean they can join in later on..
4 and e:::m and both those and others
5 who who just haven't got the time at the moment
124.1 and of.. of..
2 I would like to think
3 I don't know whether..
4 I'd like to think we could meet for.. for the week-end
5 before the term starts..
6 next term [..
RF. 125. [mmhmm

7 a week-end away together..

RF. 126. rather like we did with the confirmation.. group

JA. 127.1 yes..
2 yes..

RF. 128.1 are you running those this year [
2 or..
JA. 129.1 [yes..
2 though I don't know quite how it will e::m.. fit in the..
this atmosphere.. [
RF. 130.1 [oh
2 [have you heard from..

3 because.. getting people to be committed
4 to two different groups at the same time..
5 plus everything [else..
RF. 131. [yeah..
132. have you heard from Sue by the way..

JA. 133.1 no
2 I haven't..

Most of these units contain disfluency features such as hesitation (filled and unfilled pauses), repetitions, false starts and the like (e.g. **Unit 124** above). The features of disfluency appear mostly in the interviewer's elicitive units and comments, which is thought to be a characteristic

feature of his speech as a skilled interviewer and 'elicitor of insights'. By adhering to these features, the interviewer tries skillfully to create the best conditions for the respondents to keep on talking smoothly and spontaneously. Notice the following examples :

4.12 (INT III)

JA. 15. 1 what.. what was your::.. e:::m expectation on coming.. to Aston..
2 on.. on the level of.. student.. interests.. conce::rn involvement..
3 in::.. not just CND but issues like that..
4 which are.. e:::m.. are slightly sort of alternative issues..

DD. 16. mmhmm..
17. 1 I think it was::.. higher than.. what turned out to be.. the reality..
2 because I'd been at Sussex university before..
18. 1 I know I know Sussex has a reputation anyway..
2 of being of fairly left sort of radical.. university..
19. 1 but e:::m..
2 I think I had imagined Aston to be more rounded.. university..
20. 1 in fact I'd imagined all universities to be more sort of rounded
2 in having.. you know art sciences.. social sciences engineering and so forth..
21. 1 and I found that.. by and large it was.. basically.. science and engineering..
2 and not very much or hardly anything in the way of arts
3 and not much in the way of social sciences..

JA. 22 mmhmm..

4.13 (INT II)

JA. 76. 1 so what would you try and do..
2 I mean you've only got twelve months
3 haven't you [..
BG. 77. [mmhmm

JA. 78. it's.. *(he he)*

BG. 79. 1 well
2 I don't think we can d'..
3 we can change anything overnight like that..
80. 1 the only thing ** we can do is..
2 that we are women..
3 we are going on to those meetings..
81. 1 and we have just got to show that..
2 that we are quite capable
3 and to be respected for.. what we are..
4 our.. our roles in the Guild..
82. 1 I.. I don't see myself going in.. as an ambassador
2 to change.. to change women status..
3 I'm there to do my job
83. 1 and.. if I happen to be a woman
2 and they can see that I can do.. a good job...
3 all well and good..

Within the MEDIUM category (i.e. the between 40%-50% category), a high percentage of EXP units and IP Units can be seen. These two types of information units seem to be characteristic of students speech as they respond to the interviewer's questions, providing details of their personal experiences within the IP units and exposing and elaborating on them in subsequent units of discourse (Notice **Unit 17** as an answer question 15 in **e.g. 4.12** and **Unit 79** as an answer to question **76 in e.g. 4.13** above).

Examining these units in greater detail, it is noticed that in the majority of cases, the amount of information exposed by the students depends on the type of elicitive question that they are responding to. Certain questions need more exposition and elaboration than others. These are found to be commonly elaborated on by the interviewer himself during the period of elicitation. In the following example (4.14) the respondents give direct expository answers to the short and direct questions posed by the interviewer, (but compare with the more complex and well elaborated question of the interviewer and the relatively complex response of the interviewee in **e.g. 4.12** above).

4.14 (INT VII)

JA. 47. how much do you think people ought to work..

GM. 48. 1 enough.. to::.. get the coursework done..
2 sort of as they are given it..
3 e:::m.. and to::.. clear up any points that you're not sure about
4 from the lectures
49. 1 and I think as you've done that
2 you've done enough..
3 don't need theory or background reading.. *(he he)*

JC. 50. I'd agree with that..
51. 1 I think if you do too much work
2 and work all the time..
3 you can miss out on some of the social benefits
4 of coming to university..
52. it's a case of getting the balance right..
53. 1 if you don't do any work..
2 then you'll be thrown out of university..
3 and you won't get any social benefits..
4 so you've got to get.. a balance..
5 between working
6 and going out to enjoy yourself..

JA. 54. 1 mmhmm...
2 what do you see as the social benefits..

JC. 55. 1 championship..
2 company..
3 going out and exploring new ground..
4 doing new things..
5 things you haven't done before..

JA. 56. was making friends easier or harder than you expected..

JC. 57. 1 it was very.. easy..
2 because.. everyone was in exactly the same situation..
3 everyone...
4 well
5 no one else knew anyone..
6 so.. people had to make friends as a matter of necessity..
58. 1 although having said that..
2 the friends ** I made the first.. week
3 aren't all still my close friends..
4 so.. they change a lot..

As might well be predicted, the majority of the units in the HIGH LD category, the over 50% LD category, are highly informative units, conveying mainly general information with greater detail and elaboration, with exemplification and description (See, for example **Units 17-20, e.g. 4.12** above). Interestingly, the majority of these units have been produced by the first three pairs of informants who are postgraduate students and by pair (seven) who are members of the highest Executive Committee at the University of Aston Guild Council. This confirms the previous suggestion that there is perhaps a direct relationship between lexical density/ informativity and educational maturity, skill and training of individual speakers, as both Ure (1971) and Stubbs (1986) have suggested.

Notice also that (IG) units occur significantly more in INT than in CON, within this category, which is quite expected since during INTs the interviewees provide answers containing elaborations and general information to make their points clear (**e.g. 4.12** above). Fig. 4.7 above shows clearly that NAR and ELI Units do not usually have high lexical densities as their occurrence in the HIGH LD category is rather low. Indeed it is noticed that even ELI units occurring within this category contain questions which are highly specialised and sophisticated with an abundance of detail, exposition and description and thus demanding equally complex answers (See **Unit 15** from **e.g. 4.12** above).

Narrative units occur equally well within the three categories according to the general style of the narrator and the type of story being told (See **Units 20-29** from **e.g. 4.9** above).

The general conclusion that emerges from the results and the discussion above is that there is a greater degree of variability in the distribution of LD and informativity within the units of information of SES than that shown by the results of LD along the longer stretches of INT

and CON reported above. The units of information seem to vary in terms of lexical employment in conjunction with other linguistic, paralinguistic, cognitive and situational factors, causing an ebband flow of information, in the sense that it is unevenly distributed. However, this flow of information is not chaotic, as some might think, it is highly organised, systematic and controlled especially during INT where the organisation and control is directly manifested by the role of the interviewer.

Speakers in conversational settings exchange talk for the purpose of conveying information in the units of language they use. These units are skillfully and unconsciously delivered, with the speaker organising the distribution of his information according to the situation, topic, participants, feedback and other relevant social and cognitive factors. It seems that it is a characteristic feature of conversation then and of speakers in general to vary their units in accordance with these factors. No one can imagine a speaker using one uniform type of units such as those mentioned above. However, there seems to be a hierarchy in these units to be organised in terms of lexical and informational density as seen in Fig. 4.7 above. More research to validate these speculations is required.

4.5 Summary and conclusions

To summarise the discussion so far it seems that there are a number of questions which have to be answered about the two major parts of the chapter concerning the issue of variability in LD and the relationship between them before one can draw final conclusions. As far as the first part of the chapter is concerned, it has been observed that there is a considerable amount of discrepancy and difference between the results of this analysis and those of other researchers concerning the amount of LD in discourse (see Section 4.3 above). The question which would have to be asked now is : why do different researchers report very different percentages for apparently identical speaking tasks ? (Compare Stubbs's

54% for radio cricket commentary with Ure's 43.2% for radio football commentary; or Ure's 'Life' discussion among students (35.2%) with the figure for conversation between students (46.9%) in the present study). Does this difference really refer to real variation across types of discourse or are they due to other (perhaps extra- linguistic) factors ?

There seem to be be at least eight sources of variation :

1. **basis for calculating LD** i.e. differences in allocating items to lexical as opposed to grammatical classes.

2. **expected interruption and length of speaking turn** longer monologic texts predisposing speakers to higher LD (see figures in Stubbs (1986) and Ure (1971) where spoken texts with higher LD are monologues, such as sermons, House of Commons debates, radio commentaries, or lectures).

3. **function of component units of text** In the present study, when units with specific functions (e.g. narrative, informative, elicitive, argumentative and so on) are compared, there seems to be some sort of systematic hierarchy of these units in terms of LD, as shown in the results of Section 4.4 above. One example showing this type of variation is the difference between the amount of LD of interviewer speech (43.9%) and the MEAN LD of the Interviewees (47.8%). This is due to the nature of the interviewer's speech, which is elicitive, repetitious, and full of hedges and pauses.

4. **self-consciousness/self-monitoring** (Compare Ure's figures for lecture (39.6%) and recorded language laboratory instructions (40.9%) with the mean 48.0% in INT, 46.9% in CON, obtained in the present study).

5. **personal attributes** : maturity, educational level, confidence, etc. (cf. Stubbs (1986) who comments that the high LD obtained in his study of the London-Lund corpus could have been the product of the high educational level of the speakers). Similarly, Ure (1971) talks of the influence of the previous experience, skill and education on the performance of her subjects.

6. **group attributes** : age, sex, educational level, etc. (In the present study, undergraduates produce lower LD in the interview situation than postgraduates both in the total percentage of LD (Section 4.3) and in the specific units of information (Section 4.4). It should be noted that group attributes may not always be distinguishable from personal attributes.)

7. **planning time** (Both Ure, 1971 and Stubbs, 1986 mention this as distinguishing between spoken and written production, and it may also contribute to the monitored/unmonitored distinction.)

8. **topic** (Stubbs presents a different LD for state funeral commentary (56%) as opposed to radio cricket commentary (54%). The same 'genre' with different topic and presumably different textual sub-functions can manifest different LD levels.)

It is clearly desirable that all eight factors should be controlled in experimental studies of LD, although the difficulties of doing so are not underestimated. Ure (1971), for example, has two almost directly comparable texts : a spoken text (LD 32.2%) **'How to repot a plant'** and a written text (LD 47.1%) **'Planting and soil'**. It may be difficult to obtain a direct spoken counterpart of a written text ; or, indeed, there may be no direct spoken counterpart. (What would be the spoken counterpart of a television news text, which is normally read aloud from a teletext machine ?)

There is scope for applying algorithms such as the one developed by Stubbs (1986) to data as wide-ranging as Ure's (1971), but designed in such a way as to ensure that the same subjects produce contrasted text types, on the same topic. Until we know more about the sources of variation in LD, explanation of the functions of variation in LD will remain tentative.

Finally, it can be concluded that LD does not seem to differentiate between discourse modes in a global way. Rather, it differentiates between INT and CON for the postgraduates analysed. Undergraduates, on the other hand, perform comparably, in terms of LD, in both the INT and the CON setting. Since the population examined is a) small and b) not evenly balanced (as between undergraduates and postgraduates), it is premature to conclude that an absolute statistics for the LD of undergraduates and postgraduates can be produced. What is interesting, and worth pursuing further, is the differential between the two groups in skill and/or sensitivity at the lexical level. It would appear that postgraduates, who are more mature and have longer exposure to higher education, adjust their LD to match some perceived characteristic of the INT situation. Postgraduates may be more able to compete on an equal footing with the interviewer, and this ability may derive in part from a perception that their own status is close to that of the interviewer. Essentially, what is being suggested is an application of 'ACCOMMODATION THEORY' (cf. Giles and St Clair 1979), which is well known in social psychology, to the lexical level of linguistic control, as an explanation for the rise in LD in the INT situation. The interviewer's drop in LD can be seen as a conciliatory gesture, metaphorically the opposite of a claim to status; and this in turn facilitates the closure of the status gap of which the postgraduates are able to make use of.

Footnotes to CHAPTER FOUR

(1) See also Ellis (1966) for a detailed discussion of the notion of contextual meaning.

(2) Ure does not mention explicitly how her calculations have been carried out but it is the present researcher's assumption that they were done manually.

(3) J. Ure's invaluable comments on her 1971 work and on a joint paper (by the present author with C. Johns Lewis, presented to the 7th Sociolinguistics Symposium held in York 1988), based on the contents of this chapter came as a result of oral discussion on the issue of LD.

(4) See Stubbs (1986:36) for details of this list. His list has been adopted with certain modifications. Other lists are also available in the literature (see e.g. that of Knowles (mimeo). Farrugia (1988) adopts a comprehensive list of 321 G words, which is a modification of both Stubbs' and Knowles' mentioned above.

(5) These tests have been carried out computationally using Wilcoxon's Matched Pairs Test (Details to be found in Meddis 1975). A computer program called **'STATWORKS'** was used for all the tests examined.

(6) The unit of information referred to here should not be confused with Halliday's 'Information Unit' (See Halliday 1985a and discussion in CHAPTER THREE above). Halliday's 'Information unit is realised intonationally as a 'tone group' corresponding to other similar units found in the literature such as Crystal's (1975) 'tone unit', Grimes' (1975) 'Information Block' and Chafe's (1980) 'Idea Unit'.

CHAPTER FIVE

THE SYNTACTIC STRUCTURE OF INT AND CON

5.0 Introduction

The literature on spoken versus written variation, has reported contradictory findings on the syntactic complexity (henceforth SC) of spoken discourse, ranging between two main extremes. At the one extreme, it is believed that spoken discourse is syntactically simple, less structured and less complex than written discourse (cf. O'Donnell 1974, Kroll 1977, Chafe 1979, 1982), with many **"incomplete sentences, often simply sequences of phrases"** (Brown and Yule 1983a:15). At the other extreme, spoken discourse is considered to be on the whole structurally complex, the degree of complexity being manifested in the greater number of clauses, their variation and combination into **"complex sentence structures"** (Halliday 1979:49) although simultaneously there is evidence of low lexical density and greater reliance on coordinated phrases and clauses.

In this chapter, a statistical analysis of INT and CON is presented. The analysis focuses on the relative frequency of all syntactic features potentially contributing to the linguistic complexity of spoken discourse. The relevant features are : clauses (finite and nonfinite) and their combinations (clause complexes) via the different syntactic and semantic processes of embedding, subordination, coordination, apposition and the like. Such features which are characteristic features of SES like ellipsis, discourse markers and backchannels will also be looked at later in the chapter to see how far they contribute to the complexity (or otherwise) simplicity of sequential spoken discourse. The results will be displayed as differences between INT and CON and will be discussed accordingly. And since there is no one-to-one relationship between the Major Units in the

data and any one single grammatical unit (e.g. clause), the percentages and OPTWs of the grammatical features that are found within the Major Units will be looked at and compared to the total number of these units (displayed in Table 3.4 and 3.5 above). The differences between INT and CON in as far as the number and nature of these syntactic features are concerned will give a clue to the amount of SC in SES in general and to the differences, if any, between speakers' output in INT and CON, in particular.

The literature on spoken versus written discourse variation and other related work has been reviewed in CHAPTER ONE. Reference to definitions, methodological principles and evidence within this area of investigation will be made when necessary. The terminology on syntax will be mainly adopted from Quirk et al. (1985). However, for reasons which have already been discussed in CHAPTER ONE, the use of the term '**sentence**' (in their sense) will be avoided and replaced with terms mainly adopted from Halliday (1985b). For example, for a simple finite sentence with a subject, a verb with or without a complement, I shall be using the term '**clause**' and for units containing more than one clause having either paratactic and/or hypotactic relations(1), the term '**clause complex**' will be used (see Section 5.2.3 below).

5.1 Syntactic complexity in spoken discourse

5.1.1 Some preliminary considerations

As a point of departure for the discussion of SC in English SES, a vital distinction which is made by Crystal and Davy (1969) when investigating the style of discourse including that of conversation will have to be discussed. They first of all draw attention to the variable nature of syntax in spoken interaction, noting that informal conversation is characterised by a large number of loosely coordinated clauses, the coordination being

structurally ambiguous. For them it is an open question whether one takes these as **"sequences of sentences or as single compound sentences"** (ibid:110). If taken as sequences of sentences then the statement, which Crystal and Davy (ibid) make, that **"sentence length is relatively short, and in structure displays predominantly the simple type"** seems quite acceptable. On the other hand, if coordinated sequences are taken as whole units, then the analysis must point to a high proportion of longer, more complex and more varied sentence types. Crystal and Davy adopt the former position since it produces simpler descriptions. As will be seen below, this position is analytically unsatisfactory if one looks at the matter from a 'discourse analysis' or a 'conversation analysis' perspective. Both these theories take SES to consist of units sequentially and coherently organised, not just an accumulation of combined 'sentences'. It is a highly organised phenomenon which involves the multiple use of devices which are subtly organised in a sequential manner the diversity of which causes both confusion and difficulty (cf. Wardhaugh 1985:57). As such it cannot be broken into separate 'self-contained' sentences to be studied away from their context.

Although not explicitly stated in any of the works on language variation, it seems that researchers who take SES to consist of **'loosely connected'** sentences (i.e. akin to the first of Crystal and Davy's options cited above) are certainly inclined to find that the language of SES, in general terms, is syntactically simple, fragmentary and chaotic (cf. O'Donnell 1974, Kroll 1977, Chafe 1982). On the other hand, those who seem to take Crystal and Davy's second point of view (i.e. that conversational sequences are units, not fragments of loosely coordinated clauses) have come to the conclusion that spoken discourse is syntactically complex, sometimes as complex as written discourse if not more complex (cf. Poole and Field 1976, Halliday 1979, 1985b, Beaman 1984). The task in this chapter is to establish the complexity of these units in terms of syntax. What this means is that the nature of the syntactic devices which connect

clauses is the focus of the analysis.

5.1.2 Indices for syntactic complexity/simplicity

Looking back at the two contradictory points of view concerning the simplicity/complexity of spoken discourse in relation to written discourse, certain features of syntactic complexity in discourse will have to be established before one can start the actual syntactic analysis. It is noticed that proponents of the view that spoken discourse is syntactically simple and less structured than written discourse base their views on the claims that spoken discourse shows heavy reliance on **"loosely coordinated structures"** (Crystal and Davy op.cit.:110), or **"spoken discourse typically contains rather little subordination"** (Brown and Yule 1983a:15), (see also O'Donnell 1974, Kroll 1977, and Chafe 1982). On the other hand, proponents who support the view of the syntactic complexity of spoken discourse base their views on the findings that spoken discourse is characterised by **"complex structures with low lexical density (more clauses, but fewer high content words per clause"** (Halliday 1979:49), or by more **"subordination than coordination"** (Beaman 1984) and by **"structural elaboration, nesting and embeddedness and uncommon clause usage"** (Poole and Field 1976:309). These contradicting views suggest that, generally speaking, researchers tend to use the size and structure of the clause and clause connectivity by subordination as markers of syntactic complexity and they take connectivity by coordination, fragmentariness, and low lexical density as markers of syntactic simplicity.

Since there is no uniform view of complexity and simplicity in the various research work in the literature and since it is not clear how far different kinds of SES can be complex or simple in relation to each other, certain indices will have to be established for statistical research purposes. These indices will have to be based on assumptions which might be verified or otherwise contradicted by the results of the analyses that will be

presented in the following sections. Taking insights from the controversies reported above (and in CHAPTER ONE), it will be assumed that indices for syntactic complexity in INT and CON include greater frequencies of complex phrases and clauses, subordinated clause complexes, appositives and the like. The multiple use of clauses and clause complexes via embedding will be an essential index of complexity.

Syntactic simplicity, on the other hand, will be measured by lack of the conditions mentioned above i.e. by the general absence of long units containing subordination and embedding, and the greater use of loosely coordinated structures (see Section 5.3 below), incomplete and interrupted units (see Section 5.5 below) and low lexical densities (the results of which are reported in CHAPTER FOUR above). An investigation of elliptical units essentially calls for an overall analysis of these units as well as the features and factors affecting 'ellipticity' in discourse as used by speakers and listeners alike, such as interruptions and self-interruptions, discourse markers (cf. Schiffrin 1987) and backchannels (cf. Yngve 1970).

It is believed that the approach of specifying indices of complexity and simplicity will help in two ways. First, it will give a clear picture of the syntactic complexity, if at all present, in the two types of spoken discourse under examination. Secondly, it will help in making valid generalisations that can be used for comparison with findings of other research works in the literature.

It should be pointed out at this stage that such terms as 'subordination', 'coordination', 'embedding', 'ellipticals' and so on are very broad terms. As such the mere specification of the overall frequency of occurrence of each is not sufficient to characterise SES as syntactically complex or otherwise although it might give certain helpful clues. There is certainly a need to look at each of these categories in great detail and examine its main subcategories and constituents in order to be able to see exactly

where the source of complexity lies and how it is related to other (sub)categories within the list. For example, of the broad category of 'subordination', one can distinguish the subcategories 'Nominals', 'Relatives' and 'Adverbials'. Each of these subcategories contains other subcategories and each tends to have perhaps a different type of behaviour and frequency within INT and CON. This is why the analysis will cover all these categories and their major subcategories in both INT and CON.

5.2 Features of syntactic complexity

5.2.1 Syntactic form and function

The analysis of spoken interactional discourse, which is more of a process of human continuous interaction than an existing object like written discourse (cf. Halliday 1979:49), is rather intricate and tedious simply because it does not exist in one uniform form or one stable condition of language use that can be amenable to linguistic analysis the way written discourse is. And if it is made to exist as such in the form of long conversational transcripts, it loses much of its characteristic features namely its rhythmicality, spontaneity, and function as verbal communication. On the various levels of linguistic analysis such transcripts are not much different from any piece of written discourse and are subject to the same principles and methodologies that are used in analysing written texts. As such these conversational transcripts can be analysed on the level of their linguistic forms and/or on the level of the functions that these forms are used to fulfill in discourse. I am using the distinction form/function in Quirk et al's (1985:48) sense. These researchers believe that each grammatical constituent of whatever size displays a particular grammatical function within its formal use in discourse. They identify "**form**" in terms of the constituents of sentences and clauses i.e. in terms of the "**internal structure of sentences and clauses as a noun phrase, or as a verb phrase**" and so on. For them "**function**"

refers to the :

> "unit's 'privilege of occurrence' in terms of its position, mobility, optionality, etc. in the unit of which it is a constituent." (ibid)

The function of a unit specifies it as a subject, or an object within a particular clause. Quirk et al. (op. cit.) point out that :

> "two units which have the same privilege of occurrence may be said to be FUNCTIONALLY EQUIVALENT" (Capitals in original).

For instance, in the following example, adopted from Quirk et al. (ibid:49), the underlined words, although they belong to different formal categories (adverb phrase, noun phrase, prepositional phrase), may be said to belong to the same functional category of ADVERBIAL.

5.1 a. The weather has been very cold { **just recently**.
{ **this month**.
{ **during the past week**

Such a distinction between form and function is important for the purposes of this study. It will first help to determine the degree of complexity of verbal utterances and of the relationship between the functions of the syntactic units they consist of. Secondly, it will help in carrying out the statistical frequencies and counts and also it will help in solving some of the problems related to linguistic **'complexity'** of spoken discourse, which is the major aim of the analysis in this chapter.

5.2.2 Embedding

Embedding as an important feature of SC and organisation is defined as:

> "the occurrence of one unit as a constituent of another unit at the same rank in the grammatical hierarchy" (Quirk et al. 1985:44).

In transformational grammar, embedded 'sentences' are those that are themselves constituents of other sentences. For example in :

5.2 (CON III)

RH. 100.1 the fact that watches are disposable..
2 is one thing I'm quite pleased about..

"that watches are disposable" is embedded in the structure of **"the fact that watches are disposable"** and the clause **"I'm quite pleased about"** is embedded in the structure of **"one thing I'm quite pleased about"**.

The phenomenon of embedding accounts for the indefinite extensibility of certain units of grammar. Complex units of discourse are those that have a multiplicity of cases of embedded phrases and/or clauses within them. The more cases of embedding the unit has the more complex it is likely to become. A close examination of the data of INT and CON in this study reveals the fact that SES, in general, is characterised by a relatively high frequency of occurrence of complex cases of embedding. But the main problem is that it is virtually impossible to count the exact number of these cases simply because phrases and clauses can be embedded in other clauses which are in themselves embedded within other clauses. And it seems that with the absence of well-formed grammatical 'sentences' in SES there is no limit to the extensibility of

units containing a multiplicity of phrases and clauses embedded within each other. Quirk et al. (op.cit.) confirm this view by pointing out that **"embedding gives rise to the theoretical possibility of grammatical units having indefinite length"**. Notice the following examples from the data :

5.3 (INT I)

HC. 79. 1 e:::m.. well
2 it was different for me
3 in that I lived in a Presbytery..
80. 1 I worked in a parish
2 while I was doing.. e:::m my degree..
3 which e:::m... gave it a different aspect..
81. of course it was a philosophy degree and a theology degree..
82. 1 it sort of.. helped to.. to::.. base in reality..
2 which was.. which was interesting..
3 which was quite different from what I am doing here [..
JA. 83. [mmhmm

4 which is management [.. -- --
JA. 84. [mmhmm

5.4 (CON VIII)

GM. 121.1 the thing was if you'd gone.. e:::.. down to the kitchens..
2 from the first day..
3 well
4 from the first day of term..
5 and just carried on
6 nobody would have questioned you [..
JC. 122. [yea::h..

7 because nobody ever at any stage..
8 e:::m asked to see your ID
9 or proof of where you're living
10 or anything like that [..
JC. 123. [yeah..

These examples show the difficulty of getting a straightforward answer to

the problem of counting the number of embedding cases within SES. However, the number of FSSUs (see Section 3.2.1 above) which are clause complexes containing a multiplicity of embedded clauses has been calculated. The results are included in Tables 3.6 and 3.7 above. The high OPTWs of these units in the data suggest the presence of SC on the level of embedded clauses and phrases within units of SES. The figures also show the language of INT to be more structurally complex than that of informal CON. This is indicated by the relatively higher use of multiclausal units that have complex cases of embedding in INT than in CON.

Of course, these figures do not include all the cases of embedding within the data. There are other cases within the 2-clause units and the incomplete units (again see Table 3.6 and 3.7 above). However, these occur significantly more in CON than in INT and have less effect on the issue of SC than those cases of embedding within the multi-clause units. This variability in the use of syntactic units and their higher use in one style of speech rather than another is a strong motive for investigating the issue of SC in INT and CON to verify or otherwise challenge the hypotheses suggested by the results above.

The suggestions and speculations made above should not be taken to mean, however, that embedding is the sole source of SC in SES. There are many other factors that contribute to this type of complexity, such as ellipsis, misplaced words, phrases and clauses, word order, and so on. Some of these issues will also be touched upon throughout the analyses and discussions in the following sections.

The investigation of embedding and its relationship to the SC of discourse necessitate the study of clauses and the syntactic and semantic relations that hold between them. These relations are expressed by what syntacticians refer to as **"interdependency"** which is general to all complex

units of language starting from words, phrases and clauses alike. The present investigation in this chapter will only be confined to clause complexes, as will be outlined below.

5.2.3 Kinds of 'interdependency' between clauses

First of all, the distinction between **'independent'** and **'dependent'** clauses in syntactic terms has to be explained well in advance as most of the discussion that follows is related to it in one way or another. An ***independent clause*** is the one that is capable of constituting a simple 'sentence' by itself, and a ***dependent clause*** is a clause that makes up a grammatical sentence only if subordinate to a further clause (Quirk et al. 1972:721).

5.5 (CON VII)

JA. 1. 1 oh
2 I'm sure it will..

'I'm sure' *(Independent)* ***'it will'**.. *(Dependent)*

'I'm sure it will'.. *(Independent with Dependent)*

Most grammarians agree that syntactic relations between dependent and independent clauses within clause complexes are of two major kinds **'parataxis'** (i.e. *equal arrangement*) and **'hypotaxis'** (*underneath arrangement*) (cf. Halliday 1985a:193, Quirk et al 1985:918). 'Parataxis' refers to the relation between two like elements of equal status, and 'hypotaxis' to that between 'unequal' elements, between a dependent element and its dominant, the element on which it is dependent (Halliday:ibid). In other words, the paratactic relation is **symmetrical** whereas the hypotactic relation is **non-symmetrical** (Halliday 1985a:198). Out of this neat distinction, **'coordination'** and **'subordination'** emerge as special cases of

'parataxis' and 'hypotaxis' relations respectively. Coordination is the state of joining two or more clauses using any of the 'coordinating conjunctions' producing what has traditionally been termed 'compound sentence'. Subordination, on the other hand, refers to the joining of two clauses in such a way that one of them is dependent (subordinate) on another independent clause (complex sentence in traditional terms) using any of the 'subordinating conjunctions'. As already mentioned above, following Halliday (ibid), the term **'clause complex'** has been used in this study to refer to both cases of 'coordination' and 'subordination', so as to avoid the problems and ambiguities in the use of 'sentence' in spoken discourse.

It seems obvious, then, looking at the distinction made above, that, from the point of view of syntactic complexity, 'coordination' constitutes less complexities. It does not involve complex syntactic dependencies on adjacent or embedded units such as those displayed by subordinated clauses. This seems to be one of the reasons why some authors (e.g. Kroll 1977) relate the SC of discourse to the high frequency of subordinated and embedded clauses and syntactic 'simplicity' to the greater use of coordination than subordination. This will be one of the crucial issues to be investigated in this chapter. The issue of 'coordination' will first of all be discussed and then an overall discussion of the more complex issue of 'subordination' will follow before presenting other issues that might be linked in one way or another to the complexity of SES, such as ellipticals and so on.

5.3 Coordination in INT and CON

5.3.1 'Strong' and 'Weak' coordination

As has already been outlined, 'coordination' occurs when **"two or more units of the same status on the grammatical hierarchy may constitute a single unit of the same kind"** (Quirk et al. 1985:46). This type of construction is typically signalled by a 'coordinating conjunction' of which the most important are : ***'and', 'but', 'or'*** and ***'so'***.

It is quite obvious that the language of SES, especially 'conversation' and less so 'interview' contains a large number of these coordinated clauses. However, there is some sort of controversy concerning what constitutes real coordination. Many of the so-called coordinators (especially *'and'*, *'but'*, and *'so'*) have functions other than coordination and it is sometimes difficult to distinguish between when they are coordinators and when they are not. In this research, the answer to this question is crucially important in determining the Major Unit boundaries as well as in determining many issues related to the SC of spoken discourse. In the segmentation of data into Major Units, (see Section 3.2 above) clauses syntactically joined by a coordinating conjunction represent a clause complex, which might stand as one Major Unit depending on its semantic coherence. Accordingly, when any of these conjunctions functions as a 'real coordinator', it should not by definition appear at unit boundaries but within them. This state of coordination has been termed **'strong coordination'** (see the use of *'and'* in **Units 28.4 and 28.5, e.g. 5.6** below). However, within running speech, these connectives sometimes do not function as coordinators in the real sense of the term as outlined above; they might commonly be used with some 'pragmatic' functions (cf. van Dijk 1977, Beaman 1984, Stubbs 1983, and see also Farag (1986:66) for a detailed discussion), for example, when they occur as 'fillers' or as initiators of new topics (see **e.g. 5.6, Unit 29**). For ease of reference, this type

of joining will be called **'weak coordination'**. Conjunctions used within this function do not syntactically join separate units although from the point of view of spontaneity and sequentiality of discourse and at a more pragmatic level of discourse they do. These, in the segmentation of the data of INT and CON into Major Units, occur at unit boundaries (see Section 3.2 above), separating one unit from the other since they have no real semantic or syntactic value. Sometimes, they may be prolonged or accompanied by a long filler pause, in which case they functionally become equivalent to a pause. Notice the uses of *'and'* in **Units 28.4, 28.5** and **28.8** (Strong Coordination), and the use of *'but'* in **Unit 28.1** and that of *'and'* in **Unit 29** as 'weak coordinators', in the following example.

5.6 (INT II)

JH. 26. 1 well
2 I don't think we'll accept that
3 because you might as well give up..
4 if you.. if you think that change will never happen [..

JA. 27. [mmhmm

28. 1 e:::m... but... I think it's.. really educating.. students here.
2 to realise that e::m.. women.. are as equal as men here..
3 e:::m.. that they do gain access to engineering degrees..
4 and therefore they should be respected
5 and.. should not be treated as some sort of.. weird person sitting in the corner
6 who doesn't..
7 can't really contribute to the course
8 and can't really.. take the pressure of the course or whatever..

29. e:::m.. a:nd I think the Guild of Students can play a large part in this education..

One of the effective measures for distinguishing between the functions of these conjunctions as 'strong' or 'weak' coordinators is that their omission in the former case (strong coordination) would certainly cause

some grammatical or semantic anomaly whereas in the latter they do not cause such an anomaly. See, for example, the effect of omitting *'and'* in **Unit 28.5, and 28.8 of e.g. 5.6** above, in which it is functioning as a strong coordinator, but its omission in **Unit 29** of the same example does not result in an anomaly of any sort.

Another way of distinguishing between strong and weak coordination may be found in the deletion of the subject following the conjunction. In strong coordination the subject of the clause may be deleted if it is co-referential with that of the preceding linked clause without any grammatical or semantic anomaly (see **e.g. 5.6, again, Unit 28.8**), whereas in weak coordination, since the clause that follows the conjunction is supposed to be a new unit and sometimes even a new topic or episode altogether, the subject may not be deleted, and if it is it will result in a grammatically defective unit (see, for instance the effect of deleting the subject of **Unit 29, e.g. 5.6** above).

The coordinating conjunctions have a relatively great number of functions and uses within SES, which makes their description and characterisation rather difficult within one short chapter like the one at hand. For example, it has been noticed that the conjunction *'and'* can be used to fulfil the following functions in addition to that of 'strong' coordination: filler, additive (i.e. also), continuative (sometimes with *'then'*), episodic (i.e. initiating a new episode), summative, resultative (sometimes with *'therefore'*), contrastive and many others. The conjunction *'but'* has also many functions besides that of coordination such as 'filler' and continuation (see e.g. 5.6 and 5.7). *'So'* can also be used as a coordinator in addition to its use as a continuative or summative conjunction (cf. Farag 1986:193).

Chafe (1982:39) takes the conjunction *'because'* to be a coordinator rather than a subordinator, in which case the frequency of coordinated

clause complexes becomes even higher, because, as will be seen in the following sections, this conjunction occurs relatively frequently in SES. While it is confirmed that certain uses of *'because'* and particularly its contracted informal form *'cos'* may behave as a weak coordinator in running SES, functioning sometimes as an initiator of new topics or simply as a 'filler', the majority of other uses have a clearly subordinating function in that **"the information is presupposed rather than asserted"** (Beaman 1984:78). As such it has been included in the present analysis under subordination (see Section 5.4.2.4.1 below) except when it has a coordinative function, in which case it has been included under 'OTHER' coordinators (see Section 5.3.3 below).

No more is going to be said about the various uses of the coordinators in this chapter as this has been dealt with in detail elsewhere (see Farag 1986). Next the relationship between coordination and SC will be discussed. This will be followed by a presentation of the statistical analysis of these uses in both INT and CON before moving to an overall discussion of these results on the lines outlined above.

5.3.2 Coordination and syntactic complexity

The issue which will have to be discussed now is the relationship between the types of coordination distinguished above and the SC of spoken discourse. It has already been mentioned that several researchers refer what they believe to be syntactic 'simplicity' and 'fragmentariness' to the high frequency of coordinated structures obviously present in spoken discourse compared to written discourse (e.g. Kroll 1977; Chafe 1982). However, it should be made clear here that the findings reported by these researchers depend on absolute frequencies of the occurrence of coordinating conjunctions without regard to the context and semantics of the combined units. The distinction 'weak' and 'strong' coordination seems very useful here. Loosely connected units, i.e. those connected by

weak coordination, can certainly be markers of syntactic untidiness and fragmentariness. As discussed above, they do not have major syntactic functions. As such they cannot contribute in any real sense to the issue of SC in general terms. But what about 'strong coordination' ? And how does it relate to the issues of complexity/simplicity of SES ? Evidence from the data of sequential conversational discourse (both from INT and CON) confirms that, although coordination per se as a syntactic relationship between clauses does not increase or decrease the SC of discourse, it has a role in joining clauses and clause complexes that might in themselves be already syntactically complex. Very often in SES speakers join whole series of units of various complexity levels using one of the coordinating conjunctions available to them. This creates long units of discourse that are more difficult to handle than shorter ones. Notice, for instance, the following example from INT IV :

5.7 (INT IV)

JA. 82. 1 I s:: suppose it may be that the lines are only only d.. drawn.. much later on..
2 for example those that you keep up with
3 or.. or.. you know that.. that you're still seeing in ten year's time..
4 something like that
5 and.. maybe the line is in retrospect.. e:::m..

PM. 83. 1 there are groups of people who you can discuss things with.. [
JA. 84. [mmhmm

2 e:::m.. but certainly aren't.. more..
3 not.. more than above ten people
4 who I'd go and discuss things with..
85. 1 there are people who I'd chat with..
2 people ** I'd discuss things with..
3 and there's people
4 who sort of.. perhaps I'd discuss more.. personal things with [..
JA. 86. [mmhmm

Notice that in the above example, the conjunctions *'or'* (**Unit 82.3**), *'but'* (**Unit 83.2**), and *'and'* (**Unit 85.3**) are joining units at various complexity levels. So it seems rather a simplistic point to talk about 'coordination' as a feature of 'syntactically uncomplex' discourse. These claims are mainly made by researchers whose work is done within the field of comparing spoken with written discourse, which could be one of the reasons for their results to point to such conclusions. But certainly one cannot generalise such claims without first examining many differentiated types of spoken discourse, which is why part of this study is devoted to investigating such claims against a body of INT and CON.

Coordination may be taken then to be a syntactic process that is related to SC if it is found to link complex multi-clause units rather than simple single phrases or clauses. Only when the type of discourse investigated contains a greater use of conjoins of the simple phrase and clause structures than multi-subordinated structures one can characterise that type of discourse as 'syntactically simple'. However, I tend to believe that all these are relative standards of complexity and simplicity and that there is more to it than the simple enumeration of coordinated and subordinated structures, as discussed in Section 5.1.2 above and as will be seen in the analysis sections that follow.

5.3.3 Results

For the purpose of calculating the frequencies of coordinated structures in the data of INT and CON, all clause complexes joined by coordinators (strong and weak) were divided into three main subcategories : (1) those joined by *'and'*; (2) those joined by *'but'* and (3) those joined by other coordinators (i.e. all other coordinators except *'and'* and *'but'*, e.g. *'or','either or'*, *'so'*, *'for'*, *'zero coordinator'*, etc. The following figure shows the general frequency of occurrence of these subcategories in INT and CON in relation to each other (raw figures and percentages are to be

found in Appendix II, VOL. II of this thesis) :

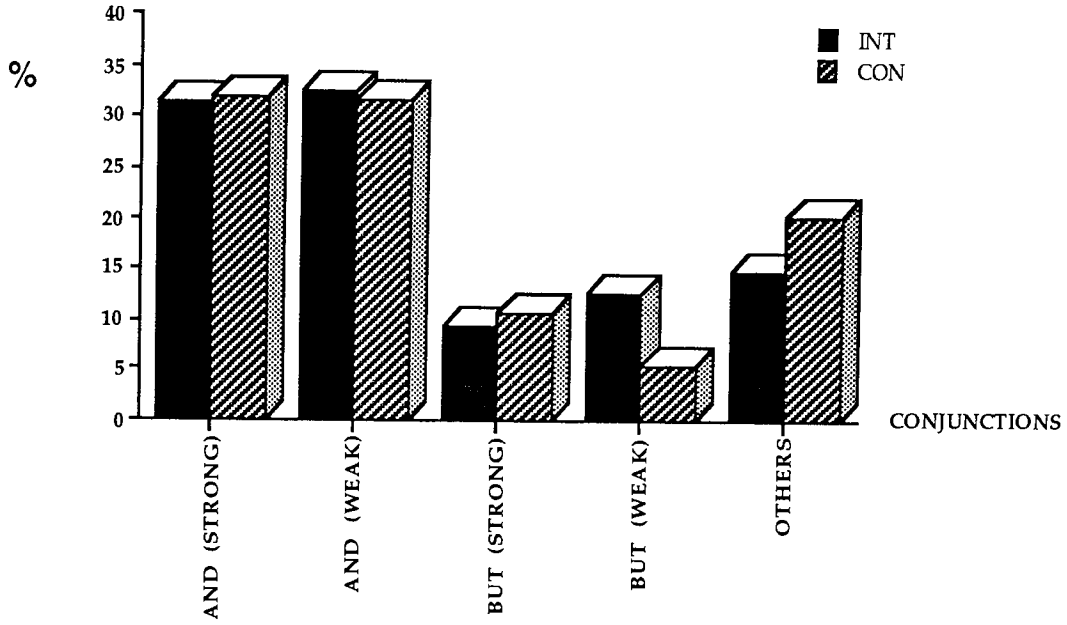


Fig. 5.1 Frequency Indices for Coordinating Conjunctions in INT and CON

The figure shows *'and'* to have the highest frequency of occurrence among other coordinating conjunctions (63.3% in INT and 63.8% in CON altogether for weak and strong coordination), *'but'* has relatively fewer occurrences (21.8% in INT and 16% in CON), and all other conjunctions, including *'zero coordinating linker'* have a frequency occurrence of 14.7% in INT and 20.2% in CON. The above percentages show clearly that at least 50% or more of *'and'* and *'but'* occur as real coordinators and the rest occur in the other functions e.g. fillers, continuation, addition, etc. As far as the differences between INT and CON are concerned, there does not seem to be much difference in the overall use of these conjunctions. Table 5.1 below shows OPTWs of each of these coordinating conjunctions in the speech of each of the subjects in INT and CON :

PARTICIPANTS	AND		BUT		OTHERS	
	INT	CON	INT	CON	INT	CON
HC	13.7	24.8	6.9	8.9	2.3	6.9
KW	24.3	32.8	8.7	5.6	5.2	13.2
JH	32.3	20.9	9.0	10.4	9.0	13.6
BG	42.4	33.1	7.3	8.3	8.8	8.3
RH	25.2	14.8	14.2	7.4	7.9	5.9
DD	34.1	40.9	11.4	12.5	11.4	19.6
RF	30.0	30.7	8.1	6.4	14.2	7.8
PM	20.8	25.5	14.4	15.3	7.2	0.0
HK	17.1	20.3	19.7	16.9	9.2	10.1
KSH	24.3	27.3	19.6	0.0	4.6	10.2
AM	48.7	47.2	3.9	4.6	18.5	9.2
CB	21.3	19.4	8.9	5.7	6.9	8.0
JC	29.8	18.9	7.0	5.0	7.0	7.9
GM	51.0	39.0	4.9	2.4	9.7	9.8
AB	39.1	21.9	9.4	5.5	1.3	3.9
KS	26.4	18.9	8.8	14.8	0.0	8.7
MEAN	30.03	27.3	10.1	8.1	7.7	8.9
SD	10.78	9.25	4.6	4.7	4.7	4.3
LB	24.10	22.21	7.57	5.51	5.11	6.53
UB	35.96	32.39	12.63	10.69	10.29	11.27
V	37.3%	36.9%	45.5%	58.02%	61.04%	48.3%

Table 5.1 OPTWs of clause complexes having 'and', 'but' and other coordinating conjunctions for each speaker in INT and CON

The figures above show that *'and'* and *'but'* occur significantly more in INT than in CON (WSRT $p < 0.05$) for both conjunctions. The figures of *'other'* conjunctions do not show a significant difference between the two discourse types. The results above give collective values of *'and'* and *'but'* in their 'weak' and 'strong' forms of coordination. The results seem to vary uniformly between the two types of speech under investigation. Notice, however, the degree of variation in the use of the three types of coordinatives in general in INT and CON. This is shown by the fairly high values of SD and V calculated for the results above. Whereas there is no significant difference between speakers in the use of *'and'* in INT and CON, there is a significant difference between them in the use of *'but'* and *'others'* in the two styles of speech. The SD figures show that there is greater spread of values in the use of *'and'* than in the use of other

coordinatives examined. The results of the following table show the difference between speakers' use of strong and weak coordination in INT and CON :

PARTICIPANTS	STRONG COORD'N		WEAK COORD'N	
	INT	CON	INT	CON
HC	2.3	6.9	18.3	17.8
KW	10.4	16.4	13.9	16.4
JH	23.3	12.5	9.0	8.4
BG	20.5	20.7	21.9	12.4
RH	22.0	5.9	3.1	8.9
DD	17.6	24.9	16.5	14.2
RF	13.5	15.0	16.5	15.7
PM	16.4	15.3	4.5	10.2
HK	5.3	6.8	11.8	10.1
KSH	13.1	13.7	11.2	13.7
AM	21.4	23.0	27.3	24.3
CB	11.0	10.3	13.0	9.1
JC	21.0	8.9	8.8	9.9
GM	17.0	16.3	34.0	22.8
AB	16.2	14.1	22.9	7.8
KS	7.5	8.7	18.8	10.1
MEAN	**14.9**	**13.7**	**15.7**	**13.2**
SD	**6.3**	**5.7**	**8.2**	**5.04**
LB	**11.43**	**10.56**	**11.19**	**10.43**
UB	**18.73**	**16.84**	**20.21**	**15.97**
V	**42.3%**	**41.6%**	**52.2%**	**38.2%**

Table 5.2 OPTWs of 'strong' and 'weak' coordinated clause complexes in INT and CON

The table above generally shows that nearly 50% of the coordinators used in INT and CON are of the 'strong' type, the other 50% falling under the category referred to above as 'weak coordination'. Notice also the relatively greater variability between speakers in the use of 'weak coordination' as opposed to its great similarity in the use of 'strong coordination'.

Following are the results of the overall uses of all coordinators for each speaker in INT and CON :

PARTICIPANTS	INT	CON
HC	22.9	**40.6**
KW	38.2	**51.7**
JH	50.3	**44.9**
BG	58.5	**49.7**
RH	47.3	**28.1**
DD	56.9	**73.0**
RF	52.3	**44.9**
PM	42.4	**40.8**
HK	46.0	**47.3**
KSH	48.5	**37.5**
AM	71.1	**61.0**
CB	37.1	**33.1**
JC	43.8	**31.8**
GM	65.6	**51.2**
AB	49.8	**31.3**
KS	35.2	**42.4**
MEAN	**47.9**	**44.3**
SD	**11.9**	**11.6**
LB	**41.35**	**37.92**
UB	**54.45**	**50.68**
V	**24.8%**	**26.2%**

Table 5.3 OPTWs of all coordinatives in INT and CON

The results above show that coordinative clause complexes have a slightly higher frequency of occurrence in INT than in CON, but the difference, again, is not statistically significant (WSRT $P < 0.107$). Altogether, the results show that the frequency of occurrence of coordinative clause complexes is relatively high in both INT and CON, which tends to confirm other people's findings in this respect.

5.3.4 Discussion

The analysis of INT and CON in terms of the frequency of occurrence of the coordinated structures that are employed in both types of spoken discourse suggests the presence of a binary distinction interpretable according to what one takes the term coordination to mean (see Section 5.3.2 above). If one takes coordination in the general sense used by

researchers working in the field of language variation (namely spoken versus written discourse variation), then the results show that there is a relatively high frequency of occurrence of these structures in both INT and CON. If, on the other hand, the distinction 'strong' and 'weak' coordination suggested above is maintained the results appear to show a different picture altogether. Discarding those cases of weak and loose coordination, far fewer frequencies are found in spoken discourse types than what other research work in the literature seems to imply. This binary distinction is very important to make for the purpose of getting a clearer picture of the behaviour of coordination in spoken discourse types.

I will first look at coordination in its general sense. The results above show that coordination in this sense (i.e. both weak and strong) occurs quite frequently in the data (Table 5.3 above). These results confirm the view of those researchers who claim that coordination is a characteristic feature of spoken discourse, without which any piece of continuous exchanged speech would lose much of its structural cohesion and syntactic organisation. Comparing the results above to those of Beaman (1984:61), it is found that the language of INT and CON contains less *'ands'* but more *'buts'* than that of Beaman's spoken narrative discourse. The findings in this respect are close to those of Chafe (op.cit.:39). This is expected since the spoken data he has analysed consist of dinnertable conversations which are in one way or another closer to the type of data investigated in the present study.

The high frequency of the coordinators in their general sense does not mean, however, that they have lost their **"specific coordinating function"** as some researchers seem to suggest. Beaman (1984:61), for example, quotes Martinet (1964:177) as saying that **"there is a constant and inverse relationship between the frequency of a unit and the information which it conveys"**. Coordination, as an important syntactic process of linguistic structure is a relationship that holds between phrases and clauses and its

greater frequency of occurrence does not diminish its role; on the contrary, it gives the discourse a characteristic sequential feature distinguishing it from other types of discourse in which its level of occurrence is less (e.g. written types of discourse).

This leads to the second part of the binary distinction made above (i.e. taking real coordination to be represented by what has been termed 'strong coordination' only). The results of the analysis shown above indicate that strong coordination by itself is half as much as the total amount of frequency of occurrence of coordination in its 'general sense' (see above). This casts doubts on the generally accepted view discussed above that spoken discourse types contain a high frequency of coordination. This conclusion is of vital importance when one considers the frequency of other syntactic structures (e.g. subordination) in order to determine SC in discourse. This will be discussed in the final section of this chapter.

'Weak coordination', although it is of great importance in keeping the sequential organisation of discourse and in helping to maintain both fluency and coherence, is not syntactically significant. As has already been stated above 'weak coordinators' can be omitted without causing a serious miscommunication problem, though **"their inclusion adds to the fluency of the text"** (Farag 1986:65).

Turning now to the difference between INT and CON in the employment of coordinated structures, the results do not show a significant difference both in the general (i.e. when both types of strong and weak coordination are considered) and more specific (i.e. when each type 'strong' or 'weak' coordination is considered on its own) senses. The results confirm, however, the presence of a considerable amount of inconsistency among speakers in the employment of coordination in formal and informal settings. Some speakers for example, use coordination twice as much in CON as in INT, (e.g. HC); others use it

more in INT than in CON (e.g. JH, BG, RH), while others produce similar results in both types (e.g. PM, HK). The degree of variability between one speaker and another is observable looking at the SD and V values presented in the tables above.

As far as the issue of SC of the language of INT and CON is concerned, it seems that the syntactically recognised cases of coordination (i.e. strong coordination) do not directly influence the complexity of spoken discourse (especially on the level of speech type differentiation) except in as far as they sometimes link already complex structures (e.g. subordinated clauses) to form longer stretches of discourse which make their analysis rather difficult to handle. Loose coordination helps to achieve sequentiality in SES and in its 'filling' function, it helps speakers to keep hold of the speaking floor and maintain its continuity.

In the next sections an attempt is going to be made towards a thorough examination of 'subordination' in SES. It is only after this issue has been resolved and that of fragmentary units in spoken discourse that one can assess the true complexity of INT in relation to CON, for the same speakers.

5.4 Subordination in INT and CON

5.4.1 Some introductory remarks

Subordination is a state of embedding, which is why most researchers link it with the SC of language. It occurs when **"one clause is made a constituent of another clause"** (Quirk et al. 1985:44). In other words a dependent clause (as defined above) is made a subordinate clause to an independent one, normally by the use of a subordinating conjunction, resulting in what is traditionally known as a 'complex sentence'. The following example from the data contains six cases of subordination :

5.8 (INT IV)

JA. 15. 1 but.. there would be those who say
2 Phil..
3 that.. that as a twenty-year old
4 e:::m.. there's no reason why you should be pampered
5 and.. and have a room.. [cleaned for you..
RF. [*(he he he he)*

6 and kept warm for you..
7 bec.. [because you are not an infant..
Ph. 16.1 [no but..

SES are different from other types of discourse in the employment of subordination in many respects, especially in that one can always find incomplete 'complex sentences', sometimes with other features such as repetition, structural revision, reformulation, insertion of discourse markers and vocatives, left and right dislocators, fillers etc., which add to the complexity of the linguistic output. Notice, for instance, units 19.2 and 19.4 in e.g. example 5.9 below :

5.9 (CON I)

KW. 18. 1 I was talking to this man..
2 who was catching the train down to North Wales..
3 which..isn't terribly far from where I live really..
19. 1 and he was saying he worked in the EEC..
2 which don't e:::m.. you know..
3 would you like to meet some other personnel..
4 I've just..
20. 1 I thought perhaps he might know somebody..
2 he wanted some personnel [..
JA. 21. [mmhmm

Subordinate clauses can occur initially (**Unit 26.2, e.g. 5.10** below), medially (**Unit 18.2 e.g. 5.9** above) or finally within clause complexes (**Unit 15.7, e.g. 5.8** above). The analysis undertaken here covers subordination in all possible positions and includes both finite and non-finite types of subordination.

5.10 (CON VII)

JC. 26. 1 well
2 certainly next year when I'm not catered for
3 I could come to the SCM now and again..
4 so could Richard..
5 I know that for a fact..

According to their syntactic and semantic functions, subordinate clauses are either nominal, adjectival or adverbial(2). Nominal clauses henceforth NCls) perform the function of noun phrases (henceforth NPs), i.e. they function as nominal structures within clause complexes, adjectival (relative) clauses (RCls) function as adjectival phrases i.e. they modify a noun head in the main clause, and adverbial clauses (ADCls) as adverbial phrases i.e. they function as adjuncts or disjuncts within clause complexes. Included under subordination as well are two more major categories which share with those mentioned above similar syntactic features ; they

are appositional clauses (APCls) and comment clauses (CCls). The first one will be discussed with the relative clauses since they, too, have the syntactic function of modifying a NP, and the second one with the adverbial clauses, since they behave as disjuncts in running discourse. The analysis, the results of which will be presented and discussed below, represents calculations in OPTWs of all categories and subcategories of subordination so as one could have an overview of the degree of variability between INT and CON in as far as SC is concerned. All the terminology and classification of features discussed are taken from Quirk et al. (1972, and 1985)(3) except when otherwise stated.

5.4.2 The analysis

First of all it has to be mentioned that all the different kinds of subordination cited above, whether in their finite or nonfinite occurrences, have been calculated in both INT and CON for this study. However, for lack of space, it is very difficult here to present examples and tabulations of all the results obtained. So the approach adopted is to present some examples, graphs and tables with their levels of statistical significance. Full discussion of these results and comparison with other research findings will be provided below.

The initial statistical results show that, excluding interviewer input, there are 1804 cases of subordination in INT and 2159 cases in CON. It is advisable at first to present the percentages of each of the five categories cited above out of the total number so as to have a clear idea of their size in comparison to each other. The following figure shows the percentage of each type in the speech of all the 16 subjects in INT and CON (The raw figures are presented in Appendix II, VOL. II of this thesis) :

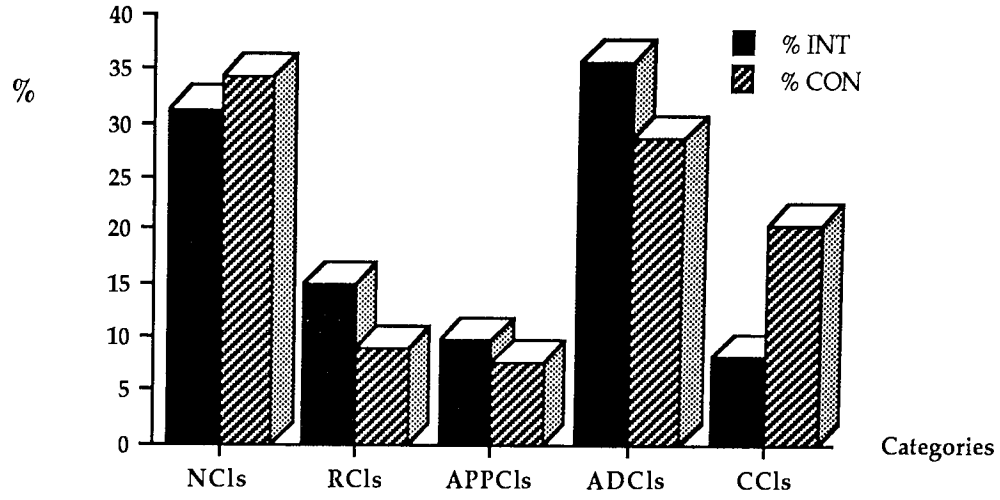


Fig. 5.2 Percentage of Frequency of Occurrence of all types of Subordinate Clauses in INT and CON

The percentages above show that CON is characterised by a greater frequency of occurrence of NCls and CCls than INT whereas there are more RCls, APCls and ADCls in INT than in CON. These results will be commented on in due course.

5.4.2.1 Nominal Clauses

NCls are those that approximate in function to NPs and thus fall into six major categories (Quirk et al. 1985:1048). The first four of these categories are normally finite NCls and the remaining two are nonfinite NCls. These are :

1. that clauses
2. Interrogative Clauses
   a. Wh-Interrogatives
   b. Yes/No Interrogatives
3. Exclamative Clauses
4. Nominal Relative Clauses
5. to-Infinitive Clauses
6. ing-Clauses

This classification has been modified in the present analysis to include three more subcategories to the types mentioned above, which are :

7. Extrapositional NCls
8. Pro-form NCls
9. Others

Extrapositioning has been considered separately from the categories 1 to 6 above for the specific complexity involved in its characteristic syntactic formulation as it involves such operations as 'postponement' and 'substitution' of constituents between phrases and clauses (cf. ibid:1390). These processes result in 'sentences' that are more syntactically complex than their original forms (i.e. before extrapositioning is applied (see elaboration and examples in Section 5.4.2.1.4 below).

The subcategories 'Pro-forms' and 'Others' have been included here as they are (unlike the Extrapositional subcategory mentioned above) used as facilitators of the SC of the spoken output. They are NPs (nouns or pronouns) that replace complex NCls. This will be elaborated on in Sections 5.4.2.1.5 and 5.4.2.1.8 below).

Another modification which was thought to be useful is the exclusion of the exclamative NCls as a subcategory and inclusion of the few examples found in the data with either the extrapositional NCls or the interrogative type where appropriate.

5.4.2.1.1 'that' Clauses (Subordinate Declarative Clauses)

This type of NCls is the most common type in spoken SES as will be seen in the analysis. Nominal declarative clauses may function as :

a. Subject (no example in the data),
b. Direct Object (**Unit 66.1 and 66.2, INT II**),
c. Subject Complement (**Unit 32.3, INT III**),
d. Appositive (**Unit 76.1 and 76.2, INT VII**)
e. Adjectival Complementation (**Unit 1.2, INT VI**).

Of all these subcategories a. and c. are of very low occurrence in the data, but subcategories b., d., and e. have a high frequency of occurrence in both INT and CON.

When the that-clause is direct object or complement, the conjunction *'that'* is frequently omitted except in formal use, leaving a zero that-clause (cf. Quirk et al.:1049) (see examples below). This type is highly recurrent in the data (see Table 5.4 below) and indeed in most types of SES as reported by other researchers (see e.g. Beaman 1984:62).

5.11 (CON VIII)

KS. 86. 1 no..
2 I didn't say I had any ideas..
3 I just said it was clear [*(he he he)*
AB. 87. 1 [well..

5.12 (INT VI)

AM. 106. I don't think there's anything in particular..

The table that follows shows the difference between INT and CON in the frequency of occurrence of *'that'* to *'zero'* NCls in the speech of each of the subjects :

PARTICIPANTS	*'that'* NCls		*'zero'* NCls	
	INT	CON	INT	CON
HC	2.3	0.0	11.4	22.8
KW	0.0	1.9	22.6	22.1
JH	7.2	1.0	16.2	20.9
BG	13.2	2.8	16.1	27.6
RH	7.9	0.0	17.3	5.9
DD	4.1	0.0	15.5	32.0
RF	4.1	5.7	12.2	15.0
PH	4.8	0.0	8.4	5.0
HK	3.9	0.0	18.4	30.4
KSH	3.7	0.0	21.5	41.0
AM	4.9	2.6	25.3	19.0
CB	6.2	4.6	18.5	16.0
JC	0.0	2.0	17.5	11.9
GM	0.0	1.0	7.3	15.4
AB	0.0	1.0	2.7	10.2
KS	12.6	2.7	23.9	24.0
MEAN	4.68	1.58	15.93	19.95
SD	4.08	1.75	6.24	9.76
LB	2.43	0.62	12.49	14.58
UB	6.93	2.55	19.36	25.32
V	87.2%	110.9%	39.2%	45.1%

Table 5.4 OPTWs of *'that'* and *'zero'* Nominal Clauses in INT and CON

The figures above show, first of all that, *'that'* NCls occur highly significantly more in INT than in CON (WSRT $p > 0.01$). The second striking result is that in both INT and CON, the speakers tend to omit *'that'* and use the *'zero'* form instead, but with a significantly higher frequency of occurrence in CON than in INT (WSRT $p > 0.05$). This means that speakers tend to omit *'that'* in NCls in informal conversation and tend to use it more in INT. The MEAN OPTWs obtained from the INT data resembles Beaman's (1984:62) results of her spoken data in the *Pear Narratives* (*'that'* 4.5 OPTWs and *'zero'* 2.1 OPTWs). Notice also that the V values in the table above suggests that there is a considerably great amount of variability in the use of *'that'* , but this variability is less so in the use of *'zero'* form in INT and CON. In both cases however, the degree of variability among speakers is greater in CON than in INT.

5.4.2.1.2 Nominal Relative Clauses

This type of NCl is basically a NP modified by an adnominal relative clause (Quirk et al. op.cit. 1056), hence the name given to it. It resembles interrogative clauses in that its examples are also introduced by a wh-element. They can be distinguished by their mobility to be paraphrased by a NP containing a noun head with general reference that is modified by a RCl (ibid). Notice the following examples from the corpus and how they are paraphrased :

5.12 (CON I)

KW. 36. 1 yes
2 it is..
37. 1 this is why I keep thinking..

(this is the reason why I keep thinking)

5.13 (CON I)

JA. 121.1 what's..
2 what what *(he he)*.. what amuses me::
3 is.....

(the thing that amuses me is..)

This type seems to have low occurrence in the data of INT and CON. Mostly recurrent are those examples that are introduced by the conjunction *'what'* (e.g. 5.13 above). Table 5.5 shows the OPTWs of this type of NCls :

PARTICIPANTS	INT	CON
HC	4.6	**0.0**
KW	3.5	**0.0**
JH	0.0	**0.0**
BG	5.8	**1.4**
RH	0.0	**1.5**
DD	1.0	**3.6**
RF	2.0	**2.9**
PM	1.2	**0.0**
HK	7.9	**10.1**
KSH	1.9	**3.4**
AM	0.0	**0.0**
CB	1.4	**4.6**
JC	0.0	**0.0**
GM	0.0	**0.0**
AB	4.0	**3.1**
KS	3.8	**4.0**
MEAN	**2.32**	**2.16**
SD	**2.39**	**2.72**
LB	**1.00**	**0.67**
UB	**3.64**	**3.66**
V	**103.2%**	**125.8%**

Table 5.5 OPTWs of Nominal Relative Clauses in INT and CON

The results yielded by the table above show that Nominal RCls occur slightly more in INT than in CON but the difference is not statistically significant. There is a considerable amount of variability in the use of this type of NCls by the subjects. Some of them have not used it in their speech at all (e.g. JH, AM, JC, and GM). Others have used it just in INT but not in CON (HK, KSH, PM). This is confirmed by the high V values of the figures in the table above.

Interestingly, Beaman (1984:63) did not find examples of such clauses in her written narratives, but found that their OPTWs in her spoken narratives have a total MEAN of 2.2, which is exactly similar to the MEAN occurrence of these clauses in the overall data. Beaman suggests that speakers use this structure to monitor the needs of the listener and provide the right amount of information (cf. Chafe's involvement factor),

but this is rather doubtful since the low occurrence of this type of clauses in the present data as well as in Beaman's does not support this hypothesis. I tend to agree with Chafe (1982) that this monitoring and provision of information is provided by the use of first person reference and such expressions as *'well', 'I mean', 'you know'* etc, as will be exemplified below.

5.4.2.1.3 Subordinate Interrogative Clauses

This subcategory of NCls recurs in the data of INT and CON quite frequently. It consists of two major types which occur in nearly the same range of functions available to the nominal that-clause exemplified above.

(A) Wh-Interrogative Nominal Clauses

Most of the examples on this type of NCls occur as direct object (e.g. 5.14 below). But they also occur in other different positions (see, for example, **Unit 25.3, INT III** as 'adjectival complementation', **Unit 81.3, INT II and Unit 83.3, INT III** as 'object to a preposition'). Subject position is rarely used in the data and object position being by far the most recurrent type of these NCls . In **Unit 3** of the following example, the Interrogative clause functions as a direct object :

5.14 (INT III)

JO. 4. 1 I think he was rung up yesterd..
2 -- e::m -- the last week..
3 and asked.. how we wished to be addressed..
4 Miss or Ms [..
JA. [(ha ha ha ha)

(B) Yes-No and Alternative Interrogative Clauses

These also occur occasionally in the data but less frequently than type A exemplified above. The yes-no clause is introduced by the subordinators *'whether'* or *'if'*. The alternative clauses are formed with the correlatives *'whether... or'* or *'if... or'*. Most of these 'direct' questions were used by the interviewer as in e.g. 5.18 below (see also **Unit 1 and 2 from INT VII, Unit 63.1 INT III, Unit 235.4 CON IV** as examples on the different positions these types of interrogative clauses occupy in running SES).

5.15 (INT IV)

JA. 1. 1 e::m.. now
2 Robin.. a:nd Phil..
3 you are living out this year..
4 having lived on campus::
5 for the.. for the first year..
2. 1 e::.. it's fairly early on in the term..
2 but I wonder you could::: give your impressions
3 of.. e:::m.. whether in fact it is a.. a.. a.. distinct disadvantage
4 both for work and for social life..
5 or in fact whether you prefer.. living off the campus..

Sometimes the speaker retains the direct question failing to use either *'if'* or *'whether'* quoting the whole question instead, as in the following example :

5.16 (CON I)

KW. 176.1 'cos they don't..
2 they don't even ask you
3 would you like to go to Mass..
4 they drag you along sort of every lunch time..

The following table shows the range of occurrence of these types of NCls in the data of INT and CON. The calculations represent the occurrences of both types of interrogative clauses:

PARTICIPANTS	INT	CON
HC	4.6	3.0
KW	0.0	10.0
JH	1.8	1.0
BG	1.5	4.4
RH	7.9	5.9
DD	3.1	0.0
RF	2.0	2.9
PH	3.6	0.0
HK	7.9	0.0
KSH	12.1	3.4
AM	9.7	5.9
CB	4.8	6.8
JC	5.3	3.0
GM	4.9	7.3
AB	1.3	7.0
KS	2.5	5.4
MEAN	4.6	4.1
SD	3.4	3.0
LB	2.73	2.45
UB	6.47	5.75
V	73.9%	73.2%

Table 5.6 OPTWs of Subordinate Interrogative Nominal Clauses in INT and CON

It has to be recalled here that the results above exclude the interviewer's use of these interrogatives. His use of these nominals is noticeable especially during INT, with a MEAN of 7 OPTWs. The results above also show that, although there is a relatively considerable number of these clauses per thousand words, their use is greatly variable among speakers with a general tendency to occur slightly more in INT than in CON though the difference is not statistically significant. The degree of variability is equally high in the two styles of speech, which is expected, since this feature did not occur in the speech of some speakers at all. Generally speaking, however, this type of NCls occur twice as often as their occurrence in spoken narrative discourse as reported by Beaman (ibid) and that is quite normal since there is more interaction and exchange of views between the speakers than with individual narrators, which increases the possibility of the use of speakers' mental processes during interaction (cf. Chafe 1982).

Included in the calculations above, is the occurrence of what Quirk et al. (1985:1055) refer to as subordinate exclamative clauses. These generally function as extraposed subject (**Unit 118.2 INT VIII**), direct object (**Unit 50.5 CON VI, Unit 31.2 INT I**), or prepositional complement (e.g. 5.20 below) and have the same form as subordinate interrogative clauses introduced by *'what'* or *'how'* (see e.g. 5.17 below) hence their inclusion within the OPTWs in Table 5.6 above :

5.17 (CON VI)

AM. 15. 1 yeah..
2 I don't know what I'm doing wrong
3 but everyone seems to.. to have totally different ideas to me..
4 about.. what to put on the bars..
5 and.. what our aims should be..

5.4.2.1.4 Extrapositional Nominal Clauses

Another category, which is quite recurrent in English SES, hence its inclusion as a separate subcategory here in this section, is the frequent use of extrapositional NCls (normally with anticipatory 'it'). This is mainly recurrent with to-infinitive clauses, 'that' clauses, interrogative clauses, etc. These clauses can be paraphrased using the ordinary NCls they stand for. Following is an example from the data (other examples from the data include **Unit 3 INT VIII, Unit 155.1 and Unit 257.2 CON I**) :

5.18 (INT VIII)

KS. 39. 1 e:::m.. I think if it hadn't been for the Christian Union
2 I'd've found it.. very difficult.. to make friends
3 because I find it.. hard to mix with people..
4 purely because I'm nervous.. about.. meeting people..

(To make friends is very difficult) and
(to mix with people is hard)

Sometimes the extrapositioning stands for a number of clauses with more than one proposition. For example :

5.19 (CON I)

KW. 258.1 it sounds strange..
2 but in lots of ways I really prefer it to be in Lampeter..
3 'cos Lampeter is so quiet..
259.1 it's really nice if you're happy
2 because I know everybody and everything's fine..

The following table shows the occurrence of the Extrapositional NCls in the data of INT and CON :

PARTICIPANTS	INT	CON
HC	6.9	**1.0**
KW	12.2	**1.9**
JH	5.4	**3.1**
BG	1.5	**2.8**
RH	4.7	**8.9**
DD	3.1	**1.8**
RF	18.3	**2.1**
PM	7.2	**5.1**
HK	0.0	**0.0**
KSH	4.7	**0.0**
AM	3.9	**0.6**
CB	4.8	**1.1**
JC	5.3	**3.0**
GM	10.9	**1.6**
AB	5.4	**0.8**
KS	8.8	**3.4**
MEAN	**6.4**	**2.3**
SD	**4.44**	**2.22**
LB	**3.96**	**1.08**
UB	**8.84**	**3.52**
V	**69.4%**	**96.5%**

Table 5.7 OPTWs of Extrapositional Nominal Clauses in INT and CON

The results of the table above show INT style to have a significantly higher OPTWs of extrapositioning than the CON style. This might well be due to the fact that interviewees tend to put more emphasis on what they are trying to express than when they are in casual conversation where the flow of talk runs smoothly and spontaneously. The results are highly significant (WSRT $p < 0.002$). Taking the feature of 'Extrapositioning' as a marker of greater SC, the results tend to show INT to be more syntactically complex than CON.

The results show that CON values are considerably more variable than the INT values, which can also be referred to the absence of this feature from the conversation of some speakers (e.g. HK and KSH).

5.4.2.1.5 Pro-forms in Nominal Clause position

One more subcategory which can be added to the above types of NCls and which represents one of the most recurrent replacements of NCls is the use of *'it'* and *'that'* as pro-forms. These, in addition to other pro-forms, such as pronouns, some time and place adverbials (e.g. *then, there*), the verb *'do'*, and also the conjunctions *'so,'such'*, *'not'* etc. are commonly used by conversants as a strategy to reduce **"grammatical complexity"** (cf. Quirk et al. op.cit.:76) of long clauses and phrases to avoid repetition of such phrases or clauses. In other types of discourse (e.g. written discourse) they are replaced by whole nominal clauses and phrases. There will be another discussion of most types of these pro-forms in CHAPTER SIX below. Some examples from the data include :

5.20 (INT I)

HC. 9. 1 yes
2 I'd agree with that..
(...with what you have said)

5.21 (CON I)

HC. 217.1I mean I think there's an argument there..
(in what we were arguing)

(See also **Unit 28, CON III on *'that'*, Unit 98.4, CON I on *'now'* and *'then'*, and also Unit 70.2, CON II on *'those'*).**

The following table presents the results of pro-forms in NCl positions as outlined above :

PARTICIPANTS	INT	CON
HC	4.6	**5.9**
KW	1.7	**4.7**
JH	9.0	**6.3**
BG	19.0	**1.4**
RH	17.3	**8.9**
DD	2.1	**7.1**
RF	12.2	**5.7**
PM	3.6	**5.1**
HK	2.6	**13.5**
KSH	4.7	**0.0**
AM	2.9	**4.6**
CB	7.5	**6.8**
JC	8.8	**12.9**
GM	4.9	**6.5**
AB	5.4	**14.1**
KS	7.5	**16.8**
MEAN	**7.1**	**7.5**
SD	**5.18**	**4.63**
LB	**3.90**	**4.95**
UB	**10.30**	**10.05**
V	**72.9%**	**61.7%**

Table 5.8 OPTWs of Pro-forms in NCL position in INT and CON

The results above show that the Pro-forms have a great degree of variability (more in INT than in CON) and are not statistically significant on the level of difference between the two contrasted styles, which might

be attributed to style constancy of speakers in using these forms. Only a few speakers tend to display a significant difference between their output in INT and CON (See especially, BG, RH, RF, AB and KS).

5.4.2.1.6 to-Infinitive Clauses

to-Infinitive clauses are non-finite NCls which can function as subject, direct object, subject complement, appositive and adjectival complement (Quirk et al op.cit.:1062). They are quite recurrent in spoken SES especially in the direct object position (Table 5.11 below); some examples from the corpus are represented by **Units 94.4 and 94.5** from the example below :

5.22 (INT I)

KW. 94. 1 I expect once it's completed
2 it will be quite nice
3 I.. I presume is to make it more like York..
4 to landscape it..
5 to make it more attractive to people coming here..

5.4.2.1.7 -ing Nominal Clauses

ing-NCls occur in positions which are quite similar to those of the to-infinitive NCls exemplified above. They are normally used by speakers as a strategy to reduce the SC of discourse (cf. Beaman 1984:66). Their occurrence in the data is displayed in Table 5.9 below. Examples from the data include the following :

5.23 (CON V)

HK 13. 1 I think e:::m the Tag Week was good..
2 e::m getting the amount of discussion..
3 e::m and e::m getting to know people..
4 in that sense..

5.4.2.1.8 Others

Included under this category are all those instances of nominals, mostly in the nonfinite form, which, like pro-forms, would otherwise be full NCls in other types of discourse. They are mostly represented by NPs and prepositional phrases. They seem also to be used by conversants as a strategy towards syntactic simplification of the SES, meanwhile relatively increasing its lexical density since they mostly employ NPs. Following are two examples :

5.24 (INT II)

BG. 81. 1 and we have just got to show that..
2 that we are quite capable
3 and to be respected for.. what we are..
4 our.. our roles in the Guild..

5.25 (CON I)

KW. 157. I just feel sorry for those little boys ** you taught [...
JA & HC [*(he he he he)*

In the following table the frequency of occurrence of all the three subcategories (to-inf, -ing, and others NCls) is presented in OPTWs as they all represent nonfinite NCls.

PARTICIPANTS	-ing		to-Inf.		Others	
	INT	CON	INT	CON	INT	CON
HC	0.0	1.0	2.1	0.0	0.0	5.0
KW	0.0	1.4	5.2	0.0	1.7	6.1
JH	1.8	0.0	0.0	0.0	10.8	8.4
BG	2.9	0.0	0.0	0.0	8.7	8.3
RH	20.4	4.4	4.7	1.5	7.9	3.0
DD	7.2	3.6	2.1	0.0	14.5	14.2
RF	0.0	1.4	6.1	0.0	2.0	0.0
PM	3.6	0.0	0.0	0.0	9.6	0.0
HK	5.3	6.8	0.0	6.8	1.3	6.8
KSH	3.7	6.8	1.0	0.0	1.0	3.4
AM	0.0	0.0	0.0	0.0	4.9	4.6
CB	0.0	0.0	0.0	1.1	5.5	5.7
JC	0.0	0.0	0.0	0.0	1.8	1.0
GM	3.6	0.0	2.4	0.0	0.0	4.1
AB	2.7	0.0	1.3	0.0	4.0	3.9
KS	3.8	1.7	0.0	2.0	6.3	6.0
MEAN	**3.4**	**1.7**	**1.6**	**0.7**	**5.0**	**5.0**
SD	**5.03**	**2.40**	**2.08**	**1.74**	**4.32**	**3.54**
LB	**0.63**	**0.36**	**0.46**	**-0.26**	**2.62**	**3.05**
UB	**6.17**	**3.02**	**2.74**	**1.66**	**7.38**	**6.95**
V	**147.9%**	**141.2%**	**130.0%**	**210.0%**	**86.4%**	**70.8%**

Table 5.9 OPTWs of Nonfinite Nominal Clauses (-ing, to-Infinitive, and Others) in INT and CON

Examining the occurrences above, it is noticed that -ing clauses occur significantly more in INT than in CON (WSRT $p < 0.03$). However, the OPTWs in the data is less than that of narrative discourse in Beaman's (op.cit.) data. As for the to-infinitive clauses, the results show no significant difference between the subjects' use of them in INT than in CON. However, the results are in agreement with Beaman's results for her spoken narratives (OPTWs 2.5). As far as the subcategory 'others' is concerned the results show their types to occur equally highly in INT and CON but with no statistical significance.

The nonfinite types of NCls have a much fewer occurrence in the data than the finite types, a result which points to the presence of SC in SES.

Examining the examples of these NCls in the data of INT and CON, it is noticed that they are normally used by speakers to reduce the influence of SC of otherwise long and complex nominal clauses. However, like the previous results there seems to be a high degree of variability in their use between one individual speaker and another, as shown in the V results. This variability might well be due to the limited size of the data obtained from the speakers.

5.4.2.1.9 Discussion

The overall OPTWs of NCls, finite and nonfinite, reveals that this type of subordination occurs relatively frequently in SES and they are highly recurrent in both INT and CON with a general MEAN of 47.8 OPTWs. The following tables and figures show the status of all finite and nonfinite NCls in INT and CON in OPTWs. All Finite NCls represent the total MEAN of the following subcategories :

a. 'that' NCls
b. 'zero' subordinator NCls
c. Nominal Relative clauses
d. Interrogative (WH- and yes-no) NCls
e. Exclamative NCls

and all Nonfinite NCls represent the total MEAN of the following subcategories :

a. Pro-forms as NCls
b. -ing NCls
c. to-infinitve NCls
d. Others NCls

First of all, the following figure shows percentage of the occurrence of each of the above subcategories in the data in relation to each other :

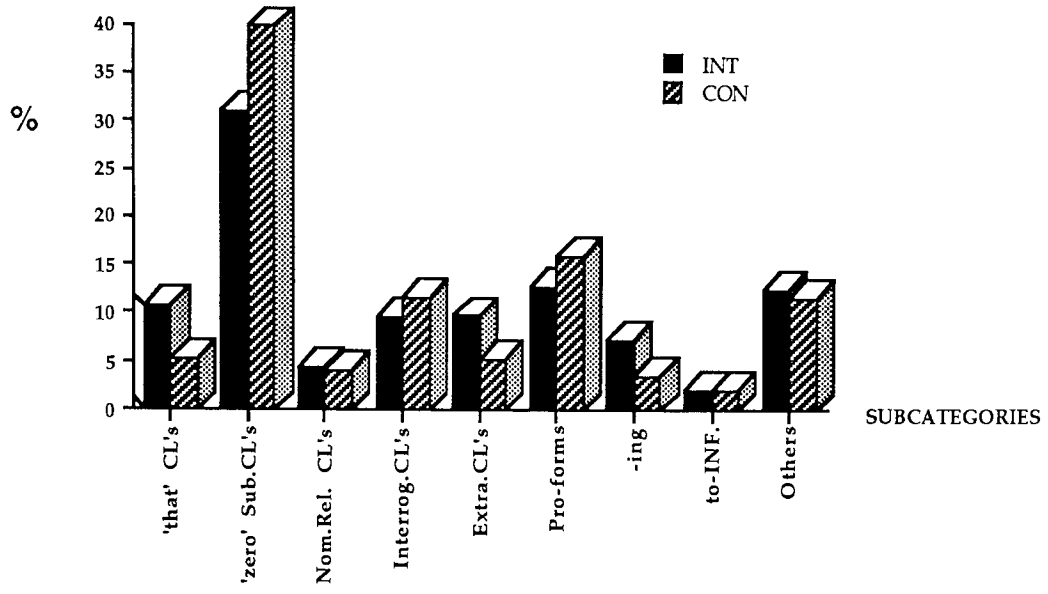


Fig. 5.3 Percentage of All Subcategories of Nominal Clauses in INT and CON

The results of this figure are clearly significant and they yield some interesting conclusions. They clearly show that NCls with *'that'* occur twice as much in INT as in CON, whereas NCls with a *'zero'* subordinator occur significantly more in CON than in INT, which is in agreement with Beaman's results mentioned above. These results in addition to the other indications shown by the greater percentage of extrapositional, -ing and 'others' NCls in INT than in CON show the INT speech to be more complex than that of CON. More complex structures are found within INT than within CON.

The following two tables show the distribution of finite and nonfinite NCls and also the overall OPTWs of NCls in the speech of each subject in INT and CON. The tables show the differences between speakers' output in both types of SES and also the degree of variability across individuals.

PARTICIPANTS	FINITE		NONFINITE	
	INT	CON	INT	CON
HC	29.7	26.8	6.9	11.9
KW	38.3	36.2	10.4	11.9
JH	37.7	26.1	21.5	14.6
BG	32.2	38.7	30.7	11
RH	36.2	22.2	50.3	17.8
DD	26.9	37.4	25.8	24.9
RF	38.6	28.5	20.3	7.1
PM	25.1	10.2	16.7	5.9
HK	38.2	74.3	9.2	33.8
KSH	43.9	47.8	10.3	6.8
AM	9.7	28.2	7.8	9.2
CB	28.1	33.0	8.9	13.7
JC	28.0	19.9	10.5	13.9
GM	23.1	25.2	10.9	10.6
AM	13.5	36.0	10.8	25.8
KS	51.5	39.6	16.3	25.5
Mean	31.3	33.1	16.7	15.4
SD	10.69	14.23	11.33	8.09
LB	25.42	25.27	10.47	10.95
UB	37.18	40.93	22.93	19.85
V	34.2%	43.0%	67.8%	52.5%

Table 5.10 OPTWs of all finite and nonfinite Nominal Clauses in INT and CON

PARTICIPANTS	ALL NOMINAL CLAUSES	
	INT	CON
HC	36.6	38.7
KW	48.7	49.4
JH	49.2	40.7
BG	62.9	49.7
RH	86.5	40.0
DD	52.7	62.3
RF	58.9	35.6
PM	41.8	16.1
HK	47.4	108.1
KSH	54.2	54.6
AM	17.5	37.4
CB	37.0	46.7
JC	38.5	33.8
GM	34.0	35.8
AB	24.3	61.8
KS	67.3	65.1
MEAN	48.0	48.5
SD	17.26	20.30
LB	38.50	37.33
UB	57.50	59.67
V	36.0%	41.9%

Table 5.11 OPTWs of all Nominals (finite and nonfinite) in INT in INT and CON

The results in the two tables above give three important conclusions. First there is a high occurrence of NCls in the data compared to other types of subordinate clauses, with a MEAN of 48.0 and 48.5 OPTWs in INT and CON respectively. Secondly, there is just a slight difference between the two modes of speech with respect to NCls occurring slightly more in CON than in INT. However, the overall results are not statistically significant. Thirdly, there seems to be some variability (slightly more in CON than in INT) among speakers as shown by the V values (see below for interpretation).

It is generally known that NPs, NCls and their combinations and the processes related to them constitute a major set of syntactic criteria on which much of the complexity of English syntax depends. Therefore, their

high or low degree of occurrence can be taken as one of the most important indices of SC, whether in written or spoken discourse. The results presented in the few preceding sections have shown NCls, especially the finite type, to occur highly in INT and CON, which suggests the presence of SC in these two types of SES. The nonfinite type, though recurring much less frequently than the finite type is also an indication for the presence of SC. Speakers tend to use them in an attempt to contract noun clauses into shorter more informationally compact noun phrases, which in turn increases their lexical density. This conclusion seems to be in agreement with Halliday's (1979) argument mentioned above about the SC of spoken discourse in comparison to written discourse.

Looking at the individual differences and variability between one speaker and another a clearer picture in the use of these NCls in INT and CON can be obtained. First of all, more nonfinite NCls occur in CON than in INT but the opposite is true for the finite type. Secondly, there is a considerable amount of variability in the use of NCls between one speaker and another as shown in the high SD figures especially in Table 5.11 above. The results show that the amount of variability is greater in CON than in INT, which is expected since conversational turns tend to be shorter and more fragmentary than INT turns. Or the reason can be (as already mentioned) the limited amount of speech obtained from the informants. More data are needed to make valid generalisations.

Also examining the OPTWs of the subcategories in greater detail shows that NCls introduced by 'that', nominal relative clauses, extrapositional, -ing, to-infinitive, and 'Others' NCls occur more in INT than in CON whereas those NCls that have zero subordinator, interrogative NCls and pro-forms occur more in the conversational encounters. Though these results are mostly not statistically significant, they still show INT to be more syntactically complex than CON. This is shown by the greater tendency of speakers to use more complex NCls

types in the language of INT than in CON; for example, the use of 'that' NCls , 'extrapositional and -ing NCls and so on.

Generally speaking the results show that NCls have a high degree of occurrence in SES both in INT and CON, much higher than the other types of subordinated clauses, as will be seen in the next sections. Altogether, they constitute around 33% of all subordinate clauses, a total of approximately 50 occurrences in each thousand words. This result is highly significant in comparison to the occurrence of other subordinate clauses in SES.

5.4.2.2 Adjectival (Relative) Clauses

Adjectival (relative) clauses represent the first type of **'noun-phrase postmodifiers'** (cf. Quirk et al. op.cit.:1244), the second type being the appositional clauses (see Section 5.4.2.3 below). RCls are those that refer back to a NP, a whole clause or sentence, or even to a whole series of sentences. Normally, they are preceded by a relative pronoun (henceforth RP), but this can be omitted in certain cases.

RCls can be restrictive or nonrestrictive according to the relation of the relative clause to its antecedent. The head in the restrictive case can be linguistically identified only through the modification that has been supplied. The head and the modification are closely connected together prosodically whereas in the nonrestrictive case the head is independently identified and any modification given to such a head is additional information (cf. Quirk et al. 1972:858). Thus, nonrestrictive RCls function as parenthetic comments which usually describe, but do not further define, the antecedent.

RCls can also be finite or nonfinite. Nonfinite RCls can be possibly recurrent with all three types of nonfinite clauses : -ing participle, -ed participle and infinitive participle. Quirk et al. (op.cit.:1263) point out that -ing and -ed types of postmodification correspond only to RCls in which the RP is subject but the infinitive type corresponds to RCls where the RP is object, adverbial, complement as well. Examples from the data will be provided below.

5.4.2.2.1 Restrictive Relative Clauses

A restrictive RCl has a similar function to that of an adjective in that it modifies an antecedent in the main clause with which it has a direct identifying link. In this type of clause the relative pronoun (RP), which can be any of the *'wh-type'*, *'that'* or *'zero'*, can function as subject, object, complement or adverbial, as in the following examples from the data. Notice that one feature which can occur frequently in SES is that the RP *'that'* can be used to replace *'who'* in both the subjective and objective cases, as in example 5.26 below :

5. 25 (CON I)

KW. 62. 1 you either get the people who are really rude to you..
2 put the phone down straight away
3 or put the phone down half the way through saying
4 no..

5. 26 (CON I)

HC. 72. 1 I worked for a department at Edinburgh
2 but I think it's unique
3 in that it it.. collected in furniture
4 from people who wanted to get rid of it..
5 and gave it to other people that needed it..
73. 1 I don't think there's any social department..
2 that's coordinated in.. doing that..

Notice that in the following example five cases of restrictive RCls are detected, all functioning as object to the main clause. They are introduced by *'who'*, *'that'*, *'which'*, and *'zero'* (marked by two asterisks) respectively :

5.27 (INT IV)

RF. 77. 1 these.. these are a whole spectrum though..
2 there's.. there's the one or two..
3 who.. you turn to
4 whenever you need assistance..
5 and there's those that you'd go and enjoy a drink with..
78. 1 and then there's those
2 which you'll.. see from time to time very regularly
3 and then there's those ** you see less frequently..
4 and those ** you really meet on your course..

In the following example the relative pronoun *'which'* is preceded by a preposition (see also **Unit 83.1 INT VII**) :

5.28 (INT VII)

JC 16. 1 one thing ** I could add to that..
2 is about the cleaners coming round..
3 the biggest nuisance of all is the time ** they come round
4 because you're having a nice lie-in..
5 the one day of the week on which you have a lie-in..
6 and the cleaners come round
7 and get you out of bed..

The following examples represent cases of nonfinite restrictive postmodification (see also **Unit 170.6, CON I and Unit 74.1 CON 1**) :

a. -ing participle

5.29 (INT I)

KW. 4. 1 they have people wandering in and out their room..
2 cleaners for instance..

b. -ed participle

5.30 (INT I)

JA. 36. 1 I mean some of the.. e::m authorities might say
2 that a few postgraduates scattered around
3 are are a good stabilising influence to..

c. infinitive participles

5.31 (CON VII)

GM. 162.
2 and I was the only person out of eight
3 to turn up..

5.4.2.2.2 Nonrestrictive Relative Clauses

Nonrestrictive RCls add further information about the head which is already identified and therefore need not be identified by the RCl. In this type of clause, the most explicit forms of RPs are typically used (e.g. *who, whom, which,* etc.). *Zero* RP cannot occur and *'that'* is very rare. Quirk et al. (op.cit.:1257) mention that the **"Nonsubject *'who'*"** within this type is objectionable. The RP can, too, function as a subject, object, complement or adverbial. (For some examples from the data see **Unit 11.2 CON.I, Unit 26.4 INT III**). The following example has both a restrictive (**Unit 18.2**) and a nonrestrictive RCl (**Unit 18.3**) :

5.32 (CON I)

KW. 18. 1 I was talking to this man..
2 who was catching the train down to North Wales..
3 which.. isn't terribly far from where I live really..

5.4.2.2.3 Sentential Relative Clauses

Sentential RCls are different from the two types described above in that the former do not have a NP as antecedent, in which case they do not function as a modifier of a NP. However, they parallel nonrestrictive RCls in that they add more information to their antecedents and are separate from them prosodically as a new 'tone unit' (cf. Crystal and Davy 1969). Like nonrestrictive RCls they can also be paraphrased by a coordinate clause. Quirk et al. (op.cit.:1118) point out that their relative items refer anaphorically to a unit larger than a phrase, usually to a clause but sometimes even to a series of sentences. For these reasons, they have been included in the calculations with the nonrestrictive type of relative clauses, whose OPTWs will be presented in Table 5.12 below.

Sentential RCls are relatively highly recurrent in SES especially with the relative pronoun *'which'*, which seems to be one of the strategies employed by speakers to elaborate and add more information about the topic discussed. The following example contains a series of sentential clauses, which is quite a common feature of SES (see also **Unit 46.3 INT VIII and Unit 76.3 CON II**):

5.33 (INT I)

JA. 78. *(he he he)*.. what about Heythrup..

HC. 79. 1 e:::m.. well
2 it was different for me

3 in that I lived in a Presbytery..
80. 1 I worked in a parish
2 while I was doing.. e:::m my degree..
3 which e:::m... gave it a different aspect..
81. of course it was a philosophy degree and a theology degree..
82. 1 it sort of.. helped to.. to::.. base in reality..
2 which was.. which was interesting..
3 which was quite different from what I am doing here [..
JA. 83. [mmhmm

4 which is management [.. -- --
JA. 84. [mmhmm

In the following example, the pronoun *'which'* is functioning as a prepositional complement :

5.34 (CON IV)

RF. 145.1 I think I was given it on a piece of paper..
2 and I'd stuck it in in my wallet [.. *(he he)*
JA. 146. [mmhmm

3 in which case I have got it..
4 but I might've taken it out of my wallet..
5 and put it in my address book of course--..

Notice that sometimes Sentential RCls can refer back to a whole unit with a multiplicity of clauses and phrases as in the following example in which the speaker is referring to the whole preceding argument :

5.35 (INT VIII)

JA. 81. mmhmm..
82. 1 which leads me on to a next question really..
2 e::m .. if::.. you had somebody here
3 who was:: going to.. start university October..
4 coming up for the first time..
5 as you did last year..

6 e:: what words of advice would you have for them..
7 before they.. even set foot in Birmingham..

5.4.2.2.4 The analysis

The following table shows the OPTWs of restrictive and nonrestrictive RCls in INT and CON. The figures on the nonrestrictive type include the OPTWs of Sentential RCls as well.

PARTICIPANTS	RES.RCls		NONRES. RCls	
	INT	CON	INT	CON
HC	32.0	9.9	16.0	5.9
KW	13.9	5.6	3.5	7.0
JH	12.6	8.4	18.0	6.3
BG	17.5	0.0	17.5	4.4
RH	6.3	7.4	9.4	8.9
DD	21.7	3.6	4.1	8.9
RF	42.7	13.6	2.0	5.7
PM	15.6	0.0	1.2	0.0
HK	6.6	3.4	2.6	3.4
KSH	13.1	10.2	5.6	6.8
AM	13.6	6.6	1.9	3.9
CB	13.7	8.0	3.4	5.7
JC	10.5	7.0	7.0	3.0
GM	8.5	9.8	3.6	0.0
AB	27.0	11.7	5.4	5.5
KS	26.4	3.4	3.8	3.4
Mean	**17.6**	**6.8**	**6.6**	**4.9**
SD	**9.99**	**3.93**	**5.65**	**2.61**
LB	**12.10**	**4.64**	**2.94**	**3.46**
UB	**23.10**	**8.96**	**10.26**	**6.34**
V	**56.8%**	**57.8%**	**85.6%**	**53.3%**

Table 5.12 OPTWs of Restrictive and Nonrestrictive Relative Clauses in INT and CON

The above table yields some interesting results. First of all, it shows the restrictive type of RCls to be considerably more recurrent in both types of SES than the nonrestrictive type. Secondly, there is a significant difference between INT and CON in the OPTWs of restrictive RCls . The results are highly significant (WSRT $p < 0.001$). The difference between them is not

significant as far as the nonrestrictive type is concerned, although there is a tendency for these clauses to occur more in INT than in CON. Thirdly, there seems to be some variation in the use of both types of RCls between the participants both in INT and in CON with a relatively higher variability in INT than in CON for the nonrestrictive type. This is shown by the high V values. These results will be confirmed by the total figures of all types of relative clauses for each of the speakers in Table 5.13 below.

PARTICIPANTS	INT	CON
HC	48.0	**15.8**
KW	17.4	**12.6**
JH	30.6	**14.7**
BG	35.0	**4.4**
RH	15.7	**16.3**
DD	25.8	**12.5**
RF	44.7	**19.3**
PM	16.8	**0.0**
HK	9.2	**6.8**
KSH	18.7	**17.0**
AM	15.5	**10.5**
CB	17.1	**13.7**
JC	17.5	**10.0**
GM	12.1	**9.8**
AB	32.4	**17.2**
KS	30.2	**6.8**
MEAN	**24.2**	**11.7**
SD	**11.5**	**5.3**
LB	**17.87**	**8.78**
UB	**30.53**	**14.62**
V	**47.5%**	**45.3%**

Table 5.13 OPTWs of all types of Relative Clauses in INT and CON

The results of the table above are highly significant (WSRT $p < 0.001$). They clearly show that the language of INT is characterised by the use of a great number of RCls, significantly more than in CON, which, in turn confirms the previous findings that the language of the speakers in INT is more syntactically complex than that of their speech in casual

conversation.

It is interesting to see that speakers use the RPs *'that'*, *'who'*, *'where'*, *'when'*, and *'zero RP'* nearly twice as much in INT as in CON. The RP *'which'* however, is used in both INT and CON, with not much difference. This suggests that the use of *'which'* is a characteristic feature of speech types of various sorts. However, this conclusion needs to be confirmed across many types of SES and in a larger corpus of speech samples.

The following figure displays the percentages of the RPs used by all speakers in the data. No instance of the RPs *'whose'* and *'whom'* were found in the data.

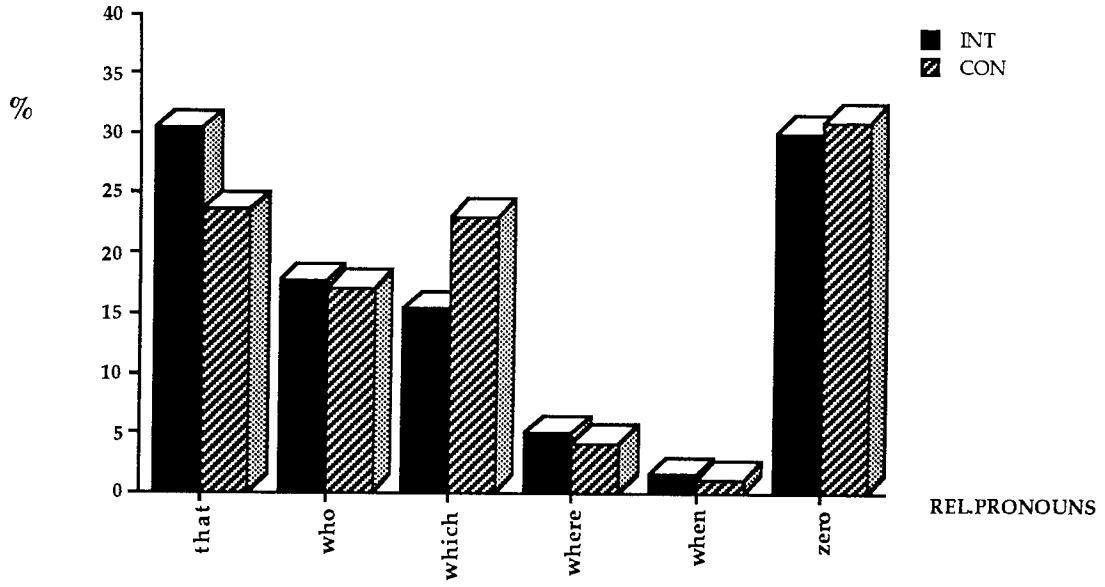


Fig. 5.4 Percentage of Occurrence of the Relative Pronouns in INT and CON

The percentages show the frequency of each RP relative to its total occurrences in the data. Examining the results more closely it appears that RCls introduced by *'that'* and *'zero'* are the most recurrent in the data in both INT and CON, which is a striking result compared to narrative discourse as reported by Beaman (1982:67) who found that RCls

introduced by *'who'* are the most frequent in her data (33%). Such clauses, in the data, have a percentage of 17.7% in INT and 17.1% in CON. Replacing the RPs *'who'* and *'whom'* with *'that'* has been quite common in the data (See examples 5.26 and 5.27 above). *'that'* RCls also have a high frequency of occurrence in Beaman's spoken narratives. But her *'zero RP'* clauses are very much less common in her data compared to their occurrence in the data of INT and CON. In this data speakers tend to omit the RP slightly more in CON than in INT and it seems that they employ the zero RP as often as they use the pronoun *'that'*.

5.4.2.2.5 Discussion

Relative clauses, which have the basic function of NP identification and modification (restrictive) and of giving additional information about an antecedent NP or clause or a sequence of them (nonrestrictive), are clearly recurrent in the language of INT and CON. They represent 15.2% of the total number of clause complexes containing subordinate clauses in INT and 9.14% of their total number in CON (See Fig. 5.2 above). However, the results show that their use varies considerably from one individual to another and from one style of speech to another, with some individuals tending to use them as often as 48 OPTWs (HC. in INT) to nil use (RF. in CON). Most speakers used them significantly more in INT than in CON (13 speakers out of 16). Some speakers, however, showed style constancy (cf. Farag 1986) in their use of these clauses in INT and CON (e.g. RH, KSH, GM).

Comparing the results above with those of Beaman's (1984), it can be seen that the MEAN OPTWs of RCls in CON (11.7) is similar to Beaman's OPTWs of those clauses in narrative discourse (also 11.7), which is expected since speakers in casual conversation very often tend to tell stories, thus employing strategies used by narrators (See particularly CONs I, II, III, VI). During INTs , however, speakers are more concerned to

explain about topics raised by the interviewer's questions and comments that they use too many modifications and add more information for explanatory and clarificatory purposes, hence the recurrence of these clauses within their speech during INT encounters.

In his investigation of dinnertable conversation (as opposed to written discourse), Chafe (1982) reports similar findings to those of the present research. His calculations included just the restrictive type of RCls as he thinks they represent one of the devices by which **"integration"** (i.e. **"the packing of more information into an idea unit"** 'ibid:39') within discourse is achieved. This is why he finds that written discourse contains more of these devices than spoken SES. The OPTWs of restrictive RCls within the latter type are 9.7 (compare to the results in Table 5.12 above, which is just 6.8 for the restrictive type of RCls in CON).

The notion of **"integration"** is important for the purposes here. Chafe claims that writers 'integrate' with their audiences (i.e. readers) through 'integration devices' some of which are nominalisation, participles, attributive adjectives, restrictive relative clauses and so on. It seems though that, taking evidence from the greater frequency of restrictive RCls in the INTs (i.e. 17.6 OPTWs , which is even more than their OPTWs in Chafe's written discourse), the INTs have certain characteristics which tend to approximate to those of written discourse (see also the results on LD reported in the previous chapter). This conclusion needs to be verified by investigating the other 'integration' devices mentioned above.

Finally, it seems that the considerable use of RCls in the data of INT and CON (see Table 5.12 above) confirms that SES, and more particularly INT discourse, display features of SC (cf. Halliday 1979). However, it is hazardous to make such a generalisation at this point before looking at the frequency of all types of subordinated clauses in the data. This will be shown in the following sections of this chapter.

5.4.2.3 Appositive Clauses

Apposition is primarily and typically a relation-ship between noun clauses. It also involves the addition of information to an antecedent NP. In this respect, its syntactic function is similar to that of nonrestrictive RCls discussed above (See Section 5.4.2.2.2 above). Indeed, in the work of some scholars, a nonrestrictive RCl is called **"an appositive relative clause"** (cf. Matthews 1981:229). However, there are major differences between the two types of clauses (see Quirk et al. 1985:1260, for details of and examples on these differences).

It is necessary to mention at this point that apposition does not only occur at the level of clauses and clause complexes but it also occurs widely at the level of words and phrases, which makes it an enormous area of grammatical research which cannot be sufficiently handled in this short section. The concern here is not very much with apposition at the level of words and phrases, but rather at that of clause complexes. Quirk et al's (op.cit.:1300-1321) model will be followed. The reader is referred to explanations made in that work. We will just specify the types of apposition that are normally recurrent in SES, giving examples from the data where necessary.

Apposition can also be restrictive or non-restrictive (ibid:1303). The appositives in the former are in separate tone units whereas in the latter they are in the same tone unit. The former contribute relatively independent information, **"with the first appositive acting as the DEFINED expression, and the second appositive having a DEFINING role"** (ibid). In the latter, on the other hand, the reference of the head in the first appositive is a member of a class which can be identified only through the second appositive. For example, the apposition is nonrestrictive in **e.g. 5.35** but restrictive in **e.g. 5.36** (Examples from Quirk et al.: ibid) (emphasis in

original).

5.35 **Mr Campbell, a lawyer,** was here last night.

5.36 **Mr Campbell the lawyer** was here last night.

In the data investigated here, there is clearly a higher occurrence of the nonrestrictive types of appositive clauses than the restrictive types, as will be seen from the results displayed in the following sections. In the calculations of the nonrestrictive types, Quirk et al's (ibid:1308) semantic scale has been taken into consideration. This runs from **'equivalence'** (i.e. most appositive) to **'loose and unequal relationship'** (least appositive), such as exemplification. Examples from the data on each type of the appositive clauses in the scale are as follows :

1. Nonrestrictive Apposition

(A) EQUIVALENCE :

(Ai) appellation (*naming relation*) : (*that is to say, that is*)

5.37 (INT I)

JA. 62. where where was your first degree..

KW. 63. 1 I did my first degree in Lampeter..
2 in West Wales..

(Aii) identification (*namely*) : The most recurrent type of equivalence apposition in the data.

5.38 (INT II)

BG. 81. 1 and we've just got to show that.. that..
2 we are quite capable
3 and to be respected.. what we are..
4 our.. our roles in the Guild..

Two types of 'identification' within appositional clauses are worth mentioning, which occasionally occur in SES. They are **"postponed"** or **"right dislocation"** and **"anticipated"** or **"left dislocation"** identification. The former **"involves placing a pro- form earlier in the sentence while the noun phrase to which it refers is placed finally as an amplificatory tag"** (ibid:1310), whereas in the latter **"the noun phrase is positioned initially and a reinforcing pronoun stands 'proxy' for it in the relevant position in the sentences"**. Following are two examples on right and left dislocations respectively :

5.39 (INT IV)

RF. 7. 1 if you..
2 I mean you'd know better
3 Phil
4 about.. public transport..
5 that's a real killer..
6 that is..

5.40 (CON III)

RH. 88. 1 m.. m.. my watch..
2 it it stopped two months ago..
3 so.. I took it
4 thinking it needed a new battery
5 and it still didn't work properly..

(Aiii) <u>designation</u>

5.41 (INT IV)

PM. 51. 1 I think that's.. probably difficult for some people.. [
RF. 52. [oh yes..

2 but.. it's the best opportunity ** they've got.. [anyway..
JA. 53. [mmhmm

3 to s.. to::.. begin to make friends..
4 to begin to be.. become more.. outgoing..

5.42 (INT V)

KSH. 4. 1 and I've put Aston fourth on my list
2 'cos I didn't really want to come here
3 just.. the thought of it..
4 of coming to *(he he)* Aston..

(Aiv) <u>reformulation</u> <u>*(rewording)*</u>

5.43 (INT IV)

PM. 101.1 in fact the only people who I..
2 the main group of friends I have now
3 are at the university..
4 here..
5 so.. I haven't got any links e::m in my old home..

This category includes **'mistake editing'**, which is a common feature of spoken SES. For example

5.44 (INT VIII)

KS. 68. 1 but maybe in::.. that you are expected to be in the Guild..
2 in.. I mean in the first month or so anyway
3 at least five times a night..
4 e:: five times a week [*(he he)*..
JA. 69. [mmhmm

(B) <u>Attribution</u>

5.45 (INT IV)

RF. 4. 1 well
2 I think as far as accom.. accommodation goes..
3 it's a lot easier..
4 or much nicer living off campus
5 because.. you got.. a bigger room..
6 usually a more comfortable place to live in [..
JA. 5. [mmhmm

(C) <u>Inclusion</u>
(ci) <u>exemplification</u>

5.46 (INT I)

KW. 4. 1 they have people wandering in and out their room..
2 cleaners for instance..

(cii) <u>particularisation</u>

5.47 (INT V)

KSH. 86. 1 but I think it's missing..
2 home and coursework sort of thing..
3 especially in holidays..

2. Restrictive Apposition

5.48 (V)

KSH. 27. 1 I.. I just sort of came
2 and completely had no real expectations
3 of how good or bad it was going to be..

This type of apposition is common with such general nouns as ***the fact, the idea, the duty,*** .. etc. as in the following example :

5. (CON VI)

CB. 35. 1 I mean I I I disagree with.. with the fact
2 that the Guild could be run like a business sort of type.. [venture..
36. [mmhmm

5.4.2.3.1 The analysis

Again, like that of the RCls , the calculations include both types of APCLs, restrictive and non- restrictive. The following table shows the OPTWs of nonrestrictive and restrictive APCls whereas Table 5.15 gives the results for all types of apposition in the data of INT and CON.

PARTICIPANTS	NONREST. APP.		REST. APP.	
	INT	CON	INT	CON
HC	18.3	21.8	2.3	1.0
KW	17.4	17.4	15.7	0.5
JH	21.5	21.9	3.6	0
BG	21.9	24.9	1.5	4.4
RH	26.7	19.3	1.6	5.9
DD	18.6	33.8	5.2	0.0
RF	34.7	30.7	6.1	1.4
PM	23.9	10.2	2.4	0.0
HK	15.8	67.6	1.3	0.0
KSH	15.9	27.3	3.7	3.4
AM	20.5	10.5	2.9	0.0
CB	21.9	29.7	6.2	1.9
JC	22.8	14.0	7.0	0.0
GM	12.2	12.2	9.7	0.8
AB	24.3	18.8	5.4	3.1
KS	27.6	14.8	1.3	3.4
MEAN	21.5	23.4	4.7	1.6
SD	5.42	13.86	3.79	1.88
LB	18.52	15.78	2.61	0.57
UB	24.48	31.02	6.79	2.63
V	25.2%	59.2%	80.7%	117.2%

Table 5.14 OPTWs of Nonrestrictive and Restrictive Apposition in INT and CON

The table above shows clearly that nonrestrictive APCls occur significantly more in the data than restrictive APCls, the reason being obviously the greater number of subtypes the nonrestrictive type has, most of which have the function of giving additional specific information (e.g. appellation, identification) or of adding nonspecific information (e.g. designation), reformulation, exemplification and the like. All these features are known to occur in SES and the results confirm this to a great deal.

The results show the nonrestrictive type to occur slightly more in CON than in INT, but the results are statistically insignificant. However, there is some variation in the use of this type of apposition on the individual

level, apparently more in CON than in INT. This is shown by the greater V values the informal conversational data has over INT.

The restrictive type of apposition seems to behave differently from the nonrestrictive type though it has a far lower frequency of occurrence. In INT there are more examples of this type than in CON. The results are highly statistically significant (WRST $p < 0.01$). This seems to correspond with the occurrence of restricted RCls (Table 5.12), which in turn suggests that 'restrictedness' is a common feature of sustained and formal speech. Notice also that the restrictive type shows a greater deal of variation among speakers than the nonrestrictive type, with that of CON being significantly more apparent. This is due to the absence of this feature from the speech of some speakers (e.g. JH, DD, PM, HK, AM, JC).

The following table represents calculations of all APCls in the data :

PARTICIPANTS	INT	CON
HC	20.6	**22.8**
KW	33.1	**17.9**
JH	25.1	21.9
BG	23.4	**29.3**
RH	28.3	**25.2**
DD	23.8	**33.8**
RF	40.8	**32.1**
PM	26.3	**10.2**
HK	17.1	**67.6**
KSH	19.6	**30.7**
AM	23.4	**10.5**
CB	28.1	**31.6**
JC	29.8	**14.0**
GM	21.9	**13.0**
AB	29.7	**21.9**
KS	28.9	**18.2**
MEAN	**26.2**	**25.0**
SD	**5.77**	**13.79**
LB	**23.03**	**17.41**
UB	**29.37**	**32.59**
V	**22.02%**	**55.2%**

Table 5.15 OPTWs of all Appositive Clauses in INT and CON

Although there is a great frequency of occurrence of APCls in INT and CON, the results are not statistically significant. There is no significant difference between the two conversational activities except in the degree of variability on the individual level, which is significantly apparent in CON more than in INT (see the V values in the table above).

5.4.2.3.2 Discussion

Again, if it is assumed that SC is manifested in the type and greater number of subordinate clauses within discourse, the results above provide considerable support to the view that SES is syntactically complex, which is clear in the OPTWs of APCls in both INT and CON, slightly more in the

former than the latter. However, I believe that the presence of APCls does not characterise SES as being syntactically complex as much as it provides evidence that because SES is syntactically complex, speakers use clauses in apposition as a strategy to reduce this complexity and facilitate comprehension by the listener. As it has already been mentioned above, APCls, unlike other types of subordinate clauses, represent highly influential syntactic and semantic devices whereby speakers provide more specific information to less specific previously mentioned NPs or vice versa. They are also available for language users as means of simplification, elaboration, exemplification, particularisation, reformulation, and so on. Notice, for example, the following extract from one of the INTs whereby one interviewee is responding to a question put forward by the interviewer :

5.50 (INT I)

KW. 3. I think it's the fact they have no privacy really..
4. 1 they have people wandering in and out their room..
2 cleaners for instance..
5. 1 and when you are asked to move
2 you're given.. absolutely no notice..
6. 1 it's as if.. the housekeeper has no idea
2 that when you're doing a research project
3 that you do actually..
4 'cos.. alter your days..
5 that you fit everything in around it..
7. 1 only when you're given a forty-eight hour notice to move
2 it's difficult to rearrange everything..

JA. 8. mmhmm...

Notice that in this extract, the speaker is providing many elaborative APCls , as an attempt to provide more obvious information or to clarify a point she has initially said. **Unit 4.2** is an exemplificatory APCl, and there are two cases of restrictive apposition in **Units 3** and **6.1**, which elaborate

on the general NPs *'the fact'* and *'the idea'* respectively.

So it seems that apposition can be looked at as one of the important strategies used by speakers to simplify their spoken output and reduce the amount of SC of discourse. As such its greater frequency of occurrence can be taken to be a good indication of the presence of SC. This is why it is often noticed that SES is rich with such appositional devices as right and left dislocations, reformulation, exemplification, clarifications and giving more specific details of previously mentioned NPs and so on. It seems that these strategies are part and parcel of human inherent interactional abilities, which speakers use, in addition to other strategies of discourse processing, to facilitate comprehension on the part of the listener so as to make the conversational encounter as interesting and comprehensible as possible.

5.4.2.4 Adverbial Clauses

Adverbial clauses (ADCls) function syntactically as adjuncts or disjuncts (Quirk et al.op.cit.: 1068). The former denote circumstances of the situation in the matrix clause and they are peripheral to it, whereas the latter comment on the style or form of what is said in the matrix clause or on its content (ibid). From the syntactic point of view, there is no clear manifestation in form or in position of the two functions, for example they mostly share the same subordinator, and they may both be positioned initially or finally. This makes their classification rather complex for statistical purposes. For this reason, the semantic classification will be adopted, as it is more straightforward and more clearly distinguishable than the syntactic one.

Winter (1982:5) thinks that ADCls constitute the central problem in studying subordination in English. This being so it is important for any study of SC, whether concerning written or spoken texts, to find out the

nature and frequency of use of these clauses in relation to other subordinated clauses. One can imagine subordination to have two extremes (cf. Winter , ibid) one representing clauses related to NP description and modification, namely relative NCls, RCls and APCls, on the one hand, and all types of adverbial clauses on the other, the latter having **"a very much more detailed description"** (ibid).

From the semantic point of view, ADCls are classified by Quirk et al. (ibid:1045/1111) into the following categories :

Clauses of time
Clauses of contingency
Clauses of place
Clauses of condition, concession and contrast
Clauses of exception
Clauses of reason or cause
Clauses of purpose
Clauses of result
Clauses of similarity and comparison
Clauses of proportion
Clauses of preference

Because space does not allow one to elaborate on the use of these types in English, there will be just an elaboration on their use in SES as manifested in the data of INT and CON, with examples where necessary(4). In the approach adopted here the subcategories above have been modified so as to include counts of the following subcategories :

Clauses of time
Clauses of place
Clauses of condition
Clauses of concession
Clauses of reason
Clauses of purpose
Clauses of Result
Clauses of Comparison
Others

Quirk et al's 'clauses of contingency' are implicitly included within clauses

of time, place and condition; clauses of concession and contrast have been amalgamated into one subcategory and clauses of exception, proportion and preference have been counted under 'others' as their occurrence in the data is limited and restricted to particular speakers rather than others.

5.4.2.4.1 Characterisation of Adverbial Clauses in the data

The results shown in Fig. 5.2 above show ADCls to have almost the highest frequency of occurrence compared to the other categories of subordination studied in this work. The percentage of occurrence of ADCls out of the total number of subordinated clauses in the data, is approximately 35% in INT and 29% in CON. This is a highly significant result in as far as SC of SES is concerned. But, before making any generalisations about these results there seems to be a need to look at the nature and occurrence of the subcategories within the data in some detail.

Initially, again excluding the interviewer speech from the calculations, there are 639 adverbial clauses in INT and 597 in CON. The following figure presents the percentages of each type of ADCls in both conversational types :

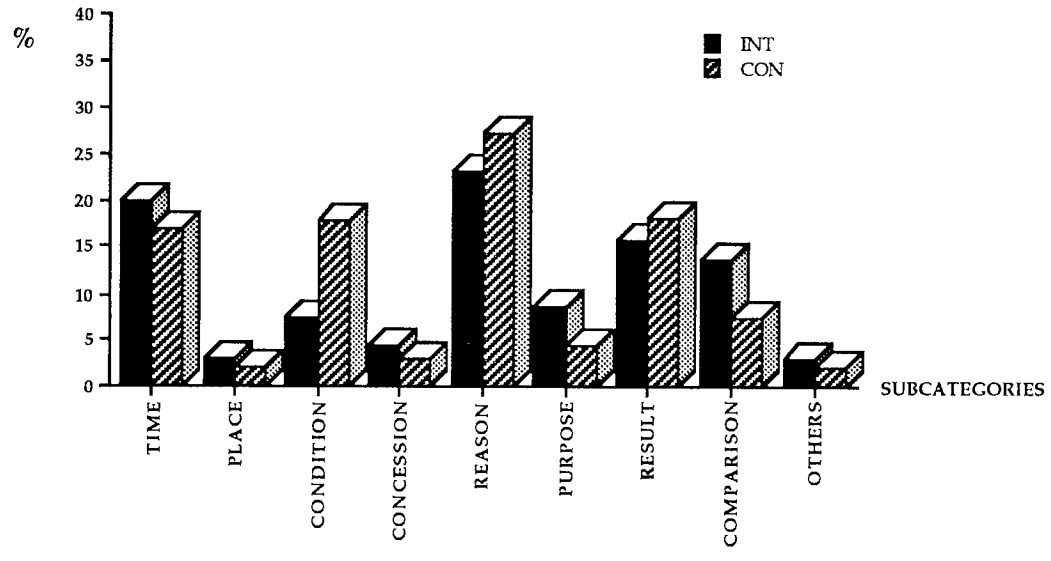


Fig. 5.5 Frequency of Occurrence of All Adverbial Clauses in INT and CON

The above percentages give us three types of occurrences : high, medium and low. Time and reason ADCls are among the highest in occurrence; result, comparison, condition and purpose clauses are among those that have a medium range of occurrence, and finally place, concession and 'others' ADCls represent the lowest types in occurrence amongst the other types. **Time clauses** which are more recurrent than others are the ones introduced by 'when', and less commonly those introduced by *'until'*, *'since'*, *'after' and 'before'*. Nonfinite clauses of time are of high occurrence in the data. Following are two examples from the data on finite and nonfinite time clauses respectively (for more examples see **Unit 96.1 CON I and Unit 15.1 INT II**) :

5.51 (CON I)

KW. 297.1 it..
2 it's really nice when you're there..

5.52 (INT IV)

PM. 23. 1 also the amount of time you waste [.. travelling..
JA. 24. [mmhmm

2 from campus to.. your flat
3 and backwards and forwards like that..

<u>Clauses of reason</u>, which are also among the highest in occurrence, are normally introduced by *'because'* (e.g. 5.53) or often by the contracted form *'cos'* (e.g. 5.54). It is worth recalling here that Chafe (1982) considers *'because'* (and it's contracted form *'cos'*) as a coordinator and studies it along with the other coordinating conjunctions *'and'*, *'but'* and *'so'*. While it is true that *'because'* can act as a weak coordinator in running

SES, a considerable number of its uses have a subordinating function of cause or reason (see Section 5.3.1 above). This is the type that has been counted under subordination in this section. This is clear in the following examples :

5.53 (INT II)

JH. 26. 1 well
2 I don't think we'll accept that
3 because you might as well give up..
4 if you.. if you think that change will never happen [..
JA. 27. [mmhmm

5.54 (INT I)

KW. 46. 1 and.. I find it just..
2 it's so much cheaper to live here..
3 and so much more convenient..
4 'cos everything is around you..
5 whereas if I lived out
6 I'd have to take into account
7 that I'd be paying my own heating expenses..
8 and bus fare as well..
9 'cos I don't have a car[..
JA. 47. [mmhmm

Also recurrent in the data of both INT and CON are examples containing causal relations but having no explicit subordinators i.e. *'zero subordinator'* (e.g. 5.55) or having subordinators which are not commonly used to introduce cause clauses such as *'that'* in example 5.56 below.

5.55 (CON III)

DD. 86. 1 I mean it's..
2 it's really a problem though..
3 you.. you can now buy stuff so cheaply..
4 it's not worth repairing other things [in terms..
JA. 87. [mmhmm

5.56 (INT I)

HC. 79. 1 e:::m.. well
2 it was different for me
3 in that I lived in a Presbytery..

The commonest types of **conditional clauses** are those introduced by *'if'* (e.g. 5.53 and e.g. 5.54 above), **result clauses** are those introduced by *'so'* (e.g. 5.57), *'therefore'* (e.g. 5.58), and *'zero subordinator'*. As far as **purpose clauses** are concerned, the 'that' clauses and the 'to-infinitive nonfinite' clauses (see e.g. 5.59 and 5.60 below) are recurrent in the data. **Clauses of comparison** are mostly those that contain *'than'* (e.g. 5.61). Following are the examples referred to above. Other examples can be found in the data transcripts in the appendices :

<u>Clauses of Result</u> **:**

5.57 (CON I)

HC. 214.1 are we educating them the capitalist system
2 to go back and.. keep.. keep enough people poor
3 so they can make money..
4 or are we saying there's something wrong with us..

5.58 (INT II)

JH. 28. 1 e:::m... but... I think it's.. really educating.. students here..
2 to realise that e::m.. women .. are as equal as men here..
3 e:::m.. that they do gain access to engineering degrees..
4 and therefore they should be respected..

Clauses of Purpose

5.59 (INT II)

JH. 99. 1 e:::m.. but we've got to try and make some progress in certain areas..
2 so that.. even if say next year
3 we have male sabbaticals..
4 e:::m.. the situation doesn't revert to what it might have been before..

JA. 100. mmhmm..

5.60 (INT I)

KW. 94. 1 I expect once it's completed
2 it will be quite nice
3 I.. I presume is to make it more like York..
4 to landscape it..
5 to make it more attractive to people coming here..

Clauses of Comparison

5.61 (INT III)

RH. 17. 1 I think it was::.. higher than.. what turned out to be.. the reality..
2 because I'd been at Sussex University before..

However, it should be noted that these examples represent explicit cases of the categories they represent. SES contain many other examples of the above mentioned and other categories that are not as easy and as explicit as they seem to be, which makes the language of conversation even more syntactically and discoursally complex. Unfortunately space does not allow me to comment on them in greater detail. Further research along these lines is forthcoming.

The use of the different kinds of ADCls exemplified above seems to vary considerably from one speaker to another according to the situation they are in (INT or CON) and also according to the personal style of the speaker. These findings are quite obvious examining the OPTWs of all ADCls results in Table 5.16 below :

PARTICIPANTS	INT	CON
HC	27.5	**36.7**
KW	90.4	**45.4**
JH	55.7	**31.3**
BG	58.5	**64.9**
RH	36.2	**40.0**
DD	20.7	**60.5**
RF	77.2	**37.8**
PM	49.0	**45.9**
HK	68.4	**27.0**
KSH	51.4	**30.7**
AM	47.8	**40.0**
CB	47.3	**36.5**
JC	71.8	**39.7**
GM	65.6	**28.5**
AB	44.5	**36.0**
KS	67.8	**31.5**
MEAN	**55.0**	**39.7**
SD	**18.35**	**10.72**
LB	**44.90**	**33.80**
UB	**65.10**	**45.60**
V	**33.4%**	**27.0%**

Table 5.16 OPTWs of Adverbial Clauses in INT and CON

The results above show the high OPTWs of ADCls within subjects' speech both in INT and CON. However, they also show that INT style is characterised by a greater use of these clauses significantly so with nearly 75% of the speakers. The results are highly statistically significant (WSRT $p < 0.01$). This result clearly confirms the first assumption put forward at the beginning of the chapter that the language of SES is highly syntactically complex.

5.4.2.4.2 Discussion

Although the results above are not directly comparable with those of Beaman (1984), there seems to be some general agreement between the two sets of results. First of all, the results support her finding that time adverbials are by far the most common type of ADCls. However, the main difference between her work and the present one lies in the frequency of occurrence of reason clauses, which have low occurrence in her spoken narratives compared to the findings above (see Table 5.16 above). SES seem to have a greater amount of reason clauses than narrative discourse. The reason for this might perhaps be due to the fact that speakers always tend to provide explicit reasons and justifications for their claims (see examples 5.53 and 5.54 above). It is interesting to observe that Beaman's results show reason clauses to occur more in spoken than in written narratives. In the present results, fewer reason clauses have been found to occur in INT than in CON, which seems to suggest that reason clauses increase with the informal types of discourse and decrease with the formal types.

One more interesting finding which greatly corresponds to Beaman's results discussed above is in the percentage of conditional clauses in both investigations. Beaman's findings show this type of ADCls to have a percentage of 17% in her spoken narratives (10% in her written narratives) compared to the other types of ADCls . The above results (Table 5.16) show that CON style is strikingly similar to oral narratives in that 17.9% of the adverbial clauses are conditional compared to just 7.5% in INT, which also indicates that this type of ADCls occur more with the informal types of discourse.

The general conclusion that can be obtained from the findings above is that ADCls , which represent an important factor in determining the SC of discourse, have a considerable occurrence within SES as manifested in

their OPTWs in INT and CON. Generally INT has more of these adverbials than CON. However, certain types of adverbial clauses seem to be characteristic of informal casual conversation, such as conditional and causal clauses. Time, purpose, and comparison clauses tend to occur more with INT than CON.

5.4.2.5 Comment Clauses

Comment clauses (CCls) represent a collection of parenthetical clauses, which can be finite or nonfinite, whose main function is to maintain the continuity and spontaneity of speech and also to help in softening or altering the effect of the stylistic force of speech units (cf. Crystal and Davy 1975:90). Syntactically, they act as disjuncts, thus resembling ADCls in that they may either express the speakers' comments on the content of the matrix clause or may convey the speakers' views on the way they are speaking (Quirk et al. 1985:1112). They are mostly introduced into speech within separate 'tone units' (Crystal and Davy 1969) and are thus marked prosodically from the rest of the clause complex they happen to be used within, having increased speed and low pitch.

In this research, CCls are introduced along with other categories of subordination as they share many characteristic features of subordination with the other types as will be seen from the types and examples below. According to whether they contain finite or nonfinite clauses two major types can be distinguished(5). The finite type is by far the most recurrent in the data. Within this type CCls may either occur like main clauses, in which case they are mostly confused with NCls (Compare **e.g. 5.62 and 5.63** below), or they may function as ADCls (**Unit 54.2, e.g. 5.64**) or like nominal RCls (**Unit 121.2 e.g. 5.65**). For ease of reference, these three types will be referred to as Type (i), (ii) and (iii), following Quirk et al. (ibid). Tag questions, which are quite recurrent in the data, are included within Type (i) (see **e.g. 5.66**).

5.62 (INT I) (NCl)

KW. 20. 1 I think it's really a big advantage..

5.63 (INT II) (CCl)

BG. 13. 1 I think.. when.. when we first started here..

5.64 (INT II)

JH. 54. 1 I think management student here..
2 as you say..
3 do.. get fairly good positions
4 when they leave..

5.65 (CON I)

JA. 121.1 what's..
2 what what *(he he)*.. what amuses me::
3 is..

5.66 (CON IV)

JA. 6. 1 you're in the Monday groups
2 aren't you..

The nonfinite types, which have low occurrence in the data, may contain a to-infinitive clause [Type (iv)], an -ing clause [Type (v)] or an -ed clause, [Type (vi)] as in the following examples :

5.67 (INT VIII)

AB. 113.1 e:::m.. he wasn't on the same course
2 which sort of.. broke him off us a bit..
3 to start with..

5.68 (INT II)

JH. 6. 1 but e::m.. speaking purely personally
2 I'm very happy this year..

No example of Type (vi) has been found in the data. The following example is taken from Quirk et al. (ibid:1113) :

5.69 Stated bluntly, he had no chance of winning.

5.4.2.5.1 Results

The following figure shows clearly the percentage of each of the above-mentioned types in INT and CON :

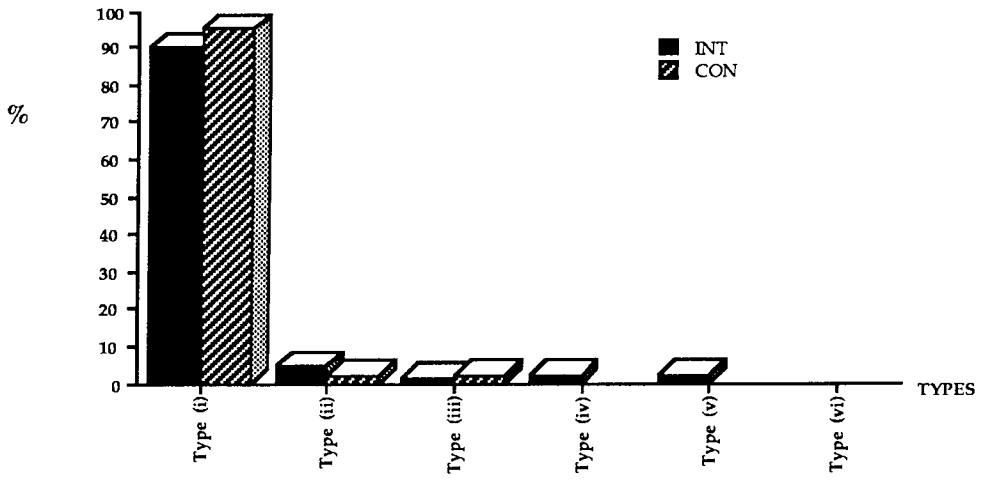


FIG. 5.6 Percentage of Types of Comment Clauses in INT and CON

According to the above results Type (i) is the most recurrent in the data significantly more in CON than in INT. This is expected since the language of informal conversation is known to have a great number of *'you know's', 'I mean's', 'I believe's' 'tag questions'* and the like. The rest

of the types (except Type iii), and most noticeably the nonfinte types seem to accompany more formal discourse. They have no examples whatsoever in the CONs.

The following table gives the OPTWs of all types of CCls in the data of INT and CON :

PARTICIPANTS	INT	CON
HC	22.9	8.9
KW	7.0	18.3
JH	9.0	26.1
BG	4.4	18.0
RH	18.9	7.4
DD	21.7	40.9
RF	8.1	6.4
PM	6.0	10.2
HK	25.0	13.5
KSH	14.0	30.7
AM	9.7	9.2
CB	12.3	6.8
JC	14.0	10.9
GM	8.5	13.0
AB	4.0	12.5
KS	12.6	18.8
MEAN	**12.4**	**15.7**
SD	**6.63**	**9.62**
LB	**8.75**	**10.41**
UB	**16.05**	**20.99**
V	**53.5%**	**61.3%**

Table 5.17 OPTWs of Comment Clauses in INT and CON

The results above show the CCls to occur more in CON than in INT. However, the difference is not statistically significant. Certainly with some speakers the difference is quite significant (e.g. KW, JH, BG, PM, KSH, AB, KS), with others the opposite is true (e.g. HC, RH,HK), while with others there is no noticeable difference. This clearly shows that there is variability in these results, which is also confirmed by the SD and V results. With some speakers the use of CCls in speech can be taken as a characteristic feature of common individual style. It can also be attributed to the level of

formality of the encounter. Recall that all the eight pairs of speakers are close friends and although during INTs a certain level of formality was maintained, there is some degree of informality when the two friends direct their talk to each other during the INT encounters and this might have influenced the results. This will be discussed again in the final chapter of this thesis.

5.4.2.5.2 Discussion

One of the unique features of CCls is that they are mostly used in spoken discourse and more specifically in SES. They do not normally occur in written discourse. They tend to have a more stylistic semantic force than a syntactic one. Their characterisation within the syntactic organisation of discourse is, however, of great importance.

From a purely syntactic point of view, the role of the CCl in speech is not very much different from the other types of nominal, adjectival or adverbial clauses except in as far as they contribute to the issue of SC as outlined above. Although they share, as shown above, some of the syntactic functions of these clauses, they seem to have less influence on the syntax organisation though they might show a greater significant role in the characterisation and overall organisation of speech, as will be shown below.

Obviously, looking at the results above, one cannot make valid generalisations from the limited data under examination and from the limited occurrences of these clauses in the data concerning such issues as those mentioned in the previous paragraph. But it seems clear enough that CCls do not constitute a major source of syntactic complexity. Unlike the other types of clause, they do not represent highly complex structures. Since they represent certain types of parenthetical phrases and clauses, they can occur in various positions within the clause complex, sometimes

interrupting the smooth structuring of other clauses and phrases. This phenomenon can be taken to be a feature of syntactic untidiness and might lead one to think that this can be a feature of syntactic complexity per se. My view is that this is a feature of the complexity of speech organisation rather than its actual structure.

As far as their occurrence in INT and CON is concerned, CCls seem to have a strong relationship with the formality of the context of situation within which they are used. The first type which is the most recurrent type and which has most of the functions mentioned above occurs significantly frequently with informal conversation and less so with the formal types of SES such as INT. The other types seem more to be features of formal styles of spoken discourse as they are absent in the informal variety of SES. These suggestions need, too, to be verified in forthcoming research against a bigger corpus of conversational data.

5.4.3 Subordination and Syntactic Complexity in SES

Summarising the findings of the analysis of subordinated clauses reported in the preceding sections, it can be concluded that the results point to the conclusion that there is certainly a great deal of syntactic complexity in the language of INT and CON manifested in the great frequency of use and complexity of subordinated clauses as opposed to other types of clauses used in the data.

But surely, it is not just the frequency of these clauses that creat SC in SES. The fact that SES is unsystematic in many of the ways which concern the use of subordination and coordination and other syntactic structures makes it highly complex and variable from one speaker to another and from one context to another. There is apparently some difference between the SC of INT and CON as far as subordination is concerned, with certain types the difference being quite highly significant.

Nominal, relative and adverbial clauses occur frequently in both INT and CON. Their presence, in addition to the other features reported above give a clear idea of SC which SES tend to have, variant according to the situation within which speakers find themselves. However, the various results up to now suggest the presence of other features the high frequency of which might play a role in reducing this SC and its influence, especially as far as the listeners are concerned. Some of these features include 'comment clauses', the 'elaborative' types of subordinate clauses (e.g. nonrestrictive RCls and apposition and also sentential RCls), in addition to the weak type of coordination discussed above. Some other features include elliptical units, backchannels, and discourse markers. These will be discussed in the following few sections.

5.5 Elliptical Units

An additional feature of SES that needs to be looked at in relation to SC is that of the 'fragmentariness' of discourse in INT and CON. Since this is one of the major characteristics of SES and its frequent use in SES can influence the results of analysing other linguistic features, it seems equally necessary to examine it in detail along with the other features examined above. Clearly elliptical units (EUs) occur with coordinated as well as with subordinated clauses. And dismissing them from the overall analysis (as some researchers have done) might reduce the significance of research especially that which is wholly devoted to the study of SES. Other researchers (e.g. Kroll 1977, Beaman 1984, Farag 1986) are justified in excluding such units from their analyses, since their work involves comparing both written and spoken discourse, and EUs of the type studied in this section are rarely used in written discourse. However, I believe that EUs have a lot to do with the complexity of discourse and should therefore be included in any study of SES whether comparative or otherwise.

Examining dinnertable conversations in relation to written narratives, Chafe (1982:39) has realised the importance of the feature of 'fragmentariness' of spoken discourse. However, Chafe's description of this phenomenon is not equivalent to what is meant by 'elliptical units' in this study. His characterisation of speech as 'fragmentary' is due to **"the stringing together of idea units without connectives"**, akin to Crystal and Davy's (1969:110) characterisation of it as containing a large number of **"loosely connected clauses"**.

The process of ellipsis per se is taken by many grammarians and linguists as a grammatical process (cf. Quirk et al. 1985:82, Halliday and Hasan 1976:142), and is studied along with other processes of language reduction such as substitution. For Halliday and Hasan, for example, 'Ellipsis' refers to **"something left unsaid... but understood"**. As far as spoken language is concerned, this is too general and broad a category for an investigation of the present sort and hence its analysis on their lines is virtually impossible here and indeed not strictly relevant to the purposes set for this chapter.

The present analysis is concerned with those incomplete units of the data of INT and CON that have some missing element(s) either because of a structural necessity (e.g. subject deletion within clause complexes) or deliberate speaking strategy (e.g. interruption and self-interruption). The concern in this section is not to give detailed theoretical explanations of ellipsis within SES, but just to look at the frequencies of EUs recognised according to certain prespecified criteria (see below) to see how far they contribute to the issue of linguistic complexity of INT and CON.

EUs in the data of INT and CON have been characterised in terms of five different types :

(i) short responses

(ii) appended units

(iii) self-interrupted units

(iv) interrupted units

(v) others

Units involving short responses include those that speakers often use when responding to questions put to them by other speakers or as side comments during continuous talk (e.g. **Unit 64. CON VI**). They are often recurrent in INT and CON. They sometimes involve substitution with some proforms such as *'so'*, *'do'*, etc. (e.g. **Units 174.1 CON I, 4.1 INT II**). Other examples include the use of *yes, no, yeah, not,* some forms of adjectives and adverbs or adverbial phrases (e.g. *true, slightly, probably, not really,* etc. See, for example **Units : 167. CON I, 47.1 INT II, 94.2 INT VI**). Sometimes speakers use such responses to avoid repetition, or to simplify otherwise previously used long and complex syntactic structures. Notice, for instance, the following elliptical responses from the data :

5.70 (INT I)

JA. 68. 1 e:::.. what..
2 where was your first degree..

HC. 69. 1 in London..
2 at Heythrup..

Appended units(6) are those that involve noun phrases added as some form of elaboration (e.g. **Unit 8.2 INT VIII**), emphasis (e.g. **Unit 68.3 CON VIII**), paraphrase (e.g. **Unit 11.4 INT VII**), exemplification (e.g. 5.71 below), specification, and so on. They may occur in a succession of consecutive noun phrases (See for example **Units 55. INT VII, 66.3 INT VII**). Such units, too, occur relatively frequently in INT and CON. The following is an example on exemplification :

5.71 (INT I)

KW. 4. 1 they have people wandering in and out their room..
2 cleaners for instance..

Self interruption occurs when a speaker interrupts his/her own speech for some reason restructuring his/her words in a new way. It may occur in the form of structural revision, reformulation, as in **Units 74.3 and 74.6** of the following example :

5.72 (CON I)

JA. 74. 1 there's a thing in Birmingham [..
HC. 75. [is there..
2 ca::lled..
3 it's done by a thing called.. community transport [..
HC. 76. [oh yeah..
4 e:::m.. which has.. e::m..
5 yes
6 they will collect your furniture if you don't want it..
7 and e:::m [.. e::.. then provided you have a chitty
HC. 77. [mmhmm
from somebody..
8 so they don't sort of get dealers coming in..

Self editing of mistakes made during running discourse may frequently produce EUs of this sort (See for example **Units 43.2 and 68.4 INT VIII**).

5.73 (INT IV)

PM. 19. 1 and.. I wouldn't say..
2 e:::m.. I mean exactly sort of living in hardship..

EUs caused by interruption by a second party while the current speaker is talking produce the fourth type of these units. In the data these are highly recurrent in CON and less so in INT. They are marked in the transcripts by the symbol [for the current speaker and the interrupter as well (See for example **Units 74.1 74.3, 74.7, 75, and 76, and 77 from e.g. 5.72** above).

Under the fifth category of EUs all types of structural ellipsis for example those that involve the deletion of the subject (**Unit 29.4 CON IV**), subject and its verb (**Unit 68 INT V**), complement (**Unit 44.2 INT III**), and so on have been included. Sometimes they do overlap with some of the categories outlined above. When they do, they have been included under just one category. Notice that **Units 14.2,3, and 4** of the following example can be examples on this category :

5.73 (INT IV)

DD. 14. 1 yes..
2 I think quite active in.. in CND
3 and.. reasonably active in.. in Aberdeen..
4 and e:::m.. probably got myself marked down..

All these categories of EUs produced in the data form the basis for the calculation of such units wherever they occur. The results of their frequency of occurrence will be presented in the sections that follow.

5.5.1 Results

The following figure shows the general percentages of the occurrences of the five types of EUs in the data of INT and CON :

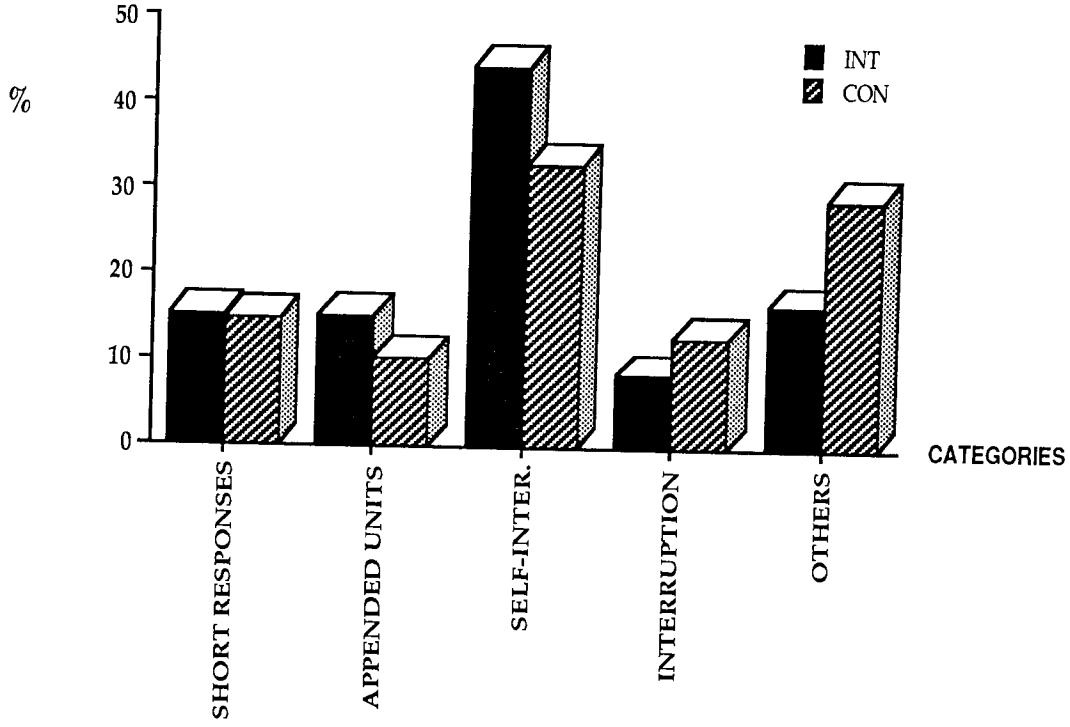


Fig 5.7 Percentage of Occurrence of Elliptical Units in INT and CON

Significantly, EUs occur more in CON than in INT, a result which is expected since casual conversation is known to have this as a characteristic aspect. The general OPTWs for EUs is almost twice as much in CON as in INT (19.4 in the latter compared to 41.2 in the former as shown in Table 5.18 below).

Table 5.18 shows the difference in EUs between each speaker's output in INT and CON and the variability between one speaker and another in this respect :

PARTICIPANT	INT	CON
HC	20.6	33.7
KW	12.2	26.8
JH	26.9	38.6
BG	13.2	29.0
RH	12.6	32.6
DD	15.5	51.6
RF	24.4	35.7
PM	26.3	56.1
HK	15.8	70.9
KSH	10.3	47.8
AM	36.1	24.9
CB	17.8	37.7
JC	15.8	21.8
GM	8.5	31.7
AB	31.0	80.5
KS	23.9	39.5
MEAN	19.4	41.2
SD	7.9	16.5
LB	15.05	32.12
UB	23.75	50.28
V	40.7%	40.0%

Table 5.18 OPTWs of Elliptical Units in INT and CON

The results are self-explanantory. With the exception of just one speaker (A), all speakers have more EUs in their speech during CON than during INT. The results are highly statistically significant (WSRT $p < 0.001$). The degree of variability between one speaker and another is relatively low but similar in both INT and CON as shown by the V figures presented above.

5.5.2 Discussion

Generally speaking, there is little to compare between the results shown above and those of other researchers since, to the best of the present author's knowledge, no one has reported statistical counts of EUs in the spoken data they have dealt with. Those who have compared spoken to written discourse (cf. Kroll 1977, Chafe 1982, Beaman 1984, Farag 1986) have simply ignored these units because of their non-occurrence in

written discourse and because the criteria they have used in the units employed in their analyses (e.g. idea units) do not include such categories. Other researchers whose work is mainly concerned with spoken discourse (e.g. Crystal 1980) have just mentioned this aspect in passing and have not produced statistical counts of frequencies to be compared with those in this investigation.

SES allow EUs to occur more frequently than written discourse and casual conversation seem to be the most likely to have the highest frequency of these types of units. However, as far as the types of EUs distinguished above are concerned, there seem to be some variable results between INT and CON in relation to their frequencies; for example, short responses, appended units and self-interrupted units occur slightly more in INT than in CON whereas the other two types (interrupted units and other types of EUs) occur significantly more in CON than in INT. The reason seems to be related to the functions of these units and the purpose behind the ellipsis. The first three types, short responses, appended units and self-interrupted units are clearly features of INT style rather than casual conversation style. Short responses are the result of speakers responding to other people's questions and queries; appended units, as characterised above, have the main function of extension and elaboration. This is also true of self-interrupted units which speakers normally use to be more specific, clear and corrective of what has been said in preceding units. The other two types, 'interruption' and 'others' are strictly choices made by speakers. Interruptive units are normally caused by listeners rather than speakers and all types of 'others' units are style-specific normally inherent in each speaker's linguistic competence.

It is apparent that most types of EUs are used by speakers, mostly unconsciously, as strategies of linguistic simplification. Therefore, their higher occurrence in a particular type of discourse indicate that complexity is prevalent in that discourse. This does not mean, however, that their

high occurrence increases syntactic (or semantic) complexity of discourse. On the contrary, they normally play a significant role in simplifying any type of linguistic complexity in discourse by using simplified and sometimes substitutory structures, incomplete (yet understood) units, elaborative and exemplificatory appended units and so on.

As far as SC is concerned, and looking at the issue from the point of high frequency of clause complexes (both coordinated and subordinated) in SES, it seems that the high presence of EUs may mean a lower use of clause complexes and vice versa. This does not necessarily mean that speakers use more EUs for the purpose of reducing the number of clause complexes in their speech but it rather means that their spontaneous use of these 'strategies' reduces the number of clause complexes in such a way that it renders their language less syntactically complex, which justifies its high occurrence in SES, and specifically more in the language of casual conversation.

5.6 Some other features of relevance

5.6.1 Discourse Markers and Reaction Signals

Two further issues need to be looked at in relation to syntactic complexity and those are normally referred to in the literature as **"Discourse Markers"** (cf. Schiffrin 1987; Sinclair and Coulthard 1975) and **"Reaction Signals"** (cf. Quirk et al 1985) or **"Backchannels"** (cf. Yngve 1971), respectively. These represent two sets of highly recurrent linguistic items in SES and are thought to have relevance to the issue of syntactic complexity in general terms and to the overall organisation of phrases and clauses within and across units of speech, hence the need to elaborate on their use and frequency in the present analysis. In the following discussion I will be using the terms **"Discourse Markers"** (DMs) and **"Reaction Signals"** (RSs) as they represent the closest possible terms for the purposes

of this study.

Quirk et al. (op.cit.) state that these two groups of items are important because of their high frequency in spoken English. However, they do not study such items in detail in their comprehensive linguistic analysis of English grammar. The reason for this might be that these items are independent of the syntactic structure of speech units (Schiffrin op.cit.:31). As such they have no direct relationship to the syntactic processes examined above. However, their linguistic realisations do tend to have influence on the linguistic complexity/simplicity as a whole. Their frequent use within and among the complex syntactic units of speech might act as a means of softening the effect of SC which the previously discussed results have shown the language of SES to have. This claim will be checked looking at the results of their occurrences in the data of INT and CON.

DMs are vaguely defined in the literature, but for the purposes of this study Schiffrin's (ibid) operational definition **"sequentially dependent elements which bracket units of talk"** has been adopted. She distinguishes (ibid:40) the following as DMs in conversational discourse :

oh, well, and, but, or, so, because, now, then, I mean, y' know.

The main difficulty with these DMs as Schiffrin herself explains is whether to consider them as members of a single word class or of many word classes. The problem becomes even greater in an analysis of the type at hand. This is because some of these items can be confused with certain syntactic functions and structures. For instance, *'and', 'but', 'or', 'so'* and *'because'* can be used as conjunctions and to connect clauses in coordination and subordination. *'I mean', 'you know'* fall within what Quirk et al. (1985:1112) call **"Comment Clauses"** (see Section 5.2.4.5 above), along with other similar expressions such as *'I suppose', 'I wonder' 'I*

expect' and so on. Sinclair and Coulthard (1975:21) distinguish a number of these items as recurrent in classroom interaction calling them **"frames"**. (e.g. *'right', 'well', 'good', 'O.K.',* and *'now'*. They point out that their function is to indicate boundaries in the lesson, the end of one stage and the beginning of the next. Stenström (1984a:81), who also calls them 'markers', mentions that their primary function is to mark the speaker's attitude to what he is saying and to what preceded. Taking these latter approaches into consideration it is difficult to see how conjunctions like coordinators and subordinators can be treated as DMs. This is why in the present analysis, those items that have syntactic functions have been treated under their syntactic categories within the treatment of coordination and subordination. Only those items that function as purely DMs in the senses cited above, along with a number of what is generally known as **'vocatives'** (cf. Stenström 1984a:81) and **'softeners'** (cf. Crystal and Davy 1975:92) (e.g. **'sort of)** have been included within this analysis. These have all been included under the category DMs in the calculations the results of which will be presented in Table 5.19 below.

RSs are those items that are produced by listeners during conversation to express their reaction towards what the speaker is saying. They have the primary function of showing the speaker that they are attentive to what is being said (cf. Crystal and Davy 1975:5). In other words, they are the listeners' feedback to the speaker's messages (cf. Yngve 1970:568, Duncan and Niederehe 1974:237, Stenström op.cit.:13). They include such items as *'mmhmm', 'yes', 'yeah', 'uhuh', 'oh yeah', 'oh no'* etc. They are often accompanied by certain paralinguistic features such as facial expressions, head nods and so on. Again, the aim of including them in the present calculations is to see whether their frequency of occurrence has anything to do with syntactic complexity of SES.

5.6.2 Results and Discussion

Following are the results of the frequency counts made of DMs and RSs for each of the speakers in INT and CON.

PARTICIPANTS	DISCOURSE MARKERS		REACTION SIGNALS	
	INT	CON	INT	CON
HC	16.0	16.8	16.0	38.7
KW	5.2	17.9	10.4	10.3
JH	10.8	35.5	12.6	26.1
BG	5.8	8.3	5.8	5.5
RH	12.6	11.9	4.7	14.8
DD	25.8	30.2	6.2	32.0
RF	8.1	6.4	14.2	24.3
PM	6.0	20.4	7.2	35.7
HK	26.3	16.9	9.2	23.6
KSH	20.4	47.8	10.3	27.3
AM	24.4	23.0	9.7	10.5
CB	15.8	20.5	15.8	37.7
JC	12.3	19.8	8.8	8.9
GM	6.1	16.3	6.1	15.9
AB	16.2	32.8	12.1	38.3
KS	11.3	24.0	11.3	25.5
MEAN	13.9	21.8	10.0	23.4
SD	7.2	10.6	3.5	11.3
LB	9.94	15.97	8.07	17.18
UB	17.86	27.63	11.92	29.62
V	51.8%	48.6%	35.0%	48.3%

Table 5.19 OPTWs of Discourse Markers and Reaction Signals in INT and CON

One of the most significant results of the occurrences above is that both DMs and RSs occur approximately twice in CON as in INT, again with a relatively low degree of variability across speakers compared with other features. DMs occur more in CON than in INT with a statistical significance of 0.004, which is considered to be highly significant. The reason for this result is obviously represented by the frequent use of self-interrupted and self-editing speech, reformulation, elaborative comments and so on, which necessitate the use of DMs. As for the RSs,

they also occur significantly more in CON than in INT (WRST $p < 0.001$). The high occurrence of this feature indicates the high degree of attention and cooperation among the conversants during CON. During INT, most of these back channelling effects have been made by the interviewer as they are characteristic of his speech. He uses them as a strategy to encourage his interviewees to carry on talking. But these were not included in the calculations.

The effects of the high frequency of use of DMs and RSs on the notion of SC in INT and CON do not seem to be clear and straightforward. However some speculations can be made looking at the figures above. They can play a great role in 'softening' the influence of the complexity of the speaker's output by breaking long units of speech into shorter and linguistically simple units. Some speakers employ them more than others. They are, therefore, style markers of speech. Notice, for example, the use of *'well'* in e.g. 5.47 and the use of *'I mean Okay'* in e.g. 5.75 below :

5.74 (INT VIII)

AB. 63. 1 e:::m.. a::nd.. you tend to::.. make friends with people..
2 as soon as you come in contact with them
3 that are a lot more similar to you anyway..
4 and so the pressure isn't necessarily there..
5 well
6 not from my sort of group..

5.75 (INT VIII)

KS. 99. 1 but it's e::
2 I mean Okay..
3 it was more difficult for you [..
AB. [*(he he he)*

Since RSs are the listener's reactions to the speaker's speech, they have no

direct role in the SC of the speaker's output itself. In speaker/hearer terms, they represent one of the links between the listener and speaker and are also used as an effective means by which the listener can signal his/her approval or otherwise of the speaker's message. They can also show the speaker the amount of attention and influence his speech is causing (or not causing) on the listeners; they can even show the speaker whether his/her speech is interesting or boring. However, all these are psychological issues and need not be gone into here as they have no direct relevance to this study. What is important to us here is the way these items are used within the linguistic structure of discourse and how they act as facilitators of the linguistic output per se since speakers are free to use them whenever they like and at any point within the linguistic flow.

One important conclusion can be made here, which is highlighted by the high frequency of DMs and RSs in the language of CON and less so in INT. This conclusion is related to the difference between spoken and written discourse on the level of SC in the two different modes. Writers often have available to them plenty of time to plan their linguistic output in such a way that they can produce, if they want to, the type of language which is syntactically and semantically simple, comprehensible and straightforward. In spoken discourse, and more particularly in SES, this planning time is not available to the speakers and most often speakers produce their conversational units so spontaneously and fluently that they do not even have time to think deeply of how they should be formulating their linguistic output. Therefore, they are liable to produce more complex language with a succession of varied phrases, clauses and clause complexes whose degree of complexity is dependent on many situational and contextual factors. At the same time, it seems that speakers normally try to use whatever strategies they can hold on to make their linguistic output 'uncomplex' and straightforward. The use of DMs and RSs in speech can be considered one of these strategies which speakers use, along with some other strategies that have been mentioned earlier. In SES,

these strategies are influential as they help to simplify the language of speakers caused by the use of complicated clauses and clause complexes. So it seems possible to speculate that their high frequency of use can indirectly point to the presence of complexity within the syntax of discourse, the same as with the other features discussed above.

5.7 Inter-individual differences in SC in INT and CON

Having looked at the results of the analyses of coordination, subordination, elliptical units, and some other linguistic features which may have some influence on the SC of SES, and following the same approach adopted in CHAPTER FOUR above, it is time now to look at the inter-individual variability in the use of syntactic complexity in both speech types. The results displayed in the chapter show that with certain features of syntactic complexity there is a considerable amount of variability between one speaker and another (see for example the V figures of the use of different types of Nominal clauses (see Tables 5.5, 5.6, 5.7, 5.8 and 5.9) ; the Relatives in Tables 5.12 and 5.13 and the restrictive type of apposition in Table 5.14. The most noticeable differences, however, are found between the two groups of PGs and UGs in the use of these features. This is going to be elaborated on as an example on the differences between individuals according to their educational status. Although the size of population of each group is limited and may not yield valid generalisations, the results can certainly point to the general tendency of speakers to use syntactic structures. It can also give clues as to the presence (or absence) of the influence of such factors as age and educational status.

First of all it must be recalled that in all the tables presented above, the first six speakers are PGs and the rest are UGs and it is easy to examine the differences in the output of the two groups of speakers in each syntactic category or subcategory should one wish to do so. However, to make the comparison easier, the general OPTWs of all the features discussed above

have been grouped into one table in an attempt to assess the influence of educational maturity on the general output in INT as opposed to CON. This is shown in the table that follows :

FEATURES	POSTGRADUATES		UNDERGRADUATES	
	INT	CON	INT	CON
SINGLE Cls	13.9	**23.4**	15.2	**19.0**
COORDIN's	43.9	**44.4**	46.8	**47.9**
NCls	54.2	**45.5**	44.2	**49.6**
RCls	26.9	**16.0**	21.6	**14.0**
APP Cls	30.5	**27.8**	27.5	**25.0**
ADV Cls	44.4	**42.8**	52.9	**34.1**
CCls	13.1	**19.0**	10.9	**13.4**
ELLIPTICALS	16.6	**40.0**	19.5	**39.1**
RSs	25.7	**28.3**	21.5	**28.6**
DMs	14.4	**26.1**	15.2	**21.5**

Table 5.20 MEAN OPTWs of all syntactic features for Postgraduates and Undergraduates in INT and CON

The level of difference between the more educationally mature students (PGs) and the less so (UGs) is determined by the variability in each group's employment of the syntactic features under analysis. First of all, the total OPTWs show that PGs tend to use more syntactically complex structures than UGs on a relatively high level of significance. This is especially manifested in the use of NCls, APCls, CCls. PGs also use more single clauses and ADCls than UGs, but the difference is not statistically significant. The only categories that are used more often in the speech of UGs than that of the PGs are the coordinatives and the ellipticals, which confirms the above conclusion as well.

As far as the differences between INT and CON in each group's employment of these syntactic structures are concerned, the results above seem to suggest that the differences are not uniform throughout. There seems to be a scale of three levels of difference ranging between 'no difference at all' (viz. coordinatives), to 'very big difference' (viz. nominal

and adverbial clauses). Between these two ends students differ, to a certain extent, in the use of such features as Single clauses, RCls, APCls, CCls, ECls, RSs and DMs. Sometimes PGs produce differences in their INT and CON outputs on parallel lines with the UGs. For example, both PGs and UGs use more RCls, APCls, ADCls but fewer single clauses, coordinatives, CCls, ellipticals, RSs and DMs in INT than in CON. This confirms the earlier results that, generally speaking, INT style differs from CON on the level of SC. Of course, the degree of difference increases or decreases according to certain factors of which the speakers' level of education is but one.

The results reported in the table above confirm the findings in CHAPTER FOUR above concerning the issue of lexical complexity of INT and CON. There the results suggest that INT has more lexical density than CON, and the output of PGs is more lexically dense than that of the UGs. The conclusion then (see Section 4.5 above) was that PGs language viz. their LD is influenced by the level of education these students have over UGs . The results in this respect tend to go on similar lines. They also show that the language of PGs is influenced by the level of education they have tending to make it slightly more syntactically complex than that of the UGs. This does not mean, however, that the UGs language does not have SC. As it has been seen in the sections above, the results confirm the presence of SC in the data in general as displayed by the relatively great and variable use of complex syntactic structures discussed above. This will be elaborated on in the conclusions below.

5.8 Summary and Conclusions

Considering first the two categories of data analysed (i.e. INT and CON), it is clear that syntactic complexity is present in both. This complexity is evident not only in the higher frequencies of embedded and subordinated structures but also in the presence of other linguistic

features, which although they have no direct link to SC, act as strategies which speakers and listeners alike use to facilitate the spoken output.

Looking at the results from a global point of view, the following table provides a summary of the MEAN OPTWs of all broad features studied in this chapter :

STRUCTURES	OPTWs		STATISTICAL SIGNIFICANCE
	INT	CON	
STRONG COORDINATION	14.9	13.7	P < 0.47 *
WEAK COORDINATION	15.7	13.2	P < 0.08 *
NOMINAL CLAUSES	48.0	48.5	P < 0.45 *
RELATIVE CLAUSES	24.2	11.7	P < 0.001 ***
APPOSITIVE CLAUSES	26.2	25.04	P < 0.12 *
ADVERBIAL CLAUSES	55.0	39.7	P < 0.01 ***
COMMENT CLAUSES	12.4	15.7	P < 0.12 *
ELLIPTICAL CLAUSES	19.4	41.2	P < 0.001 ***
DISCOURSE MARKERS	10.0	23.4	P < 0.04 ***
REACTION SIGNALS	13.9	21.8	P < 0.001 ***

Table 5.21 Total OPTWs of ALL Syntactic Features in INT and CON

* *Not Statistically Significant.*
** *Statistically Significant.*
*** *Highly Statistically Significant.*

Although it is realised here that global results per se cannot reflect the real type and nature of linguistic complexity present in the two categories of SES investigated particularly with the limited data at hand, they tend to point to the general conclusion concerning the presence or absence of SC in these spoken discourse types. They can also give general hints of the inter-individual variability as well. The results of the table above indicate clearly the presence of complexity in both types of spoken discourse studied. Whereas coordinated structures have a total OPTWs of 30.6 in INT and 26.9 in CON, subordinated structures have a total OPTWs of 165.8

in INT and 140.6 in CON. The results are highly significant. They confirm the conclusion made earlier that the language of SES (represented here by INT and CON) is on the whole syntactically complex. These results confirm the first (but not the second) part of Halliday's (1979) claims about the complexity of conversational discourse expressed in the following statement : **"In linguistic terms, spoken language is characterized by complex sentence structures with low lexical density"**. The language of INT and CON contain various complex structures, the level of complexity is variable in accordance with the type of structure used and the purpose of speech at the time the utterance is made.

The second part of Halliday's claim that the syntactic complexity of conversational discourse is always accompanied by low lexical density in the spoken output is flawed by the results obtained from the analysis conducted in CHAPTER FOUR above. The results have shown that both INT and CON have a high level of LD ranging between 45% and 55%, yet, as explained above, they are syntactically complex as well. Of course, this difference might well be due to the difference in terminology between what Halliday takes syntactic complexity and lexical density to mean and between what they are taken here to mean. This point will have to be checked against various types of conversational discourse.

Also on the global level, the language of INT and CON seems to be characterised by a relatively great number of incomplete (i.e. elliptical) units which are partly caused by the use of interruptive and self-interruptive speech, short responses and appending units. Sometimes the greater use of certain discourse markers and reaction signals accompany these elliptical units giving the impression that conversational discourse is fragmentary, elliptical and disintegrated. But it is the belief of the present researcher that the greater reliance of speakers on the use of these interrupters, markers and items is in itself an indication of the complexity of spoken discourse. They are merely used by speakers and

listeners alike as strategies to reduce the effects of syntactic complexity accompanying speech.

Turning now to the difference between INT and CON, it is clear that some of the syntactic structures (which in this research have been assumed to be markers of SC) (see Section 5.1.2 above), tend to occur more in INT than in CON sometimes with a noticeably high level of statistical significance. Examples include the greater use of relative, appositive and adverbial clauses. Notice, however, that the category of nominal clauses shows no significant difference in its use in INT and CON.

As far as the category of coordination is concerned, there does not seem to be too much difference between its global use in INT and CON. This points to the more intra-subject stability in the use of this feature in formal and informal speech especially on the level of strong coordination. A greater type of variability in INT and CON is obtained from the general results of elliptical units, discourse markers and reaction signal items. These occur almost twice as much in CON as in INT with a high level of statistical significance.

The findings that, on certain levels of linguistic complexity, the language of INT is slightly more syntactically complex than that of CON does not seem to correspond with some other findings of researchers in the literature. Milroy (1980:68), for example, has found that **"Interview Styles"** are characterised by structural simplicity in contrast to what she terms **"spontaneous styles"** of which casual conversation is a distinctive type. The results do not go in parallel lines with Halliday's (1979, 1985b) and Beaman's results that the spoken 'informal' types of discourse are more syntactically complex than the 'formal' written types of discourse. Although the types of data analysed here are not directly comparable with those of Halliday and Beaman, the results reported above seem to indicate that the more formal type of data (i.e. INT) tends to show greater SC than

the language of the informal type (i.e. CON). This does not mean, however, that the language of CON is syntactically simple. It rather means that at certain levels of SC, the language of formal speech appears to be slightly more syntactically complex than that of informal conversation.

The difference between the output of speakers in INT and CON on the level of syntactic complexity seems to the present researcher to be due to two main factors: First, it might be due to the degree of monitoring and self-consciousness the speakers have during the INT talk and their relative absence in casual chat. Interviewees seem to be tied during interviews to such situational factors as the presence of the interviewer himself, the topic they are dealing with and the recording process. These factors seem to make the interviewees more aware of what they are saying and more conscious (than in casual CON) of their talk and causes them to use a more **"integrated"** (cf. Chafe 1982:39) type of language. Evidence for this is provided by the findings (see Table 5.21 above) that in INT speakers use less elliptical speech, discourse markers, and interruptions. Secondly, there seems to be good reason to believe that speakers shift the style of their speech when talking to the interviewer from the normal and more natural way they are used to when talking to inmates and friends. This phenomenon seems to influence the linguistic output of speakers in the use of both lexical items (see Section 4.5 and Footnote 7 CHAPTER FOUR above) and syntactic structures and causes variation on various levels of complexity in different speech situations. So it seems that most of the differences between the speakers' talk in INT and CON are due to their shifting their speech style to fit the more formal style of the interviewer whereas they kept their natural way of talking during their casual chat with their fellow friends in the CON encounters. This can also be an explanation for the relatively different results obtained from examining the speech of the more and the less politically and educationally mature students (i.e. PGs and UGs) (See Section 5.7 above)(7).

The conclusions that have up to now been drawn are not just based on observing the general results of the total MEAN of the percentages and OPTWs of the broad syntactic categories (i.e. on the global level) collectively reported in Table 5.21 above. They are rather based on a very close and detailed examination of all the subcategories of the syntactic structures and their behaviour for each of the speakers. The analysis that has been carried out in the preceding sections essentially covered all the subcategories of the syntactic features studied and in the speech of each participant (excluding the interviewer). This has been done in order to obtain a clearer view of the intra- and inter speaker variability on all levels of syntactic analysis. The results and conclusions are discussed at the end of each major section of the chapter.

FOOTNOTES TO CHAPTER FIVE

1. See Halliday (1985a; Chapter 7 : 252) for explanations of these terms.

2. Winter (1982:51), however, points out that there are six kinds of subordination according to their syntactic functions. The first three are : adverbial clause, relative clause and noun clause. The fourth is confined only to non-finite verbs, and the fifth and sixth have in common a subordination whose semantics is like that of the relative clause but which must be contextually distinguished from it. These last two are the special operations clauses of cleft and pseudo-cleft sentence, and the interpolation adjunct.

3. The reader is referred to Quirk et al. 1985 for more details on these features and especially the following pages :

   Nominal Clauses 1048-1067
   Relative Clauses 1244-1260
   Appositional Clauses 1300-1321
   Adverbial Clauses 1068-1118
   Comment Clauses 1112-1118

4. See Quirk et al. (1985:1045/1111) for details and examples on these clauses.

5. For details of these types see Quirk et al. (1985:1112ff).

6. The term **'APPENDED'** is borrowed from Quirk et al. (1985:911).

7. Compare these results to those of CHAPTER FOUR above concerning the lexical density in INT and CON.

CHAPTER SIX

THE STRUCTURE OF INFORMATION IN INT AND CON

6.0 Introduction

In the previous two chapters a linguistic analysis of the most salient features related to the two formal levels of lexis and syntax has been carried out within the units of discourse in INT and CON. The aim was to investigate the debatable issue of structural complexity on these two levels in spoken discourse and their variability between the output of speakers in INT and CON. In this chapter, the analysis will be taken a step further towards investigating linguistic complexity on a more discoursal and pragmatic level of analysis, to talk about, among other things, the issue of information organisation and structure on the linguistic level. This essentially involves decisions related in one way or another to the shared assumptions between speakers and listeners as they jointly cooperate in creating and exchanging information. The study of the linguistic representation of information structure presented in this chapter is based on an extended and modified version of Prince's (1981) taxonomy of information status as it has been seen as the most suitable taxonomy for the analysis of information within SES. However, in order that the discussion is put into proper perspective, I will begin by presenting and discussing some clarificatory preliminaries and some of the available approaches to the study of information structure, some of which have been briefly commented on in CHAPTER ONE above. The aims are first to show how relevant these approaches are to a theory of linguistic representation of information structure within spoken discourse types ; secondly, to point out their limitations in relation to the present study and thirdly and perhaps most importantly, to choose a suitable framework within which the analysis of the linguistic realisations of information structure in INT and CON can be placed. But before that, and

for the purpose of avoiding any confusion in the use of terminology related to the general issue of information structure and organisation and the model of analysis adopted, the next sections will be devoted to clarifying some of the general relevant terms and issues leaving the other more specific ones to be elaborated on as they appear in the discussion of the approaches proper.

6.1 Some preliminary considerations

6.1.1 Terminology

The term **"INFORMATION STRUCTURE"** (henceforth IS), is normally used to mean the packaging of information and its organisation in terms of the distribution of **'DISCOURSE ENTITIES'** (cf. Prince op.cit.:235, and definition below) in their **'NEW'** or **'UNKNOWN'** status as regarding their **"GIVEN"** or **"KNOWN"** status (see Section 6.3.1 below for definitional comments) during information conveying and exchanging. This is a generally accepted definition (cf. Halliday 1967, 1985a; Brown 1980; Brown and Yule 1983) and it has relevance to both linguistic and cognitive psychology (cf. Prince 1981:233). Its relevance to cognitive psychology is manifested in that it deals with such matters as speakers' beliefs and assumptions about their listeners' knowledge at the time the information is exchanged and processed. This is not in the main the concern of the present chapter. What is of interest to us in this chapter is the 'linguistic relevance' to information structure not its psychological one. It is mainly the study of the linguistic realisations of the 'discourse entities' in their Given, Inferable and New status of occurrence within the units of information conveyed and exchanged between speakers and listeners; it is also the study of the various types of relationships that hold between these linguistic realisations (i.e. co-referential relationships : anaphoric, cataphoric, exophoric (cf. Halliday and Hasan 1976:14-18). The linguistic realisations of these discourse entities constitute the essence of the

structure of information and hence their frequency of occurrence and co-referential distribution can best determine the complexity or otherwise of discourse on the informational level of analysis.

The expression **"DISCOURSE ENTITY"** is used in the sense of Prince (op.cit.:235) to refer to **"a discourse model object...; it may represent an individual (existent in the real world or not), a class of individuals, an exemplar, a substance, a concept, etc."** As such, from the linguistic point of view, these discourse entities are represented by Noun Phrases (NPs) in a text. The study of information structure in this chapter, then, will be confined to the study of NPs of INT and CON (except where otherwise stated) and the relationships that hold between them.

The expression **"INFORMATION LOAD"** has already been touched upon in CHAPTER FOUR above. On the formal linguistic level of analysis carried out in that chapter the expression was used to refer to the content and packaging of information within the units of speech by virtue of their higher lexical densities. This view will be adopted but extended in this chapter to cover the functions of discourse entities as Given, New or Inferable from whatever surrounding source. The extension in the definition of the term is necessary as it will be the basis for the forthcoming argument on the status of information within INT and CON.

6.1.2 Information status in SES : An illustrative example

It is generally accepted that the primary function of language in face to face interaction is communication and communication involves the exchange of information between interlocutors. It seems easy to imagine this process to have three main, self-explaining, consecutive phases. The Production Phase, The Transmission Phase and the Reception Phase. The whole process can be simply sketched as follows :

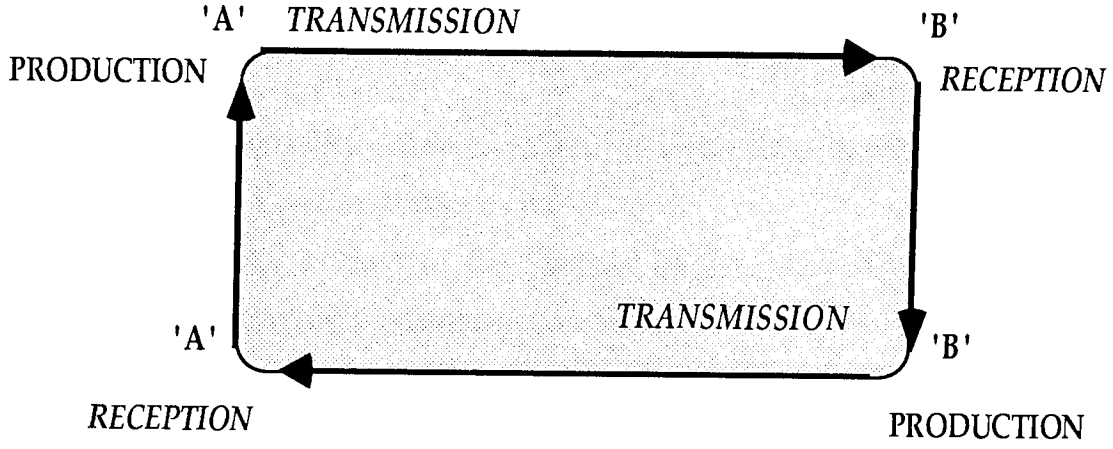


Diagram 6.1 Phases of Information Production and Exchange

The outcome of this dynamic process (Brown and Yule 1983a:23) is that information is conveyed and exchanged on cooperative grounds (Grice 1975). This information is conveyed in 'chunks' (Halliday 1967). At least parts of these chunks represent information already familiar to the listener but certain other parts may well be unknown to him/her. de Beaugrande and Dressler (1981:138) call this process **"INFORMATIVITY"**, a term which they use to designate **"the extent to which a presentation is new or unexpected for the receivers"** (ibid). Clark and Clark (1977:32) state that **"speakers tailor their sentences to fit what they think their listeners know"**. Let us consider the following example from the data (See special Transcription Conventions in Section 6.5.1 below).

6.1 (CON IV)

RF. 84. 1 a:::nd **{it}** was e::m **/Perry/** who took **[our statements]**

It is clear that the speaker's purpose in saying this sentence is to inform the listener that 'the policeman' (inferred from previous discourse) who did the act of taking '***our statements***' was ***Perry*** and supposedly '***the taking of our statements***' has already been mentioned in the discourse (i.e. known to the listener). So it seems appropriate to talk about this unit as

having :

1) Information already known to the listener (i.e. GIVEN), **(we gave statements at a police station)** and
2) Information not known to the listener at the moment the utterance is said that **(it was Perry who took our statements).**

These two parts are signalled in the structure of discourse in the right circumstances and are expressed by the speaker in **a specific order using certain lexical, syntactic cues** in addition to pitch prominence (in Halliday's views). For instance, in the example above, the speaker has used the NPs ***it, Perry, our statements*** in a special syntactic order to express the content of the message he is conveying to his listener(s). **Each of these entities has a particular status within the linguistic structure and particular relationships with other NPs within units of discourse and it is those issues in relation to the status of NPs as Given or non-Given that the present chapter is mostly involved with.** This is arrived at by studying in detail all possibilities of linguistic representation of these entities when they occur in samples of speech as will be seen below.

6.1.3 The problem

Two important questions now arise. The first question is related to the generally recognised problem of how any analyst can determine what part of these 'chunks' of discourse are to be taken as 'known' or 'unknown' to the listeners since these are decisions to be made by the interlocutors themselves. This question will be tackled below after discussing the approaches that have dealt with the status of information in discourse. An important point will have to be introduced here, however, and that is : any decision made by an analyst in this concern will not be as accurate as those made by the participants themselves within the speech situations they are involved in. Besides his/her skill and experience, (which can be a

major source of knowledge) the analyst will have to depend on certain clues (linguistic and otherwise) to determine the status of information and on (perhaps) his knowledge of the participants, the purpose of the encounter, the topics discussed and the speech situation itself.

The second question (or rather a set of questions) which is of direct relevance to the purposes of the present investigation, is concerned with the problem of how Given and New information are expressed linguistically in terms of lexis and syntax. In other words, what are the linguistic clues that express these types of information in speech ? What are the relationships that hold between them ? What types of complexities do they show within their status as Given or otherwise ? These questions and the answers available to them at present in the literature (if any) represent the point of departure for the analysis of IS in INT and CON conducted in this chapter.

It is not claimed here that an answer will be found to these questions within this short investigation but in attempting to find a plausible answer for them we will be able to arrive at the 'gist' of the problem which is the study of the linguistic representation of information structure using the data of INT and CON recorded for the investigation. Although it is realised here that this problem can be an important issue of research within such fields as psycholinguistics, cognitive psychology and pragmatics, it is mainly a linguistic problem (cf. Prince 1981:233) and can be tackled using linguistic techniques and procedures, **"if a goal of linguistics is to produce a theory of discourse that distinguishes between a random sequence of sentences and something we would intuitively call a text"** (ibid).

It is also realised here that the categories to be investigated under the above points are numerous. There might be overlap between linguistic and non-linguistic (e.g. cognitive) issues. This is inevitable with a

sensitive issue such as that of IS. So in order to avoid such overlap the following plan has been set up for the sequencing of arguments and analytical procedures adopted in this chapter :

1. Elaborating on the linguistic realisations representing information types and the relationships that can be thought to hold between them as these represent the main criteria for a linguistic theory of IS. This will be done in Section 6.2 below.
2. Reviewing the most relevant available approaches to the controversial issue of IS for the purpose of finding, as already stated in the introductory remarks above, a suitable approach that can best achieve the main purpose of the study. This will be done in Section 6.3 below.
3. Reviewing Prince's taxonomy to IS as it is believed to be the most appropriate of the ones discussed for the purposes set here. It covers all types of IS conveyed by communicators and follows what is thought to be a plausible approach which is amenable to both qualitative and quantitative research types. This is found in Section 6.4 below.
4. Applying the taxonomy to the data of INT and CON using a statistical technique (Section 6.5 below).
5. Discussing the results, implications and conclusions for a linguistic theory of IS (Sections 6.6 and 6.7 below).

It seems right to re-emphasise here that the chapter does not attempt to find solutions to all the problems discussed above though insights from the arguments pertaining to them will be made use of whenever deemed necessary. Practically, it does not address any issues related to how information is produced and how it is comprehended and processed except in as much as they relate to the linguistic analysis of information status. It is merely an attempt to fill in the gap which is thought to be found within linguistic research of the issue of the linguistic realisations of information types exchanged by communicators. To the best of my knowledge, with the exception of Prince's (1981) work no attempts have

been made to investigate such questions on types of SES such as INT and CON.

6.1.4 Levels and technique of analysis

The status of information within any particular type of discourse can be accounted for on various levels depending on the approach undertaken and the purpose of the analysis per se. One can describe the status of information on a global level looking at the text as a whole and moving to the lower level units of information akin to the approaches adopted by some psycholinguists working within discourse processing (e.g. van Dijk and Kintsch 1983; Kintsch 1977a; Meyer 1975, 1977; Frederkisen 1977; Schank 1975a, 1975b; Rumelhart 1975, 1977; Rumelhart and Ortony 1977; Rumelhrat and Norman 1975; de Beaugrande 1978, 1980, 1981) (see also de Beaugrande 1981 for a review and discussion of most of these models). Alternatively, one can look at the lower levels of discourse structure (i.e. on a local scale (cf. ibid:265), akin to the textual models that deal with the status of information on the lower units of discourse, such as 'information units' (Halliday 1967, 1985a), 'idea units' (Chafe 1980), 'the sentence' (Danes 1974; Firbas 1974) and also the psycholinguistic models that deal with information comprehension at the level of 'sentence' (e.g. Clark 1978; Clark and Clark 1977; Clark and Haviland 1977). The model that will be adopted for the analysis of information status in INT and CON in this chapter is an eclectic one in that it draws insights form the local and global models (1) in its application. It is essentially a bottom-up approach in that it starts with the lower units of discourse but it offers the results on a global level in terms of frequencies (percentages and OPTWs).

For illustrative purposes, we can think of the structure of INT and CON to have a hierarchical organisation in a similar fashion to that offered within traditional linguistics (cf. Halliday 1961) and that of discourse organisation in classroom interaction offered by Sinclair and

Coulthard (1975). In terms of grammatical organisation, Halliday's (op.cit.:253) well known rank scale is as follows :

	UNITS
	{sentence
	{clause
RANK	{group(/phrase)
	{word
	{morpheme

On the level of discourse structure, Sinclair and Coulthard (1975) provide the following hierarchy for the description of spoken discourse (classroom interaction), using the 'speech act' as the **"minimal functional unit"** (cf. Hymes 1972:52) as in the following formula :

DISCOURSE
LESSON
EXCHANGE
MOVE
ACT

None of the above organisations seems to fit the language of conversational discourse (see Section 3.1.3.2 for the difficulties of applying the grammatical notions on SES and see also Sinclair and Coulthard (op.cit.); McKnight 1976; McTear 1977; Burton 1980 for the problems of applying the discoursal hierarchy to the language of conversational discourse). However, a similar hierarchy for the analysis of IS in INT and CON can be suggested depending on the unit of analysis adopted for this research (see CHAPTER THREE). The data of INT and CON can be crudely modelled hierarchically to include the following categories from top to bottom :

INT	CON
EXCHANGE	EXCHANGE
TURN	TURN
MAJOR UNIT	MAJOR UNIT
MINOR UNIT	MINOR UNIT

Such an organisation might be helpful in solving some of the dangling problems related to the analysis, especially on the linguistic level, where none of the well-established grammatical units can tidily be applied. Both the lexical and syntactic statuses of INT and CON, for example were investigated within the lower units of the scale (see CHAPTERS FOUR and FIVE above). It is also useful on the level of information structure as has been outlined earlier.

Within the above model, each unit at the bottom of the scale is examined in terms of the discourse entities (i.e. in their linguistic realisations) and then using the same statistical approach adopted in the previous chapters, a calculation of all categories and subcategories on IS is arrived at in the form of percentages and OPTWs. The final results are presented as comparisons between the status of information in the speech of speakers' output in INT as regards that of CON and also on the inter-speaker level of comparison.

The statistical approach to IS outlined above has been adopted although it is realised that there are some reservations about the use of what de Beaugrande and Dressler (1981:140) call **"statistical probability"** in natural communication. But the fact that the analyses have been carried out on a relatively wide scale, involving a considerable number of participants (16 altogether in eight encounters of INT and CON) justifies the use of the statistical approach as de Beaugrande and Dressler (ibid:150) themselves acknowledge **"Hence frequency is useful especially if computed for a very large set of texts"**, though they still realise its insufficiency and unreliability.

In the following few sections an elaboration of the main problem outlined in Section 6.1.3 above will be presented. The major linguistic forms that are thought to express information will be looked at and exemplified. Co-referential relationships between these forms will also be

elaborated on. Then following that, a detailed discussion of what insights the available approaches give can us for studying these linguistic issues will be presented. The examples given from the data in the following sections contain special symbols, each representing a particular category within the taxonomy that will be adopted in the analysis. **All categories to be studied and calculated are bold typed**. A list of transcription conventions is found at the beginning of the thesis and those specific symbols used to represent the taxonomy values are also found in Section 6.5.1 after these values are discussed in detail.

6.2 The Linguistic representation of information structure

6.2.1 Linguistic forms and information structure

As will be discussed in Section 6.3 below, although intonational clues have been used in the realisation of IS and have been, to a certain extent , helpful for pinpointing given and non-given information, they are not the most accurate means of doing that and can even be problematic and unreliable (see also discussion in Section 6.1.2 above). Although not directly studied in this thesis, it is right to say here that there are other more reliable linguistic means such as the use of certain syntactic features and lexical items and pronouns which can help in characterising information to a greater degree of accuracy. Word order and the position of these syntactic and lexical items and also their 'first' or 'second mentions' can be helpful as well. Certain classes of lexical items and relations between them are important for that purpose too, such as **'REITERATION'** relations (cf. Halliday and Hasan 1976:278), which involve the repetition of a lexical item, the use of a general word to refer back to a lexical item, the use of synonyms, near-synonyms and superordinates etc. (ibid). These, in addition to the 'directional' relationships that hold between these items (e.g. phoric relations) (cf. ibid:14-8) can play an important part in the characterisation of information

structure in spoken discourse.

Generally speaking, the elements which represent new information in a discourse unit are lexical units which are mostly represented by NPs with any attributes that might be linked to them. Sometimes the new element in a unit can stretch to include the verb preceding the NP (cf. Chafe 1970). Generally, in the unmarked cases these elements are positioned towards the end of these units. NPs which convey new information are more consistently given strong pronunciation than verb phrases and than other elements of the unit.

One of the most distinguishing features of Given versus New information is the use of the definite and indefinite articles with nouns (Chafe 1970; Sanford and Garrod 1981; Brown and Yule 1983a). It is normally the case that the New element within the information structure is an indefinite noun and any definite or generic noun can be taken to be Given (see **e.g. 6.2 Unit 145** below). This claim is, however, a questionable issue, as will be seen in Section 6.3.2.1 below.

Brown and Yule (ibid:171) summarise some of the linguistic forms by which 'givenness' is expressed in relation to the 'New' elements they refer to. These are mostly studied by scholars under the rubric of 'reference' (cf. Halliday and Hasan 1976, Farag 1986, and also see discussion below). Following is a summary of Brown and Yule's (op.cit.) linguistic criteria for expressing 'Givenness' in discourse with examples from the data of INT and CON :

A.

i) Lexical units which are mentioned for the second time (notice the repetition of **[my wallet]** in **Unit 145.4 of e.g. 6.2** below).

ii) Lexical Units which are presented as being within the same semantic field of a previously mentioned lexical unit (see, for example, the use of the lexical items **(the impression)** in **Unit 136.3** of **e.g. 6.2** below to refer to the whole idea presented in the previous unit that *Sue Wadsworth* is ***'a bit fed up'***. Also notice the use of **(a note)** in **Unit 141.3**

to refer to the previously mentioned items of ***letter*** and ***address***).

B.

i) Pronominals used anaphorically following a lexical form in the preceding sentence (Notice the many occurrences of the pronouns **'she'** and **'her'** for **'Sue Wadsworth'** and the pronoun **'it'** for **'her address' in e.g. 6.2** below.

ii) Pronominals used exophorically (to refer to the physical context of situation) where the referent is present (**Unit 242. of e.g. 6.3 below**).

iii) Pro-verbals (**Units 238, 239 of e.g. 6.3** below).

The following examples from the data contain most of the features discussed above.

6.2 (CON IV)

RF. 132. have **[you]** heard from **/Sue/** by the way..

JA. 133.1 no
2 **[I]** haven't..

RF. 134. **{Sue.. Wadsworth}**..

JA. 135. e:::m but **(somebody)** said **{she}** was a bit fed up..

RF. 136.1 yes..
2 **{that}**'s right..
3 **{that}** was **(the impression) [I]** got..
4 **[I]**'ve been sort of composing **letter** to send to **{her}** [..

JA. 137. [have
[you] got **(an.. an address)** for **{her}** [..

PM. 138.1 [yes..
2 would **[you]** like **{it}**..
3 **[I]**'ll bring **{it}** in.. **[some time]**..

JA. 139.1 mmhmm..
2 in **/the Isle of Wight/**..

PM. 140. yes..

JA. 141.1 mmhmm..
2 yes
3 **[I]** thought **[I]**'d drop **{her} (a note)**..

RF. 142. **[I]** may have **{it}** with **[me]** actually..
143. **[I]** will look in **[my jacket]** when **[I]** go out [..
JA. 144. [mmhmm
145.1 **[I]** think **[I]** was given **{it}** on **(a piece of paper)**..
2 and **[I]**'d stuck **{it}** in in **[my wallet]** [.. *(he he)*
JA. 146. [mmhmm

3 in **{which case}** **[I]** have got **{it}**..
4 but **[I]** might've taken **{it}** out of **{my wallet}**..
5 and put **{it}** in **[my address book]** of course--..

6.3 (CON VII)

AB. 237.1 in **(the front of the Guild)**..
2 on **(the big pillar)**..

KS. 238. **{it}** does..

AB. 239 **{it}** does..

KS. 240.1 remind **[me]** to look **[next time]** ** *(he he)*
2 **[I]**'m in **{the Guild}**..

AB. 241.1 well..
2 there's **(a {Guild} opening and closing time)** up **[there]**..

KS. 242. is **[that]** **([the door])**..

AB. 243.1 well..
2 **{it}** looks like **{the one}**..

A close examination of the data of INTs and CONs on the lines discussed above shows the following general points :

1. Lexical units which are mentioned for the second time and those that are presented as being within the same semantic field (Category A above) are of rare occurrence in the data, much less than those in Category B.
2. Sometimes a series of pronominals referring to a particular antecedent occur in succession (see **e.g. 6.2** above).
3. More often than not, repeated lexical units or coreferential pronouns

may appear in the speech of the second party in the conversation and not in that of the original speaker in whose turn the antecedent has been used (see examples above). Indeed this is a recurrent feature in both types of data (i.e. INT and CON).

4. The span of referential use may vary between a number of words to a whole unit to even a long stretch of units.

These general points, in addition to many other points which will be yielded by the analysis will be discussed in the following theoretical and analytical sections.

6.2.2. Reference and referential entities in spoken discourse

Very much linked to the Given/New distinction in discourse and to the interpretation of lexical units not by virtue of their own use but in relation to other items which they refer to are those items that are often described as **'referential'** (Halliday and Hasan 1976), or sometimes **'co-referential'** (de Beaugrande 1980; Brown and Yule 1983a). The latter term will be preferred here since it offers less ambiguity than the former (cf. de Beaugrande, ibid).

For Halliday and Hasan (op.cit.), this phenomenon represents one of the important cohesive ties in spoken and written texts (p.4). They define **'Reference'** as the process whereby certain lexical items **"instead of being interpreted semantically in their own right, they make reference to something else for their interpretation".** They believe that in English these items are **'PERSONALS', 'DEMONSTRATIVES'** and **'COMPARATIVES'**. Under 'PERSONAL REFERENCE', they study some very important types of textual relations normally referred to as 'anaphoric' (i.e. referring back to a previously-mentioned antecedent), 'cataphoric' (i.e. referring forward to a following element) and 'exophoric' (i.e. referring normally to the outside environment of the text). The latter type as in :

6.4 (From Halliday and Hasan, op.cit.:18)

Did the gardener water those plants ?

is taken by Halliday and Hasan as not cohesive and therefore excludes its treatment from their study of referential entities. Other types of personal pronouns are 'cohesive' within the structure of discourse. The personal pronouns **'she'** and **'her'** in **Units 135 and 136 of e.g. 6.2** above refer back to **'Sue Wadsworth'** in **Unit 134** of the same example.

Under 'DEMONSTRATIVE REFERENCE' Halliday and Hasan study the demonstratives, deixis and deictic elements (those of 'person', 'space' and 'time' and also certain uses of 'this', 'that', 'these', and 'those' (see also Levinson (1983) for a comprehensive discussion of these elements in discourse) . The demonstrative **'that'** in **Unit 242 of e.g. 6.3** above refers cataphorically to the noun **'door'** in the same unit.

Halliday and Hasan distinguish a third category of reference which they call 'COMPARATIVE REFERENCE'. This is particularly relevant to the present discussion as it will be one of the major subcategories of Inferables distinguished within the data of INT and CON (see Section 6.4.3.1 below). They divide **'comparative reference'** into two major types : **'general comparison'** and **'particular comparison'**. By **'general comparison'** they mean comparison that is in terms of "*likeness or unlikeness*" (ibid:77), i.e. *being similar to* or *different from* another entity that has explicitly or implicitly been mentioned in discourse. '*Particular comparison*' is the one which is related to '*quantity or quality*' (ibid), expressed by adjectives, or adverbs (examples will be given in Section 6.4.3.1 below)(2).

While for some linguists, like Halliday and Hasan (op.cit.), the term 'reference' is used to refer to a relation that holds between one element of

the text and another element of the same text, for some other discourse analysts it refers to an element in the text and an entity in the 'discourse model' (cf. Farag 1986). A 'discourse model' is a discourse representation which a hearer tries to construct for himself/herself when receiving discourse (cf. Brown and Yule 1983a:206) (see also Webber 1978a for a more detailed elaboration on the notion 'discourse model'). This view of 'reference' is shared by many discourse analysts (see for example, Stenning (1978); Hawkins (1978); Webber (op.cit. and 1978b, 1981); Prince (1981). Farag (op.cit.:120) points out that this view to 'reference' is somewhat similar to that held by Lyons (1977a:660), who suggests that **"an anaphoric pronoun refers to what its antecedent refers to"**. Prince's taxonomy of referential relations are also partly based on this view.

The modification of Prince's approach to include some textual subcategories which are not explicitly mentioned by her (see discussion below) is an attempt to bring the two points of view about reference discussed above together since in both cases one is dealing with the linguistic features of discourse and since, as Karttunen (1971:366) points out :

> "the problem of coreference with a discourse is a linguistic problem and can be studied independently of any general theory of extra-linguistic reference".

As will be seen below, Prince's taxonomy does not include many of the referential relations which have been discussed by other scholars in their treatment of reference (e.g. comparative reference as discussed by Halliday and Hasan's model of cohesion). Details of the modification made will be cited below. It is believed that these modifications make the approach more comprehensive and conclusive of all features of co-referentiality within SES, as will be seen in the discussion of the approach in Section 6.4 below.

The use of reference and co-referential entities in SES is of vital importance looking at the matter from the point of view of discourse structure, complexity and interpretability on the part of the listener. Speakers exchange roles and talk about topics referring both to the outside and the immediate environment using such co-referential items as pronouns, demonstratives and so on. The first and second persons (*I, we, you*) are frequently used by speakers and listeners referring to themselves and to each other. The third person pronouns (i.e. *he, she, it, they*) are used to refer to persons or things other than the speaker or hearer (Lyons 1968:276). Other forms of co-referentials are used by the speaker to refer to things and topics inside and outside the discourse (anaphorically, cataphorically or exophorically, using Halliday and Hasan's terms mentioned above). This process seems to have a lot to do with the notion of 'complexity' as outlined in the chapters above. Indeed the complexity of one type may sometimes be greatly dependent on that of other types and sometimes the linguistic variations within one type may be attributable to variations within the other types. Farag (1986:112), for example, found that a number of grammatical differences between speech and writing are largely attributable to differences in referential strategies. This will be checked against the data of INT and CON in the analysis sections.

Due to the great number of linguistic features and co-referentials that need to be discussed as representatives of information within the structure of INT and CON, the approach adopted is to discuss and exemplify each of them in more detail in the analytical sections each within its own category. This has been done for the purpose of putting all characterisation, exemplification and analysis together to facilitate direct reference and maintain coherence of relevant issues in relation to each other. In the next few sections the main arguments within the approaches of IS will be discussed. The aim is to see how much they can offer the approach to be adopted in this investigation.

6.3 Approaches and arguments of relevance

Bringing into mind the classification of the approaches for the study of information status into three broad categories : **"Textual", "Cognitive" and "Discoursal"** mentioned already in Section 1.2.3.3.3, one can readily spot the differences and similarities between them and can evaluate them in relation to the analysis of IS in this chapter and to how much they offer a linguistic theory of information status within spoken discourse. I will first examine the major arguments within these approaches in relation to those of the present study before presenting the analysis.

6.3.1 Arguments within the 'Textual Approaches' to IS

The study of informational status within the textual approaches has always been linked to the distribution of Given and New information within the chunks of information which speakers and listeners produce and share within any speech encounter (cf. Halliday 1967, 1985a; de Beaugrande and Dressler 1981). The distinction Given/New was originally introduced by Mathesius (1939), along with other notions such as 'Theme and Rheme' and 'Communicative Dynamism', which constitute the major principles of the Czechoslovakian approach to information structure (i.e. FSP) (see especially the collection of articles in Danes 1974). Mathesius used **'Given'** to refer to what **"is known or at least obvious in the given situation"** (ibid, in Danes op.cit.:106) and he refers the **'New'** to that part of information which is unknown to the listener (see **e.g. 6.1** above).

Unlike the Prague linguists, Halliday (op.cit.) distinguishes between 'Information Structure' as the study of Given/New information and 'Thematic Structure' as the study of Theme and Rheme in the structure of discourse for the relationship between the two distinctions (see Halliday

1985a). This is not going to be elaborated on here as it is not practically relevant to the present discussion. What is really relevant here is Halliday's treatment of 'Information Structure' per se therefore it will be discussed below, albeit briefly.

Halliday (1967) talks about 'Given' information within the structure of 'information units' in terms of 'recoverability'. For him the significant variable is : **information that is presented as recoverable (Given) or not recoverable (New) to the listener.** This raises some objections as to **"what kind of procedure the speaker gets the listener to interpret what, in fact, is recoverable from it, as being "new"** (cf. Danes op.cit.). This is put more aptly by Yule (1981:42) who wonders how an analyst can, in order to discover the formal linguistic realisation of 'Given', set about deciding that a speaker can assume a particular referent, for example, is known to his hearers. In his opinion this involves not only the knowledge but also the intentions, purposes or even beliefs of speakers. This is also shared by Prince (1981:228) who suggests that **"a consideration of speakers' hypotheses about hearers' beliefs and strategies must be a primitive"**.

Secondly, Halliday puts great emphasis on the direct relationship which he firmly believes to be present between information structure and the phonological realisation of the information units of discourse (see CHAPTER THREE, Sections 3.1.3.1 and 3.1.5 for a detailed discussion). He states :

> "Information structure is realized phonologically by 'tonality", the distribution of the text into tone groups: one information unit is realized as one tone group."

This has also been argued against (cf. Brown and Yule 1983a:165) as there are instances of New information conveyed within speakers' utterances that do not necessarily have pitch prominence (cf. Chafe 1970 and discussion below). Besides, there are many other prosodic features

which are thought to determine tone group boundaries which Halliday's approach seems to neglect. Such features include : tempo, relative duration, relative pitch height and perceived pause. On the prosodic level of analysis, these represent important cues for matching information structure in spoken discourse.

Thirdly, Halliday's work seems to be all exclusive to spoken discourse (Brown 1983:65). Apparently, any application of the approach to written language examples will have to go through a process of attributing the status of "Given" to all that a reader can ever be expected to know from whatever source (ibid:68), as opposed to the status of "New" to be attributed to those linguistic forms, which inevitably attract phonological prominence if read aloud.

6.3.2 Arguments within the psycholinguistic approaches

6.3.2.1 Chafe's account of Information Structure

Although Chafe's views (see especially his work of 1970, 1974, 1976, 1977), can be more appropriately put under the psycholinguistic approaches, they will be discussed here for their relevance to those of Halliday mentioned above. First of all, Chafe relates 'Givenness' and 'Newness' of information to the addressees' consciousness at the time the utterance is produced. What this means is that chunks of information that are present in the addressees' consciousness at the time the utterance is spoken represent 'Old' (i.e. known, or given) information and those that are not present at that time, albeit perhaps already known to them, are 'New'. This is not in agreement with Halliday's characterisation of Given/New information presented above. As a matter of fact, from a linguistic point of view, this adds to the confusion already present concerning the issue. Most of what we normally say to our addressees during everyday conversation is not present in their consciousness at the

time of communication. According to Chafe's view all this is to be taken as new information to the addressees, a claim which does not seem to be in the least acceptable by many linguists in general terms.

Chafe, also believes that usually, old information is conveyed in a weaker and more attenuated manner than new information. He states that :

> "In English, it seems to be the case that given information is always pronounced with low pitch and weak stress, unless it is contrastive (1976:31) (Underlining in original).

But is this really similar to Halliday's (op.cit.) suggestion that New information is expressed by the most prominent syllable in the clause ? The answer to this question appears to be no. Let us consider an example taken from Chafe (1970:213). In answering the question **'What did David do then** ? (Underlining here signals the most strongly stressed syllables having a higher pitch than the rest of the elements in the sentence), an answerer will in the most unmarked case respond **'David emptied the box**'. Chafe argues that it is not only ***'the box'*** that signals the new information, as Halliday's treatment would suggest, but also the verb ***'emptied'*** as well although the latter is pronounced with low pitch. Similarly, Chafe implies that adjective phrases and adverb phrases can express new information to the listener (see Chafe ibid:213-215 for other examples). One important implication for this controversy is that what Halliday takes to be Given might not be the same as 'old' in Chafe's analogy and vice versa, which in turn shows the amount of controversy displayed in the various investigations of IS in discourse. Therefore, care should be taken when dealing with such terminology.

Another major difference between Halliday's and Chafe's characterisation of IS cited above is that while the former investigates this issue within his 'Information Units', the latter does not use his

well-recognised 'Idea Units' (see Section 3.1.3.3 above) which are supposedly equivalent to Halliday's 'Information Units' (cf. Beaman 1984). Actually Chafe's use of 'Idea Unit' started in his 1980's discussion of the *Pear Stories* in which he used the notion as a **'focus of consciousness'** which the human memory is able to recognise as one spurt of information at one time (cf. Chafe 1980:25)(3).

6.3.2.2 Other psycholinguistic accounts of relevance

In addition to Chafe's (op.cit.) contribution to the analysis of information in discourse, other psycholinguists such as H.H. Clark and his associates (see particularly, Clark 1975; Clark and Haviland 1974, 1977; Clark and Clark 1977; Haviland and Clark 1974) talk about Given and New information but they do this in relation to discourse comprehension (see Section 1.2.3.3.3 above). Clark (1975:245) suggests that speakers and listeners employ the production and comprehension of New and Given information as a strategy. It is a form of contract between them (ibid). For them Given information is expressed by what the speaker **"believes the listener already knows and accepts as true"** as regards information the speaker believes **"the listener does not yet know"** i.e. New (Clark and Haviland 1977:4). This view is similar to the original definition of the notions presented by Mathesius (1939). Comprehension occurs when the listener tries to find antecedents for the Given information he/she is expected to know and tries to integrate them to the bits of information which are 'news' to him/her. If the listener is not able (for any reason) to find such an antecedent in his/her memory, he/she will try to use a process which Clark (op.cit.) postulates to be present within each individual's capacity called **'BRIDGING'**, which is the result of the listener adding a set of one or more propositions to memory to help interpret the Given information. Unless the listener is successful in interpreting this information, he/she will not be able to comprehend the New information and will consequently use other strategies (ibid:247).

6.3.3 Implications for an extended theory

Generally, the approaches discussed above represent two different trends, each characterising the status of information from its own perspective. They show clearly the controversies surrounding the analysis of IS in terms of Given/New only. There are many problems involved in this distinction both on the level of information types and on the level of linguistic realisations of these types. There is certainly a need for establishing a theory that covers **(1)** all types of information conveyed not just in terms of Given and New but in terms of what speakers can infer from whatever source; and **(2)** there is a need to establish a taxonomy of the types of linguistic forms that can represent the types of information conveyed. Such a theory will be helpful in studying information structure on the level of discourse analysis and it will also help in the investigation of the linguistic structure of information as represented by the categories mentioned in Section 6.2 above. The first of these points is provided by Prince's (1981) taxonomy to IS (see below for discussion). Prince has also made some attempts at establishing the second of these points but as we shall see later she is not very clear about it. This chapter is an attempt at establishing the second of these points through an analysis of the first one. This of course needs an understanding of Prince's approach in detail so as to be able to apply it on the data of INT and CON analysed for this study.

Prince has made use of the controversies found within the other approaches. She therefore adopts an intermediate position between the two, claiming that Given and New information are the outcome of the **'SHARED KNOWLEDGE'** of speakers and listeners, though she discards the term 'shared knowledge' altogether for the confusion it has given rise to (p.232) and adopts the term **"ASSUMED FAMILIARITY"** (AF) instead. This term has also been used as a label for the taxonomy itself (cf. Brown and Yule 1983). This taxonomy will be critically discussed and adopted for the analysis of information in INT and CON.

6.4 The 'ASSUMED FAMILIARITY' taxonomy

6.4.1 Why the AF Taxonomy

The choice of the AF taxonomy as a model for the analysis of IS in INT and CON in this thesis has been made for many reasons. First of all, it is **"an extended taxonomy"** (Brown and Yule 1983a:182) in that it has made use of all arguments pertaining to the controversial issues of Given/New information not just from a linguistic perspective but also from a discoursal and cognitive one. It represents **"a richer taxonomy than the simple 'given/new' distinction'** (ibid). This is right in that it is able to describe the status of information based on the shared assumptions between interlocutors in terms of 1) Givenness and Newness of information, 2) Inferentiality, 3) Coreferential relations. These features can be investigated by examining the linguistic forms that represent them within the units of discourse. As such the approach seems to provide linguists with a reasonable framework within which spoken discourse types can be examined, analysed and compared with each other. This is not available in the other discoursal approaches and if it is available (cf. Sinclair and Coulthard's (1975) hierarchical approach to classroom interaction), it seems to present difficulties to analysts in its application to such SES types as conversational discourse, as discussed above.

The taxonomy, being applied by Prince to an extract of conversational narrative discourse, seems to fit other types of SES quite well. Its ability to handle conversational texts both on the local and the global levels is perhaps one of the reasons why the taxonomy is suitable for such types of spoken discourse as speech exchange systems. No other taxonomy has been able to handle such systems efficiently and successfully.

One major feature of the AF taxonomy that makes it suitable for the present investigation is its amenity for statistical analysis. The various

types of categories of the IS, as represented by their linguistic realisations within the units of data can be calculated and compared in INT and CON for the eight groups of speakers recorded for the analysis. The complexity of linguistic forms representing these categories will be discussed and assessed as differences between the two styles of speech under analysis. So to sum up these points, Prince's taxonomy will be used just as a means of establishing through statistical analysis the linguistic means that are used in the representation of various types of information conveyed and exchanged by speakers in conversational settings. The aim is not to establish another Given/New taxonomy akin to that used by Halliday and others because that has already been done by Prince herself. The aim is to apply her taxonomy to our data for the purposes set at the outset of this chapter.

6.4.2 General characteristics

Taking into consideration the fact that there is a need within the linguistic sciences to produce a theory of discourse that relates formal linguistic features to the types of information delivered and received by speakers and listeners (Prince 1981:233) and the confusion in the literature concerning these types of information in both use and terminology, Prince (ibid) presented her taxonomy to the characterisation of information structure, basing it on the assumption that **"a text is a set of instructions from a speaker to a hearer on how to construct a particular DISCOURSE MODEL"** (p.235). The model will contain **DISCOURSE ENTITIES, ATTRIBUTES** and **LINKS**. It is important to elaborate on **'discourse entity'** as it is the basic notion in the taxonomy. A 'discourse entity' represents discourse-model object, which may represent **"an individual, a class of individuals, an exemplar, a substance, a concept, etc."** (ibid, see also definitional comments in Section 6.1.1). All discourse entities in a particular discourse-model are represented by NPs in a text (though not all NPs in a text represent discourse entities) (ibid).

In order to understand the various components of the taxonomy, I first present a diagram adopted from Prince (ibid) which gives a summary of the basic categories of the taxonomy and which Prince thinks the taxonomy comprises :

ASSUMED FAMILIARITY		
NEW	INFERABLE	EVOKED
Brand-new	(Noncontaining)	(Textually) Evoked
(Unanchored)	Containing	Situationally Evoked
Anchored		
Unused		

Diagram 6.2 Prince's 'ASSUMED FAMILIARITY' taxonomy

As the diagram above shows, Prince distinguishes three main categories of discourse entities : **NEW, INFERABLE** and **EVOKED**. Entities which are introduced for the first time into a discourse are **NEW**. A **NEW** entity can either be **BRAND NEW** in which case it has no referent in the hearer's mind, thus urging him to create a new entity for its interpretation, or it can be **UNUSED,** in which case the hearer is assumed to have some knowledge of it but it is not present in his 'consciousness' at the moment the utterance is being produced (cf. Chafe 1976:30). BRAND NEW Entities can be of two types **ANCHORED** or **UNANCHORED**. The NP representing the **ANCHORED** type is linked to some other discourse entity in the discourse whereas for the one representing the **UNANCHORED** it is not.

EVOKED entities are those that are already in the discourse model. These can be either **TEXTUALLY EVOKED** or **SITUATIONALLY EVOKED** depending on whether they are evoked by the hearer on textual grounds following instructions from the speaker, or the hearer was able to evoke them all by himself/herself for situational reasons (ibid:236).

SITUATIONALLY EVOKED entities represent discourse participants and salient features of the extratextual context, which includes the text itself.

The **INFERABLES** are those entities that can be inferred by the speaker via logical-or more commonly, plausible-reasoning, from discourse entities already evoked or other inferables. A special sub-class of inferables is what Prince terms **"CONTAINING INFERABLES", where what is inferenced off is properly contained within the Inferable NP itself"** (ibid:236).

Yule (1981) and more recently Brown (1983) and Brown and Yule (1983a) introduce a further distinction into Prince's category of **"Textually Evoked Entities"** (see discussion above). They believe that those entities that are textually evoked can either be **"current"** or they can be **"displaced"** according to their position of occurrence. The **CURRENT EVOKED ENTITY** is **"the most recent 'new' entity to be established in the discourse"** (Yule op.cit.:47). **DISPLACED EVOKED ENTITIES** are those that **"have been established previously in the discourse"** (ibid). The following examples taken from Brown and Yule (op.cit.:1983:173) make this distinction clear :

6.5 (Current and Displaced Evoked Entities)

a. at the end of *the lines* 1 draw *a square* 4
b. underneath *the red line* 4 write *IN* 7
c. draw *a diameter* 2 across *it* 1
d. draw *a straight line* 5 across *the circle* 1

(Italics in original)

In this example, the numerical subscripts are used to refer to the number of occurrences the entity has been used in the text. As each new entity expression is introduced, it can be given a numerical subscript. Thus the

NP with the highest number attached to it will mention the most 'new' entity and the NPs with lower numbers are those that refer to those previously mentioned. Thus in the example above, the expression **'it'** is the current evoked entity as it is used to refer to an entity which was introduced immediately before **'a diameter'** was mentioned; so it has the number (1). The expressions : **the line, the red line, the circle**, represent the displaced entities as they were used to refer to new entities which occurred prior to the most recent new entity (ie IN^7).

The following diagram represents Prince's taxonomy of information status with Brown and Yule's suggested modification of the Textually Evoked Entities :

<u>ASSUMED FAMILIARITY</u>

<u>NEW</u>	<u>INFERABLE</u>	<u>EVOKED</u>
Brand-new	(Noncontaining)	Situationally Evoked
(Unanchored)	Containing	(Textually) Evoked
Anchored		current
Unused		displaced

Diagram 6.3 Brown and Yule's modification of Prince's 'ASSUMED FAMILIARITY' taxonomy

This extension to Prince's taxonomy is quite useful especially in the specification of information structure in terms of New and Given with relation to the form and function of discourse. However, as Brown and Yule (ibid:188) themselves admit, this extension is not of great help in establishing a taxonomy of information status **"independently of the forms of expressions used by speakers"** and that it seems that the only safe access to information status is provided by the form of the expressions used by the speaker/writer (ibid). This is why this extension has not been employed within the application of the taxonomy on the data of INT and CON.

6.4.3 A critical assessment of the taxonomy

While it is acknowledged (cf. Brown and Yule 1983a; Farag 1986) that Prince's AF taxonomy is one of the richest so far present on information status, there remains a number of points that have to be raised, some of which may be taken as arguments against the plausibility of the taxonomy and its effectiveness when applied to different discourse types. Most of these points arise from the actual application of the taxonomy to SES. The first and most serious difficulty seems to lie in the characterisation of elements as 'Given' and 'New' per se. As Yule (1981:42) rightly argues, a 'Given' or a 'New' entity within any speaker's utterance is determined by the speaker as 'New' according to his/her expectations about the hearer's state of knowledge at the time the utterance is said. The analyst's decision about what is 'Given' and what is 'New' may or may not tally with the speaker's processing. There is no simple and direct way of validating the analyst's decisions, although clearly there are cues (e.g. prosodic : pitch, pause, syntactic : definiteness and indefiniteness) which are the principle bases of the analyst's decision. It is not simply that errors of analytical judgement occur. It is that analytical activity and processing activity cannot be assumed to match in any direct one-to-one fashion.

Further, as Brown (1983:70) has noted as well, Prince totally ignores the phenomenon of intonation although she analyses an extended spoken text and although a major part of her argument of the Given/New dichotomy (p.226) is based on the distinction proposed by Halliday (1967). The characterisation of discourse entities as 'New' or 'Evoked' in Prince's taxonomy is wholly based on linguistic (lexical and syntactic) and discoursal criteria. This is not in itself a serious drawback in the taxonomy but its avoidance should be well accounted for, which Prince did not actually do within her discussion of the taxonomy.

There seems to be a good degree of overlap between the categories proposed by Prince, which might well be due to the great number of categories involved and the possibility of having more than one interpretation for the entities under examination. It is not quite clear, even to Prince herself (p. 252) whether the category termed **'INFERABLE'** should be counted as **'NEW'** or **'EVOKED'**. The answer to this latter question might solve the other problem put forward by Prince herself of whether the taxonomy is **"binary"** or **"ternary"** (ibid). This question will be tackled in more detail in the discussion sections that follow.

Another point about the taxonomy is related to the syntactic realisations of the 'discourse entities' themselves. Prince takes it that all discourse entities are represented by NPs in a text (p.235). However, this does not seem to be always applicable. While it is true that the majority of 'discourse entities' are represented formally by NPs and these can either be 'New' or 'Given', **'attributes'** and **'links'** can also be 'New' to the hearer. This means that other types of phrases (e.g. verb phrases, adjectival phrases, adverbial phrases) (see also Chafe 1970:112) can have the property of 'Newness' or 'non-Newness'. For instance, in the following Unit :

(CON VIII)

JA. 25. **[I]** used to be hopeless on **{coursework}**..

the discourse entity **{coursework}** is textually evoked as it has been mentioned many times in the previous discourse. Accompanied by heavy prosodic prominence the adjective **'hopeless'** seems to be expressing the New element. Also many types of co-referentials can be represented formally not by NPs but by adjectival phrases or even adverbial phrases (cf., for example Halliday and Hasan's 1976 comparative reference) (see the use of the comparatives **'easier', nicer', 'more comfortable'** in **e.g. 6.7** below).

6.7 (INT I)

RF. 4. 1 well
2 [I] think as far as **(accom.. accommodation)** goes..
3 **{it}**'s **(a lot easier)**..
4 or **(much nicer) {living off campus}**
5 because **[you]**'ve got.. **(a bigger room)**..
6 usually **(a more comfortable place)** to live in..

In addition to the points mentioned above, the AF approach ignores altogether the phenomenon of 'Ellipsis' which, for some types of spoken discourse (e.g. conversational discourse) is one of the frequently recurrent features that help in achieving cohesion as 'Ellipsis' is **"normally an anaphoric relation"** (Halliday and Hasan 1976:144). So it seems that any taxonomy that deals with coreferential relations must deal with 'ellipticals', especially if it were to be applied to SES types.

Two points in favour of the taxonomy, however, should be mentioned at this point : First, Prince has left the taxonomy open to improvement, to **"refine/revise/replace"** (P.252) as deemed necessary; secondly, the taxonomy can very appropriately be used to differentiate **"styles of discourse, by characterizing the nature and complexity of the entities and inferences involved and the morphology and syntax of the NPs representing those entities"** (ibid). This is why the approach seems to be most suitable for the purposes set out for the comparison made between interview style and conversation style undertaken in this study.

6.4.4 Prince's application of the taxonomy

Prince applied her taxonomy to two **"naturally occurring texts"**, representing two different styles : an informal oral narrative and a formal written text (the beginning chapter in Hymes 1974). The purpose was to see

how the taxonomy works and what patterns emerge. What is interesting to us here is the results of her analysis on the linguistic level. The analysis of the data showed a preferred hierarchy or scale for what type of entity used, more or less as in the following formula(4) :

Familiarity Scale

{E }

$$> U > I > I^{c} > BN^{A} > BN$$

{E^{s}}

where : E = Textually Evoked; Es = Situationally Evoked; U = Unused; I = Inferables; Ic = Containing Inferables; BNA = Brand New Anchored; BN = Brand New (Unanchored).

As far as the differences between her informal and formal data are concerned, Prince found that the informal type showed a greater tendency than the formal type to use Evoked entities but less so of the Inferable type. The formal text is characterised by a higher degree of METALINGUISTC inferencing, and abstractness than the informal narrative, which, in Prince's belief, are features of the linguistic complexity present in the former but not in the latter. The two types of discourse differ also in the way they call for what Prince terms "cultural Assumptions" required for the inferencing. There is another difference which concerns the size of the entities themselves. Many of the entities in the written text, but not in the oral text are themselves made up of entities and attributes, with the composing entities themselves being complex (ibid:251).

These differences seem to suggest that the taxonomy can be very illuminating and insightful when applied to different discourse types, both spoken and/or written. The application of the taxonomy to the two types of SES under investigation may provide the type of extension which Prince calls for.

6.4.5 Some necessary modifications to the AF taxonomy

Some modifications of Prince's taxonomy were deemed to be necessary for the present analysis. The modifications involve both the 'macro' structure of the 'Assumed Familiarity' categories themselves and its 'micro' structure (i.e. the type and structure of the subcategories that constitute the five major categories mentioned above), and these are now discussed.

On the macro level, as will be seen from the results and the discussion below, the data suggest that the Inferable category of entities can be of two types NEW and EVOKED, depending on the way the inference is being made and on how to interpret the terms 'Given' and 'New' (see discussion in Section 6.2.1 above). According to this modification, the classification of the taxonomy will be 'ternary' rather than binary (cf. ibid:252), consisting of three major categories (NEW, INFERABLE and EVOKED) respectively, as will be shown in diagram 6.3 below.

The modifications on the micro levels concern the actual application of the approach to the discourse entities themselves. There is certainly a need to distinguish within each of the three categories of the taxonomy a number of subcategories, the specification and description of which constitute the internal structure of the taxonomy. This makes explicit what is missing in the original taxonomy. For example, all types of comparative reference have been included within the Inferable category since these represent a recurrent entity whose meaning and interpretation can be inferenced from the entity being compared to (see Section 6.4.3.1 below).

The main subcategories distinguished in addition to those already existing in Prince's taxonomy fall within the Inferable and Evoked categories. The following diagram shows all the categories and subcategories of the AF taxonomy as will be applied in the analysis of INT and CON :

ASSUMED FAMILIARITY

NEW
Brand-new
(Unanchored)
Anchored
Unused

INFERABLE

1. Comparative Inferables
2. Inferables from Previous or Subsequent discourse
3. Inferables from the 'Outside World'
4. Inferables from Logical Interpretation
5. Containing Inferables
6. Generic Inferables
7. Descriptive Inferables
8. Other Inferable

EVOKED

TEXTUALLY EVOKED
1. 3rd Person Pronouns
2. Extrapositional 'it'
3. Repeated Entities
4. Relative Entities
5. Deictic Demonstratives
6. Others

SITUATIONALLY EVOKED
1. 1st Person Pronouns
2. 2nd Person Pronouns
3. Others

Diagram 6.4 Prince's 'ASSUMED FAMILIARITY' taxonomy with the proposed modifications

In the following sections an attempt will be made towards the characterisation of each of the categories and subcategories of the taxonomy with examples from the data provided of each type as necessity arises.

6.5 The present investigation

6.5.1 Methodology

Drawing on the statistical approach employed by Prince (1981) in the application of her AF taxonomy to two types of discourse (informal oral narrative and formal written excerpt from linguistic literature), in each FSSU, all the 'discourse entities' were analysed in terms of their New, Evoked and Inferable status of occurrence. No systematic prosodic analysis was employed. However, in certain (problematic) cases, both pitch and prominence within syllables were consulted.

A notational system was employed in which each category was given a special symbol to facilitate the process of its location and computation in the data analysed. The notational conventions used particularly for this chapter are as follows :

<u>UNDERLINING</u>	**BRAND NEW ENTITIES (BNEs)**
/ /	**UNUSED ENTITIES (UEs)**
( )	**INFERABLE ENTITIES (IEs)**
[ ]	**SITUATIONALLY EVOKED ENTITIES (SEEs)**
{ }	**CONTEXTUALLY EVOKED ENTITIES (TEEs)**
{**}	**ZERO RELATIVE PRONOUN**

Because of the great number of categories and subcategories of the modified taxonomy, two points should be noted. First, a detailed description of each of the categories and exemplification from the data was not possible within the previous explanatory sections. So both the description and exemplification have had to be incorporated within the analysis sections that follow. The approach adopted in the following

sections is to talk about the category (or subcategory) under examination in relation to the data of INT and CON with exemplification where necessary. Then the results of the statistical analysis will be presented in tables and graphs in percentages and OPTWs . The results represent calculations of just the students speech (i.e. with the speech of the interviewer excluded from the calculations). This has been done to get a balanced set of data and speakers in INT and CON. At the end of each main category examined a discussion of the results will follow. A general conclusion of all the results and implications will be provided at the end of the chapter.

The second point that has to be borne in mind in the following sections is that there are bound to be certain cases of overlap between the subcategories themselves (especially so within the Inferables) (see, for example, e.g. 6.8 below). This is not surprising, given the fact that they are so numerous. This means that an entity can be taken to belong to more than one subcategory at the same time, as it may fit the conditions necessary for both subcategories. The figures themselves do not reveal the extent of overlap, and there are cases where the entity under consideration was considered under the two overlapping subcategories. However, these cases are few in number and do not, in any case, influence, the general significance of the results.

As a point of departure for the study of the AF categories, the following table presents the raw figures and percentages of the occurrences of the three general major categories New, Inferable and Evoked. This is presented to show the general frequencies of these categories as they will always be needed as reference in the following sections. No attempt will be made at this stage at interpreting the results of this table. The percentages will be displayed in graph form as Fig. 6.6 and there will be a full discussion of these results in the final 'Summary and Conclusion Section'.

	RAW FIGURES		PERCENTAGES	
	INT	CON	INT	CON
BNEs	145	181	4.0%	3.9%
UEs	45	118	1.2%	2.6%
IEs	1080	1255	29.8%	27.2%
TEEs	1101	1501	30.4%	32.6%
SEEs	1251	1554	34.5%	33.7%
TOTAL	3622	4609	100.00%	100.00%

Table 6.1 Percentage of Occurrence of each of the AF categories in relation to each other in INT and CON

Notice that the Evoked categories have the highest percentage and the New entities the lowest. The Inferables, though high in occurrence, are slightly lower than the Evoked categories but are much higher in occurrence than the New (see Section 6.6 below for a full discussion of these results).

6.5.2 New Entities

6.5.2.1 Brand New Entities

It is important to re-emphasise here that what is taken to be 'Brand New' is dependent on the analyst's observation of the linguistic behaviour and environment within which the discourse is used and my previous knowledge of the topics, participants and purpose of the speech events, although, as shown in Section 6.2.1 above making such decisions may be incorrect. The analyst can never be sure of the speakers' intentions and decisions regarding the information structure.

In Prince's terms BNEs are those entities that are assumed to be unknown to the hearer, i.e. introduced into the discourse for the first time. For example, the entity **'a bus'** in **"I got on a bus yesterday and the driver was drunk"**(5) is brand new from the hearer's viewpoint. Later entities which refer to 'a bus' either through repetition or pronominalisation will be treated as 'non- new' (Evoked) or simply 'Given' in Halliday's terms.

It seems to be normally the case that, in SES, the production of BNEs into discourse is linked in one way or another to the topics being exchanged and discussed. The more new topics are raised, the more likely it is that BNEs will be presented into the speaking floor. It is also likely that BNEs in SES appear more often within the initial units of discourse within which new topics are presented than within later units. This needs a lot of controlled experimentation before one can draw valid conclusions. Expectedly, however, the frequency of occurrence of BNEs within any conversant's speech might be quite different from one speech style to another according to the purpose of the conversation encounter, the audience, the topics discussed, and formality of the situation. For example, within interviews, the interviewer's initiating question mostly includes at least one BNE over which subsequent discussion might circulate. Later BNEs are presented and subsequently cohesively referred to as the conversation continues. In conversational encounters of the informal type between friends, new topics, within which new entities appear, might leap into the discourse. It might also be the case that a conversation drags on one particular topic with no 'NEW' element presented by any of the speakers until a decision is made by one of them to 'activate' the floor by the presentation of a new topic which might contain a BNE. Notice, for example, the distribution of BNEs in the following two examples (recall that BNEs in the data are printed in boldface and underlined) :

6.8 (INT I)

JA. 1. 1 right..
2 --..--..
2. 1 **[you]**'ve.. **[you]**'ve both lived on **/campus/** for **[a year] [now]**..
2 e:::.. w.. w.. what would **[you]** see as **(the main difficulties)**..
3 for **(postgraduates)**..
4 of of living.. on **({a campus} like {this})**..

KW. 3. **[I]** think **{it}**'s **(the fact) {they}** have no **privacy really**..
4. 1 **{they}** have **people** wandering in and out **{their room}**..
2 **(cleaners for instance)**..
5. 1 and when **[you]** are asked to move
2 **[you]**'re given.. absolutely **no notice**..
6. 1 it's as if.. **(the housekeeper)** has no **(idea)**
2 that when **[you]**'re doing **a research project**
3 that **[you]** do actually..
4 'cos.. alter **[your days]**..
5 that **[you]** fit **(everything)** in around **{it}**..

A great deal of the discourse that follows in this particular interview will circulate around the New entities issued in these initiating units. The following example represents a chunk of chat in which the speaker (KW) interrupts her previous discourse to start a new episode which contains BNEs :

6.9 (CON I)

KW. 11. 1 **[I]** got **an interview** at **/Trust House Forte/**..
2 e:::.. **{which}** is absolutely awful..
3 **(people)** were just dreadful [..
HC. 12. [yeah

13. 1 **[you]**'d never believe that **{they}**'d actually done **this course**..
2 because **{they}** just seemed to have no **idea** of how to interview **[you]**..
14. 1 when **[I]** asked what **salary structure {they}** had in **(mind)**..
2 **{they}** said

3 o:::h
4 **[we]** haven't really thought about **{it}** yet..
15. so **[I]**'d..gone for **a couple of drinks** with **a friend**..
16. 1 and **[I]**'m suddenly at **{Euston Station}**
2 trying desperately hard to look.. civilised and.. not.. tipsy.. [*(he he)* --
JA. 17. [*(he he)* yes..

Notice that in these two excerpts the apparent similarity between the frequency and distribution of BNEs might be due to the fact that both examples have been produced by the same speaker with the difference that in the first one she is acting as a respondent to the interviewer's questions and in the second one she is engaged in unplanned casual chat. Whether such similarity would obtain between any two speakers' speech in INT and CON is yet to be seen after the results of the analysis will be presented and discussed below.

As has already been mentioned above BNEs can either be anchored or unanchored depending on whether or not they are linked to another entity properly contained in it. Nearly all the instances of BNEs in the above examples are unanchored. An example of an 'anchored BNE' is found in unit **6.3 of e.g. 6.10** below :

6.10 (INT II)

JH. 6. 1 but e::m.. speaking purely personally **[I]** am very happy **[this year]**
2 that **[we]** have **four female sabbaticals**..
3 and **(a number of.. female executive officers)**..
4 and hopefully **[we]** can e::.. try and make **some progress**..
5 for **{women}**..
6 for **{all women}** at **{the university}**..

The following table shows the OPTWs of all BNEs (anchored and unanchored) in the data of INT and CON :

PARTICIPANTS	INT			CON		
	BNA	BNU	TOTAL	BNA	BNU	TOTAL
HC	9.2	11.4	**20.6**	5.9	8.9	**14.8**
KW	3.5	10.4	**13.9**	4.7	13.6	**18.3**
JH	5.4	9.0	**14.4**	6.3	4.2	**10.5**
BG	5.8	2.9	**8.7**	4.1	11.0	**15.1**
RH	3.1	4.7	**7.8**	3.0	4.4	**7.4**
DD	10.3	7.2	**17.5**	3.6	8.9	**12.5**
RF	7.5	0.0	**7.5**	4.3	0.0	**4.3**
PM	1.5	1.5	**3.0**	5.1	17.1	**22.2**
HK	5.3	2.6	**7.9**	0.0	10.5	**10.5**
KSH	3.9	2.8	**6.7**	3.4	6.8	**10.2**
AM	3.8	7.8	**11.6**	2.0	7.8	**9.8**
CB	2.7	5.5	**8.2**	1.1	6.8	**7.9**
JC	7.0	28.0	**35.0**	2.0	15.9	**17.9**
GM	7.3	8.5	**15.8**	4.1	17.1	**21.2**
AB	6.7	6.7	**13.4**	0.8	10.2	**11.0**
KS	5.0	5.0	**10.0**	0.7	6.7	**7.4**
MEAN	**5.5**	**7.1**	**12.6**	**3.2**	**9.4**	**12.6**
SD	**2.41**	**6.45**	**7.51**	**1.92**	**4.81**	**5.21**
LB	**4.17**	**3.55**	**8.47**	**2.14**	**6.75**	**9.73**
UB	**6.83**	**10.65**	**16.73**	**4.62**	**12.05**	**15.47**
V	**43.8%**	**90.8%**	**59.6%**	**60.0%**	**51.2%**	**41.3%**

Table 6.2 OPTWs of Brand New Entities in INT and CON

Although the MEAN of the total number of OPTWs is the same for INT and CON, the tendency, generally speaking, is for speakers to use more BNEs in INT than in CON. Ten of the speakers out of 16 used more BNEs in the INT encounters, one used an approximately similar number whereas only 5 used more BNEs in CON than in INT. The results, however, are not statistically significant (WSRT $P < 0.45$).

6.5.2.2 Unused entities

This type of discourse entity represents another sub-variety of New entities. However, the difference lies in the fact that while BNEs (anchored and unanchored) are assumed to be totally unknown to the hearer, UEs are known to the hearer in his background knowledge but at

the time they were mentioned were not in his/her 'consciousness' (cf. Chafe 1976:30), as in Prince's example **"Noam Chomsky went to Penn,** where **'Noam Chomsky'** is the unused entity.

Following are two examples from the INT and CON data containing some UEs of the type outlined above (marked with slashes / /) :

6.11 (INT III)

DD. 6. 1 **[I]** was working in **/Aberdeen/..**
2 as **a trainee.. certified accountant**.. for **/Shell/..**
7. 1 and then..-
2 well
3 **[I]**'ve been thinking about **(moving jobs)** for **(some time)**..
4 but.. wasn't.. particularly looking for **(a move)**..
8. 1 and then.. e:::m.. **[my flat mate]** showed **an advert** in.. in **/the Guardian/..**
2 e:::m.. **Ph.D. with a difference** **{it}** said [..
JA. [*(he he)*

6.12 (CON IV)

RF. 32. 1 **[I]**'ve been to **(things)** **[you]**'ve done [..
JA. 33. [mmhmm
2 and **/Adrienne/and/Mary/..**
3 so.. [..
JA. 34. [mmhmm
35. 1 **{it}** was different to go to **(something)**
2 **{that}** **{Teddy}** was.. organising and running..
36. 1 a::nd.. **[I]**.. in fact **[I]** learned a lot about what.. **{the Catholics}** believe..
2 and how **{their faith}** is.. related..
3 **{which}** **[I]** haven't realised..
37. 1 **[I]** haven't realised that often **{they}** don't..
2 **[I]** mean rather than **(the priests)**
3 **{they}** don't tend to read **/the Bible/..**

JA. 38. yes..

UEs seem also to be featured in the initiating questions of the interviewer as he invites the interviewees to talk about a subject which they know of but which is not in their 'consciousness' at the time. The answer the interviewees attempt to give might also include an UE, as in **e.g. 6.13** below. In most cases a change of topic might very likely carry with it an UE or a BNE as in **e.g. 6.14** below :

6.13 (INT V)

JA. 1. 1 O.K..
2 e::m.. **[Karen]**.. **[Helen]**
3 welcome..
2. 1 e::m.. can **[I]** ask first..
2 be..fo::re **[you]**.. ever came to **/university/**..
3 e:::m.. perhaps why **[you]** chose **/Aston/** in **(the first place)**..
4 maybe **[you]** didn't choose **{Aston}** **[I]** don't know..
5 why did **[you]** choose **{Aston}**..

HK. 3. 1 well
2 **[I]**.. e::m.. applied because **[I]**'m doing **/Pharmacy/**
3 and there are only **(a limited number of universities)**
4 **{that}** **[you]** can apply to..

6.14 (CON II)

JH. 20. 1 **(another thing)** **{that}** was funny was
2 e:::m.. at **/the May Ball/**..
3 you know **[we]** have <u>a line-out</u>
4 with [.. **/the V.C./** and.. **/Ch.. the Chancellor/** and..
JA. 21. [mmhmm

/the outgoing President/ and **/the incoming President/**..

The results of the calculations (see **Table 6.3** below) show that speakers use UEs as frequently as they use the other types of New entities, especially so in the conversational encounters. This claim is clearly manifested in the

results of the following table :

PARTICIPANTS	INT	CON
HC	6.9	13.9
KW	5.2	7.0
JH	3.6	13.6
BG	8.8	6.9
RH	1.6	0.0
DD	8.3	1.8
RF	3.0	7.1
PM	1.5	5.1
HK	2.6	0.0
KSH	2.8	27.3
AM	4.9	5.2
CB	0.7	5.7
JC	0.0	10.9
GM	3.6	5.7
AB	0.0	9.4
KS	6.3	6.7
MEAN	**3.7**	**7.9**
SD	**2.78**	**6.57**
LB	**2.14**	**1.29**
UB	**5.23**	**8.51**
V	**75.2%**	**83.1%**

Table 6.3 OPTWs of Unused Entities in INT and CON

Interestingly, unlike the findings in **Table 6.2** above, speakers use UEs considerably more in CON than in INT, indeed more than twice as much. The results are highly statistically significant (WRST $P < 0.01$). The results generally show the tendency of speakers to frequently shift topics and introduce UEs into the speaking floor more in the CON encounters where the turns and units are relatively shorter and new topics are more likely to be introduced at any point during sequential discourse.

6.5.2.3 New Entities in INT and CON : Discussion

Examining the results of the analysis of the occurrence of New entities in the data, which according to Table 6.1 above, represents a general percentage of 5.2% in INT and 6.15% in CON in comparison to the other categories of the taxonomy, it can be seen that the use of New elements in INT and CON is relatively similar as far as the frequency of occurrence is concerned but quite different in their distribution in the two sets of data. As previously mentioned, there is a tendency for speakers to use more New elements in the initiation units of INT and new topics in CON. This does not mean, however, that these entities do not occur at later stages in the discourse. They do occur in various positions but their occurrence in any such position is tied to other factors such as in the case of interruption or self-interruption and the presentation of new ideas into the discourse, as in the following example in which the speaker (JA) suddenly interrupts the silence by saying :

6.15 (CON I)

JA. 251.1 **[you]** were brought up in **/Wales/**
2 weren't **[you]**..

It is noticed that the postgraduate group of students (PGs) used more BNEs than the undergraduate group (UGs). The UGs seem to use more UEs instead. Whether this result has any relationship with the level of educational maturity the PGs have in comparison to the UGs needs a larger size of population than that recorded and analysed for this study. My own intuition, looking at the results above and knowing the participants personally, is that the PGs and their level of advanced political and religious training and the seriousness of talk they handled in the encounters justify their higher use of New topics and new entities in the data.

The V values in the tables above show the presence of variability among speakers in the use of New entities, more with the UEs than the BNEs. Again this is just a tentative suggestion that needs to be verified on a larger body of data.

Since the entity referred to as **"INFERABLE"** can be 'New' in the senses discussed above and since the analysis of data has shown this category to be frequently used by speakers, it is going to be presented in the following sections.

6.5.3 Inferable Entities

Inferable Entities (IEs) represent the second group of entities which are related to the speaker's assumptions about the listener's knowledge. However, they are different from the class of BNEs above in that the speaker assumes that the hearer can infer them by common sense reasoning or logical interpretation. As such, as far as discourse processing is concerned, they posit the greatest load on the processor as they can only be inferred by reference to another entity or group of entities, event, people or situation. These expressions would be totally incomprehensible if such a link could not be made by the listener, after which some form of explanation would be required for them to be understood. Prince (1981:236) considers this type of discourse entity to be the most difficult of all types. In the following example, **(a new battery)** is inferred by reasoning, that a watch (mentioned in previous discourse) needs a new battery. The adjective **'new'** is an **"ATTRIBUTE"** in Prince's terms:

6.16 (CON III)

DD. 88. 1 m.. m.. **[my watch]**..
2 **{it}** **{it}** stopped **[two months ago]**..
3 so.. **[I]** took **{it}**
4 thinking **{it}** needed **(a new battery)**
5 and **{it}** still didn't work properly..

A special subclass of Inferables which Prince distinguishes is the **"Containing Inferables"** (see Section 6.5.3.8 below) where the act of inferencing is done within the Inferable item itself. An example on this type is provided in **Unit 6.1 (the problem of getting in)** of the example below :

6.17 (INT IV)

RF. 4. 1 well
2 [I] think as far as **(accom.. accommodation)** goes..
3 {it}'s **(a lot easier)**..
4 or **(much nicer)** {**living off campus**}
5 because.. **[you]**'ve got.. **(a bigger room)**..
6 usually **(a more comfortable place)** to live in [..
JA. 5. [mmhmm

6. 1 but:: socially {**it**}'s **(the problem of getting in)**..
2 just to pop in for **[half an hour]** to see **(someone)**..
3 is going to take **[you] (an hour and a half round trip)**
4 even.. using **(the car)** as **[I]** do..

The examples of IEs that have up to now been referred to contain many discourse entities that can be either New or non-New according to the position they occupy in the running discourse. They can be inferenced via logical reasoning from subsequent or previous discourse. However, logical interpretation is not the only available means by which discourse entities can be inferenced. There seem to be many other strategies that are available to the listener for interpreting a particular chunk of discourse. These have been considered to be subcategories of the Inferable type of discourse entities. It is worth mentioning here, however, that Prince (op.cit.) does not make such a subcategorisation of the data. Following is a brief explanation of these subcategories with examples from the data for each type :

6.5.3.1 Comparative Inferables (CIs)

There is no explicit mention in Prince's (op.cit.) paper of how comparative discourse entities (i.e. entities that can be inferred by comparing them to other entities in the discourse) are to be treated within the taxonomy although one expects that, like all other discourse entities, they can fall within any of the New or Evoked categories mentioned above. Since in SES the phenomenon of comparison and the expressions

used to represent them are of high recurrence (see for instance **e.g. 6.18** below), it is believed that a special treatment of them is necessary within the taxonomy.

The main reason for considering the phenomenon of comparison under the 'Inferable' category of the AF taxonomy employed in this chapter is that any comparison cannot be drawn without relating one entity to another entity present within the discourse domain. For instance, in **Unit 74.1 of e.g. 6.18** below the speaker KW is comparing her life in Lampeter to her present life at Aston, Birmingham. The listener can infer the entity **'much more fun'** in relation to the entity the comparison is being made with. The same thing applies to the comparisons made in **Units 74.2, 74.4, 75.1, 2, 3, 4** respectively :

6.18 (INT I)

KW. 74. 1 **[I]** think **[I]** had **(much more fun)** in **{Lampeter}**
2 **[I}** had **(much more of a social life)**
3 because.. **[you]** knew **(everybody)**
4 and **{everybody}** went to **(the same parties)**
5 and there was only **(one place)** to go..
6 so **[you]** always made **(the most of what [you] have)**..
75. 1 but workwise **[I]** think **{it}**'s **(so much better)** to be in **(a bigger place)**..
2 because **(everyone)**'s **(more motivated)** to work
3 **(more geared)** towards **(passing {their exams})**..
4 whereas in **{Lampeter}** it was just **(a whole different world)** [..
J. 76. [mmhmm

In the calculations of the CIs in INT and CON the distinction general/particular drawn by Halliday and Hasan (1976:77) (see Section 6.2.2.2) has been adopted but in some more detail. Under **'general comparison'**, for example are a) all instances of identity and similarity between discourse entities that commonly occur with such words as **'same', 'similar to', 'like'**, etc. (e.g. **Unit 75.4 of 6.18** above). b) all instances

of difference that come with such items as **'different', unlike'**, etc. (e.g. **Unit 75.4 of e.g. 6.18** above) c) all instances of exemplification using such connectives as **'like', 'as', 'such as'**, etc. This type has been included under this category although it is realised that they are not cases of comparison proper. But, like comparison, exemplification is usually made in relation to another entity within adjacent discourse and is inferenced accordingly. This is illustrated in the example below :

6.19 (INT VI)

CB. 110.1 {it}'s **(like e:::m.. just a week of.. good fun and entertainment)**..

Included under **'particular comparison'**, which is related to **'quantity or quality'** (ibid), are all other forms of comparison expressed by adjectives, or adverbs (**e.g. Units 74.1, 2 and 75.1, 2 and 3 in e.g. 6.18** above).

As shown in Fig. 6.1 (p. 380), the percentage of CIs from the total number of Inferables is 17.7% in INT and 11.2% in CON, which is relatively high compared to that of the other subcategories of Inferables in the data. As far as the differences between each speaker's output in INT and CON is concerned, the following table shows the OPTWs of all CIs in the two types of speech :

PARTICIPANTS	INT	CON
HC	16.0	7.9
KW	36.5	8.0
JH	12.6	13.6
BG	13.2	10.0
RH	25.2	3.0
DD	14.5	7.1
RF	15.0	8.6
PM	14.9	10.2
HK	10.5	10.1
KSH	11.2	13.7
AM	19.5	4.6
CB	8.9	5.7
JC	12.3	8.9
GM	10.9	4.1
AB	21.6	8.6
KS	15.1	14.1
MEAN	**16.1**	**8.6**
SD	**6.91**	**3.34**
LB	**12.30**	**6.76**
UB	**19.90**	**10.44**
V	**42.91%**	**38.8%**

Table 6.4 OPTWs of Comparative Inferable in INT and CON

The results show the language of INT to contain nearly twice as much of the CIs as in CON. These results are highly significant (WSRT $P < 0.001$). The difference can be ascribed to the interviewees' need to compare and demonstrate the factual entities and topics they are being interviewed about. It also shows the great emphasis the interviewees put on elaborating their speech so as to provide the sort of clarity sought by the interviewer. This phenomenon can occur in conversational discourse but apparently with a lower frequency than in INT.

6.5.3.2 Inferables from previous and subsequent discourse

These represent those entities which can be inferred from the other adjacent entities (hence AIs) within the co-text. The importance of this subcategory, which is highly recurrent in the speech of the informants, lies in the relationships that exist between entities of sequential discourse

and that cause its cohesion and unity as text. However, like that of comparative co-reference, an AI is not in one-to-one relationship with the chunk of discourse from which its reference can be inferenced, as, for instance, in the Evoked Entities (see Section 6.5.4 below), but is related to it in one way or another through class membership, possession and so on. Following are two examples, in the first of which **(e.g. 6.20) 'the baby' (Unit 88.5)** can be inferred from previous discourse and in the second **(e.g. 6.21) 'the problem' (Unit 50.1)** can be inferred from the subsequent discourse, respectively :

6.20 (CON VIII)

KS. 88. 1 o:::h
2 **[I]** don't know..
3 **[I]**'d like to see **(more of /Adrienne/) [next year]**
4 but.. **[I]** mean **{that}** was purely that..
5 **[I]** didn't see **{her}** in **(the first term)** because of **(the baby)**..

6.21 (INT I)

KW. 50. 1 **[I]** think **(the problem)** is **{it}**..
2 **[you]** can only really stick to **(that sort of life)** for **[a year]**..

As will be seen from Fig. 6.1 (p. 380), AIs are highly recurrent in the data of INT and CON. There are altogether 223 instances in INT and 223 in CON, which represent a total percentage of 20.6% of Inferables in INT and 17.8% in CON. This high frequency can be ascribed to the strategies employed by speakers to maintain coherence within discourse by always relating entities to previous and subsequent discourse. It could also be a strategy whereby speakers keep their listeners attentive.

Table 6.5 below shows the OPTWs of AIs in the output of speakers in INT and CON :

PARTICIPANTS	INT	CON
HC	6.9	18.8
KW	15.7	12.2
JH	23.3	15.7
BG	17.5	6.9
RH	28.3	10.4
DD	16.5	28.5
RF	9.0	17.8
PM	11.9	5.1
HK	10.5	13.5
KSH	13.1	13.7
AM	27.3	19.7
CB	12.3	20.5
JC	19.3	10.9
GM	27.9	4.9
AB	22.9	12.5
KS	23.9	13.4
MEAN	**17.9**	**14.0**
SD	**7.03**	**6.15**
LB	**14.03**	**10.62**
UB	**21.77**	**17.38**
V	**39.3%**	**43.9%**

Table 6.5 OPTWs of Adjacent Inferables in INT and CON

Again, the results of the table above suggest that there are more of these entities in INT than CON. 10 speakers out of 16 used more of these entities in INT than in CON, which confirms the suggestion mentioned above about the use of this feature as a strategy to keep listeners alert and attentive. The results are not statistically significant (WSRT $P < 0.12$ level) though. However, there seems to be some variability in the use of these Inferables from one speaker to another similar to that of Table 6.4 above.

6.5.3.3 Inferables from the 'Outside World'

Included under this heading (Situational Inferables (SIs) for short) are those entities that can be arrived at by inference from the 'outside world', i.e. outside the text, which includes the immediate situation within which the talk is taking place. Halliday and Hasan (1976:18) refer to this type as **"Exophoric Reference"** but they exclude it from their analysis of cohesion

in English on the grounds that it is not 'cohesive' since **"it does not bind the two elements together into a text"** (see Section 6.2.2.2 and example thereinwith). This type of reference includes, of course, both First and Second Person Reference (i.e. **I, we, you**) except in certain cases (e.g. quoted speech) (cf. ibid:50), but those are **'Evoked'** rather than **'Inferable'** and have, therefore, been included within the Situationally Evoked Entities (see Section 6.5.4.2 below). The type included in this section covers only those entities that can be inferred by the listener from the surrounding context of situation and the world in general. Following are some examples from the data of INT and CON. Notice first of all the use of **(they)** in the following example :

6.22 (INT I)

KW. 77. 1 **[I]** mean if **(they)** dropped **a nuclear bomb**
2 **[you]** wouldn't have known **{that}**..

In the above example, **(they)** is not 'evoked' in the sense that it does not refer anaphorically or cataphorically to 'people' within the environment of speech. The speaker is using it to refer to some entity in the outside world *(perhaps a foreign nation, an enemy, or some unknown mad aggressor)*. In the following example, the speaker, JH, is presumably reminding the listener of an event with which they were both once involved, which would make him infer the referent of **'the library person'** she had in mind:

6.23 (CON II)

JH. 1. 1 but.. e::m for instance **[I]** liked..
2 **[I]** was pleased **[you]** were **[there]**
3 when **[we]** spoke to **(the library person)**..

This subcategory of Inferables has a total percentage of occurrence of 9.7% in INT and 11.6% in CON, a result which confirms the influence of

situational clues in the interpretation of Inferables within discourse. The following table shows the OPTWs in the speech of each informant in both types of speech :

PARTICIPANTS	INT	CON
HC	4.6	5.0
KW	17.4	10.8
JH	9.0	7.3
BG	7.3	16.6
RH	11.0	10.4
DD	17.6	16.0
RF	6.0	10.7
PM	8.9	15.3
HK	3.9	13.5
KSH	5.6	17.1
AM	8.8	3.9
CB	8.2	10.3
JC	7.0	10.9
GM	7.3	2.4
AB	8.1	11.7
KS	3.8	8.0
MEAN	**8.4**	**10.6**
SD	**4.6**	**4.48**
LB	**5.78**	**8.14**
UB	**10.93**	**13.03**
V	**54.8%**	**42.3%**

Table 6.6 OPTWs of Situational Inferables in INT and CON

Generally, the figures in the table above show fewer occurrences of these Inferables in INT than in CON. 10 of the 16 speakers used less of these Inferables in their INT encounters, the rest using more in their CON speech. These results are not statistically significant on (WSRT $P < 0.12$ level).

6.5.3.4 Inferables from Logical Interpretation

This subcategory of Inferables (Logical Inferables (LIs) for short) is the most recurrent of all in the data of INT and CON. It invloves a group of entities the reference of which can be inferred by logical interpretation. Of the total number of occurrences, it has a percentage of 21.2% in INT and 24.7% in CON (see Fig. 6.1 below). This result is highly substantial and indicative of the involvement of speakers and listeners with each other in communication. It clearly shows the role of certain factors in the process of interpretation of speakers' messages, such as **common sense, reason, logic, shared knowledge** and sometimes, **level of education and specialisation of participants**. Prince (ibid:242) discusses these factors in considerable detail describing them as **"culture-based, involving the Stereotypic Assumptions discussed in Prince 1978b in relation to existential presupposition (e.g. *Houses have doors, People have sides, Kitchens have sinks)*"** (*Italics* in original). Prince also includes in this category of logical inferences **'set to subset'** as in **'*the steps*'** to **'*four*'** in **'*four flights of steps*'**, **'set to member'** as in **'*they* '** to **'*one of them* '**, **'member to set'** as in **'*Jan*'** as a member of some relevant set, presumably a family (ibid:243).

Following are some examples from the data. In the following example, the entity **(the students)** can be inferred by logical interpretation from the fact that the speaker is one of the executive members of the Guild of Students:

6.24 (CON VI)

AM. 8. never..--
9. 1 **[I]** was **[last week]**..
2 **[we]**'ve started **[now]**..
3 sort of every..
4 <u>all the staff</u> are back..
5 so **{it}**'s **[time]**.. **[time]** to work..

CB. 10. yea::h..

AM. 11. 1 so instead of arguing with **(the students)**
2 **[we]** argue with **{the staff}** as well [..
CB. 12. [*(he he he)* yes..
3 **(a great situation)** to be in..

In the following example, **(the children)** can be inferred by logical interpretation from the fact that **'nurseries have children'**.

6.25 (INT II)

BG. 73. 1 **[we]** had **{them}** suggesting..
2 that.. to have **{the nursery}** next to **(the lake)**..
3 where **{the children}** could play..

Following is a table representing OPTWs of Logical Inferables in the subjects' speech in INT and CON :

PARTICIPANTS	INT	CON
HC	16.0	23.8
KW	26.1	23.0
JH	26.9	14.6
BG	19.0	20.7
RH	17.3	20.7
DD	18.6	23.1
RF	13.5	14.3
PM	14.9	30.6
HK	17.1	27.0
KSH	24.3	27.3
AM	23.4	26.9
CB	13.7	14.8
JC	45.5	8.9
GM	18.2	13.0
AB	2.7	28.1
KS	6.3	16.1
MEAN	**19.0**	**20.8**
SD	**9.59**	**6.47**
LB	**13.73**	**17.24**
UB	**24.27**	**24.36**
V	**50.5%**	**31.1%**

Table 6.7 OPTWs of Logical Inferables in INT and CON

Although the MEAN of the two sets of figures suggests that the speakers use this subcategory similarly in INT and CON, observing the results in greater detail, we can see that 12 of the 16 speakers use such Inferables less in their INT speech than in their CON speech, which might be due to the already existing familiarity between each pair of speakers and their shared knowledge of topics discussed within their conversational encounters. However, the results are not statistically significant (WSRT $P < 0.09$).

6.5.3.5. Generic Inferables

This subcategory (GIs) includes all entities of discourse that have a generic sense referring to creatures in general (e.g. **all people are created equal** 'Quirk et al 1985:260), also to things and notions (e.g. **making friends is easy**). They are quite recurrent in SES and may be expressed by nouns or sometimes pronouns. Examples of words used in this sense include *'people', 'humans', 'men', 'women' 'thing'* etc. (see examples below). Related to this subcategory are some pronouns and compound pronouns which might express a generic (or sometimes a specific) meaning, especially those that end with *'body', 'thing'* such as *'something', 'somebody', 'anybody', 'anything', 'everything', 'everybody'* etc. In the following example the entities **'everyone else', 'people', 'new friends'** are GIs.

6.26 (INT VI)

CB. 73. 1 but then of course when **[you]** get **[here]**
2 and realise that **(everyone else)** is on **{their own}**
3 and **{they}**'re all looking for **(people)** to go around with
4 and make.. **(n.. n.. new friends)** and things
5 so.. [**{it}** wasn't ..

In the following example, the entities 'people', 'money' and 'something' can also be considered (GIs) :

6.27 (CON I)

HC. 214.1 are **[we]** educating **{them} (the capitalist system)**
2 to go back and.. keep.. keep **(enough people)** poor
3 so **{they}** can make **(money)**..
4 or are **[we]** saying there's **(something)** wrong with **[us]**..

The following table shows the OPTWs of such entities in INT and CON :

PARTICIPANTS	INT	CON
HC	0.0	10.9
KW	3.5	5.6
JH	1.8	6.3
BG	4.4	8.3
RH	4.7	4.4
DD	4.1	10.7
RF	7.5	5.7
PM	10.4	5.1
HK	3.9	10.1
KSH	3.7	0.0
AM	4.9	4.6
CB	1.4	8.0
JC	7.0	5.0
GM	8.5	10.6
AB	5.4	2.3
KS	2.5	4.7
MEAN	**4.6**	**6.4**
SD	**2.71**	**3.15**
LB	**3.11**	**4.67**
UB	**6.09**	**8.13**
V	**58.9%**	**49.2%**

Table 6.8 OPTWs of Generic Inferables in INT and CON

The results above seem to suggest that speakers favour using this feature more in their informal chat than in INT. The degree of variation as shown by the SD and V figures is more or less similar to those of the other subcategories discussed above. The results are not statistically significant (WSRT $P < 0.06$).

6.5.3.6 Descriptive Inferables

The entities that are included under this rubric (DIs) include modified NPs, the modification itself acting as a clue to the listener to infer the NP reference. This includes those NPs that have vivid descriptions and attributes, especially the ones that normally occur as subject complement (mainly of copular verbs) (cf. Quirk et al. 1985:1173) (Unit **194.3 of e.g. 6.28** below). Another set of examples calculated within this subcategory include those nouns that are figuratively and metaphorically used. Sometimes they occur with exclamative and with imaginative speech (see for example **Unit 31.2, e.g. 6.29** below).

6.28 (INT I)

KW. 194.1 **[I]** had to remember **(all these things)**..
2 because **{it}** gets up on **[your reference]**..
3 being **(a nasty and unsupported little child)**..

6.29 (INT I)

HC. 31. 1 and **[I]**'m trying to remember[*(ha ha)*..
JA. 32. [mmhmm

2 e::m.. that **[I]**'d forgotten how **(big a mountain)****
({a first degree}) can be..
3 when **[you]** are at **(the other end of {it})**..

Also included within the DIs are proverbial and parabolical entities, as in the following two examples :

6.30 (INT VIII)

AB. 121.1 yeah
2 **{he}** wouldn't say **(a boo to a goose of {his own})**..

6.31 (INT VI)

JA. 8. 1 **[I]** thought actually that..
2 e::m.. **[I]** mean **(killing two birds with one stone)** in fact..

A counting of the various features mentioned above reveals the following results :

PARTICIPANTS	INT	CON
HC	16.0	2.0
KW	3.5	5.2
JH	14.4	3.1
BG	4.3	5.5
RH	9.4	0.0
DD	1.0	5.3
RF	7.5	9.3
PM	1.5	5.1
HK	3.9	13.5
KSH	2.8	13.7
AM	4.9	19.7
CB	4.8	20.5
JC	3.5	10.9
GM	4.9	4.9
AB	2.7	12.5
KS	2.5	13.4
MEAN	**5.5**	**9.0**
SD	**4.34**	**6.12**
LB	**3.11**	**5.63**
UB	**7.89**	**12.37**
V	**78.9%**	**68.0%**

Table 6.9 OPTWs of Descriptive Inferables in INT and CON

The results quite obviously indicate that the features of DIs are fewer in INT than in CON. 13 of the 16 speakers tend to follow this pattern. The results are statistically significant (WSRT $P < 0.05$). This seems to suggest that speakers tend to use more vivid and imaginative pictures in their informal chat than when they talk formally in interviews, where they are more controlled with specific topics and are not as free as in conversation. The degree of variation between speakers seems to be relatively higher

than the rest of the subcategories of the Inferables.

6.5.3.7 Other Inferables

Included under this subcategory (OIs) are all other types of Inferables which are not included within any of the above-mentioned subcategories of Inferables. For instance, items that denote **'universality'** and **'partitiveness'** (see Quirk et al 1985:376) have been included. These include items such as *'all', 'the whole of..', 'both', 'another', 'the other..', 'the rest of..', 'part of', 'some of..', 'a bit of..',* etc. These items can be inferred by the listener in relation to other entities or concepts within the discourse realms, which might or might not be present in adjacent discourse. For example the speaker in **e.g. 6.29** above **Unit 31.3** uses the compound entity **(the other end of it)** which can be interpreted in relation to the entity representing the other end (although in this particular example **"the other end of it"** can be a good example of Contained Inferable discussed in Section 6.5.3.8 below. The speaker in the following example had been telling a funny story before she decided to tell another one :

6.32 (CON II)

JH. 20. 1 **(another thing)** that was funny was...

So the entity **'another thing'** can be interpreted in relation to all her previous funny episodes. Notice the use of the expression **'everyone else'** in **Unit 41.7** of the following example, which can be interpreted in relation to all the people who were present in the social gathering the speaker is talking about, although this has not explicitly been mentioned before :

6.33 (INT VI)

AM. 41. 1 {it}'s sort of..
2 if **[you]**'re fairly shy
3 **[you]**'d.. sort of..
4 **(one thing) [you]** can do
5 is just go on.. and say.. [..ts
JA. 42. [mmhmm

6 who are **[you]** [..
CB. 43. [yeah

7 'cos **(everyone else)** is saying exactly **(the same thing)**
8 so **[you]**'d.. don't feel put out..

This type of Inferables is frequent as is seen from the following figures :

PARTICIPANTS	INT	CON
HC	9.2	5.9
KW	0.0	3.8
JH	1.8	12.5
BG	8.8	8.3
RH	25.2	16.3
DD	5.2	10.7
RF	6.0	13.6
PM	11.9	30.6
HK	6.6	16.9
KSH	5.6	13.7
AM	5.8	15.7
CB	9.6	8.0
JC	3.5	4.0
GM	10.9	6.5
AB	1.3	14.1
KS	6.3	13.4
MEAN	**7.4**	**12.1**
SD	**5.84**	**6.55**
LB	**4.19**	**8.50**
UB	**10.61**	**15.70**
V	**78.9%**	**54.1%**

Table 6.10 OPTWs of 'Others' Inferables in INT and CON

Again the results show quite clearly that there is a tendency for speakers to use this subcategory with the informal conversation rather than with the

formal interview. This is to be expected with the expressions that have been calculated for this subcategory as they are characteristic of conversational discourse in general. The results are highly statistically significant (WSRT $P < 0.01$).

6.5.3.8 Containing Inferables

The term 'Containing Inferables' (CONT.Is) is considered by Prince (1981:236) as a special subclass of Inferables, **"where what is inferenced off is properly contained within the Inferable NP"**, i.e. is uncontroversially assumed to normally have what is inferred as a feature. For instance, **'one of these eggs'** (ibid) is inferable, by a set-member inference from **'these eggs'**, which is contained within the NP and which, in the usual case, is Situationally Evoked (ibid).

In the data of INT and CON, a considerable number of the 'Contained Inferables' can be found. From the total number of Inferables in the data analysed, a percentage of 11.7% is found in INT and 7.6% in CON (see Fig. 6.1 below). In the following example from the data, **(an interesting way of doing a degree)** is an example of 'Containing Inferable' :

6.34 (INT I)

HC. 88. 1 e:::m.. {it} was **(an interesting way of doing a a degree)** [..
JA. 89. [mmhmm

2 **[I]** think..

In the following example from CON III **(this machine with a condenser mike)** is an example of Containing Inferable whereas **(the recorder)** in **Unit 58.2** is Inferable from previous discourse or it can also be Inferable by logical interpretation from **(the machine with a condenser mike)** mentioned earlier.

6.35 (CON III)

DD. 58. 1 **[I].. [I]** had **(this machine with a condenser-mike)** on **{it}..**
2 and e:::.. **(the recording)** was awful..

Speakers vary in the way they use this subcategory. Some use it more than others. Others use it in certain conversational settings rather than others. The figures in the following table give a clear picture of this variation. The table represents OPTWs of the Containing Inferables in INT and CON :

PARTICIPANTS	INT	CON
HC	13.7	15.9
KW	8.7	8.9
JH	25.1	6.2
BG	14.6	0.0
RH	17.3	1.5
DD	15.5	3.6
RF	12.0	7.1
PM	10.4	0.0
HK	9.2	3.4
KSH	4.7	6.8
AM	16.6	2.6
CB	2.7	3.4
JC	3.5	11.9
GM	6.1	11.4
AB	10.8	1.6
KS	2.5	2.7
MEAN	**10.8**	**5.4**
SD	**6.22**	**4.62**
LB	**7.38**	**2.86**
UB	**14.22**	**7.94**
V	**57.6%**	**85.6%**

Table 6.11 OPTWs of Containing Inferables in INT and CON

It is interesting to see that Containing Inferables are used twice as much in INT as in CON, which shows, according to the standards set by Prince (op.cit.:250), a greater degree of **linguistic complexity**. The more of these

Inferables are found in a text, the more 'linguistically complex' it is. This will be elaborated on in more detail below. The results in the table above also show a greater of variability in the use of this category between one speaker and another than in the other categories. Two speakers ('BG' and 'PM') made practically no use of this feature in their CON speech although they used it frequently in their INT speech. Altogether 10 of the 16 speakers used it more in their INT output than in their CONs . The V figures in the table above point to the presence of variability between speakers. The results of the table are statistically significant (WSRT P <0.02).

Before leaving the issue of Inferable category, in the next section a summary of the results obtained from the tables above will be given with an account of the frequencies occurring in INT and CON and between one speaker and another.

6.5.3.9 'Inferable Entities' in INT and CON : Summary and discussion

A close observation of the OPTWs of all the Inferables in INT and CON shows that there seems to be at least one inferable entity in almost every unit of discourse. This is a considerably high level of occurrence. The results in the tables above, however, are not consistently similar in the line of findings they offer. They vary in their degree of occurrence first between INT and CON and secondly between one speaker and another as shown in the SD and V figures of the tables. Certain subcategories occur significantly more in INT than in CON, others more in the latter rather than the former. The following figure gives a general summary (in percentage) of all subcategory occurrences in relation to each other :

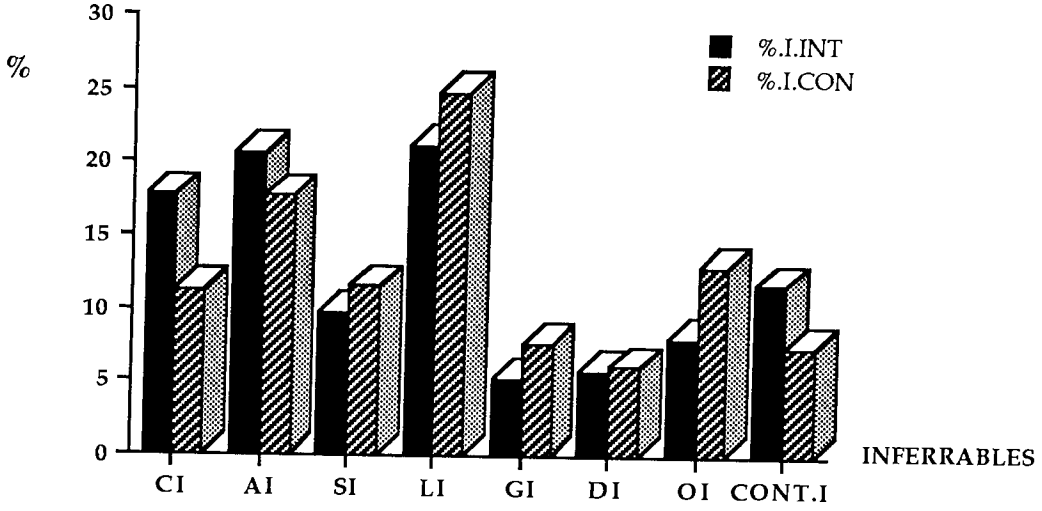


Fig. 6.1 Percentage of occurrence of Inferrable subcategories in relation to each other in INT and CON

It is obvious that three of the subcategories have a considerably higher OPTWs in INT than CON. These are : the **'Comparative Inferables'**, **'Adjacent Inferables'** and the **'Containing Inferables'**. It is interesting to notice that these three subcategories (especially the Containing type) (with a significantly higher occurrence in INT than in CON) represent the highest on the linguistic complexity scale as they all involve the use of complex expressions and phrases. However, their higher use in INT than in CON can be taken as meaning that the convention in INT is to spell out or make more explicit, information to be currently processed. CON, on the other hand, prefers a less explicit style, with the hearer left to infer comparatives, adjacent, and containing entities.

Five subcategories occur slightly more in CON than INT : **(Situational, Logical, Generic, Descriptive** and **Others)**, which seems to suggest that such features can be characteristic of informal conversation. These results are expected with these subcategories since they involve 'shared knowledge' more than the others. Intuitively, one would expect more of shared knowledge within informal conversational talk than the other types of talk.

The linguistic representations of these subcategories do not seem to involve as highly complex linguistic structures as, for example, the 'Containing' or 'Comparative' Inferables. An exception to this may be the 'Descriptive Inferables' which can be represented by 'compounds' or sequences of 'discourse entities and attributes' of the type **"a nasty and unsupported little child"** (e.g. **6.28** above).

Generally, the overall results show that the total OPTWs of **ALL INFERRABLES** in INT is **86.8** and in CON **80.2**. Two important conclusions can be made here. First, that there is a considerable amount of **INFERENCING** in both types of speech under investigation as represented by the high figures above. Secondly, the results show that there is more **INFERENCING** work to be done in INT than CON. This points to another important conclusion and that is related to complexity on a different level from that discussed in various parts of the thesis. The results above show that the language of INT is more complex than that of CON from the point of view of information processing as it lays a heavier processing load on the participants. This in turn suggests that there might be a link between complexity on the linguistic level and on the processing level. But to make such a generalisation with a limited corpus such as the one analysed here is practically implausible and should be left to future research.

Generally speaking, the PG subjects used more Inferables than the UGs in the two discourse types under analysis, which suggests there may be a relationship between educational status and linguistic sophistication in the form of complex structures. They also used more of these Inferables in their INT encounters than in their CON encounters. The UGs, however, yielded results in the opposite direction (i.e. they used slightly fewer Inferables in the INT encounters).

Another interesting finding concerning the PG/UG distinction is that

the PGs showed more statistical significance than the UGs in using more complex structures in INT than in CON. This might be accidental but it can also be due to the greater political or religious training the PGs have (personal communication with the interviewer). Clearly, however, with such a small population, it would be injudicious to claim that this is more than a possibility for future reference.

Taken altogether, the results of the calculations on the Inferable category seem to go in the same direction as those reported by Prince (1981) concerning the (linguistic) complexity of the formal type of discourse in her analysis (i.e. the written text) over the informal oral narrative text. The formal set of data in the present analysis (i.e. INT) displayed more linguistic complexity than the informal set (i.e. CON). This is reflected in the great use of subcategories of Inferables that employ more compound and complex linguistic structures such as the Comparatives and Containing types.

6.5.4 Evoked Entities

As has already been discussed, Evoked Entities are those that co-refer in a direct relationship with other entities in the text or in the context of situation within which the discourse is actually being produced. Halliday and Hasan (1976:14-8) refer to the relationship that holds between an entity and its referent within the text as **"endophoric"** relationship and to that held with an entity and its referent 'outside the text' or in the situation as **exophoric"**. Endophoric relationships can be **"anaphoric"** referring to items in previous discourse or they can be **"cataphoric"** in which case the referent occurs after the entity. Prince (1981) modified this terminology to suit her taxonomy calling the first type (i.e. endophoric) **"Textually Evoked Entities"** and the exophoric type **"Situationally Evoked Entities"** (see Section 6.4 above). In the following section these types of entities will be discussed with reference to the data of INT and CON analysed for this study.

6.5.4.1 Textually Evoked Entities

The entities which fall in the category of 'Textually Evoked Entities' are usually co-referential (Brown and Yule 1983a:192) with another entity already existent in the text. The 'Evoked' entities then are 'Given', referring to another entity that has been 'New' when first mentioned. In the calculations of the TEEs the following subcategories have been distinguished : **Third Person Pronouns, Extrapositional 'it', Repeated NPs , Relative Pronouns and Zero Relative Pronouns, and Others**. Following is a characterisation of these subcategories with examples from the data and statistical counts :

6.5.4.1.1 Third Person Pronouns

This subcategory represents a number of personal, possessive and reflexive pronouns, which are used to refer, mostly anaphorically, to people and objects other than the speaker and the addressee in spoken interactional discourse. That is why, in terms of the functional roles they perform in the communication process, they are referred to as **"OTHER ROLES"** pronouns (Halliday and Hasan op.cit.:45). The following table, adapted from Quirk et al (1985:346), summarise the major forms of these pronouns according to their syntactic functions :

	3rd PERSON PERSONAL PRONOUNS		REFLEXIVE PRONOUNS	POSSESSIVE PRONOUNS	
	subjective case	**objective case**		**determinative**	**independent**
singular masculine	he	him	himself	his	his
singular feminine	she	her	herself	her	hers
singular non-personal	it	it	itself	its	
plural	they	them	themselves	their	theirs

Table 6.12 Third Person personal, reflexive and possessive pronouns

Halliday and Hasan (op.cit.:44) have another category added to the list above, which they call "Generalized Human" (one, ones) as in "**one never knows...**". This has been calculated with the subcategory 'Others' (see Section 6.5.4.1.6 below).

Following are some examples of Third Person Pronouns as used in SES. In **e.g. 6.36**, there are four consecutive occurrences of **'he'** referring to **'this man'** in the first Unit :

6.36 (CON II)

KW. 18. 1 **[I]** was talking to **this man**..
2 **{who}** was catching **(the train)** down to **/North Wales/**..
3 **{which}**.. isn't terribly far from **{where}** **[I]** live really..
19. 1 and **{he}** was saying **{he}** worked in **/the EEC/**..
2 **{which}** don't e:::m.. you know..
3 would **[you]** like to meet **(some other personnel)**..
4 **[I]**'ve just..
20. 1 **[I]** thought perhaps **{he}** might know **(somebody)**..
2 **{he}** wanted **(some personnel)** [..
JA. 21. [mmhmm

Notice the use of the pronouns **'their'** and **'they'** in the following example:

6.37 (INT II)

BG. 88. 1 **[I]** mean **[I]** don't think **{it}**'s..
2 **[I]** think **{it}**'s when.. **{the women}** aren't **[there]**
3 that.. **{they}**'re **(more)**..
4 yeah..
5 **(sexist)** in **{their attitudes}**
6 isn't **{it}** [..
JH. 89. 1 [oh yes..
2 **{they}** **{they}** do make **an effort**..
3 when there's **{a woman}** present [..
J. 90. [mmhmm

Notice also the use of the possessive pronoun **'his'** in **Unit 24.1** which refers anaphorically to **'Don'**

6.38 (CON II)

JH. 24. 1 so.. e:::m.. /**Don**/ comes on **(the other side)** with **{his girl-friend}**..
2 a::nd **[I]** was on **(the other side)**
3 with.. **a bloke** **{who}**.. **[I]** wasn't going out with at **[the time]**
4 but **[we]** were just **(good friends)**..

It has to be emphasised that in the calculations of TEEs all these pronouns were considered in their 'Subjective case', 'Objective case', 'Reflexive' and 'Possessive Determiner'. In the latter case both 'the possessive determiner' and 'the possessed' were treated as evoked entity enclosing both in one type of brackets, { }, as in the use of 'his girl-friend' in the example above.

The subcategory of 3rd Person Evoked Entities is the most frequent of all the TEEs subcategories. The calculations of the total occurrences of this category (see Fig. 6.3 p. 403) show that it has a percentage of 36.5% in INT and 50% in CON. Besides the implication these results yield concerning the 'cohesiveness' (cf. Halliday and Hasan's (1976:52) of the two types of SES under analysis, they also show the tendency of speakers to involve other persons (other than the speaker and addressee(s) in the conversational roles being exchanged during the talk. This might also be an interpretation for the higher use of these pronouns in CON than in INT, as will be seen from the results of the table below :

PARTICIPANTS	INT	CON
HC	27.5	59.5
KW	47.0	47.5
JH	16.2	34.4
BG	26.3	29.0
RH	30.0	26.3
DD	35.1	78.3
RF	25.5	63.5
PM	25.3	71.4
HK	25.0	23.6
KSH	35.5	27.3
AM	40.9	46.6
CB	21.9	62.8
JC	45.5	60.6
GM	42.5	32.5
AB	53.9	39.9
KS	21.4	38.2
MEAN	**32.5**	**46.34**
SD	**10.81**	**17.5**
LB	**26.55**	**36.71**
UB	**38.45**	**55.97**
V	**33.3%**	**37.8%**

Table 6.13 OPTWs of Third Person Pronouns in INT and CON

The results above indicate that the Third Person Pronouns occur significantly more often in CON than in INT. 11 of the speakers above use these pronouns more in their CON talk than in their INT talk. The results are highly significant (WSRT $P < 0.01$). As suggested above, the results confirm the speakers' general tendency in conversational discourse to involve persons not present in the speaking situation with the talk as against the more situationally-tied style of INT where more involvement is to be found between the interviewer and interviewees themselves. In other words there is a tendency for speakers to use 'OTHER ROLES' more often than 'SPEECH ROLES' (cf. Halliday and Hasan op.cit.:45) in CON than INT. This will be checked against the results of the Situationally Evoked Entities presented in Section 6.5.4.2 below.

The degree of variability, though comparatively low in both styles of

speech, is slightly higher within the speech of conversationalists than within that of the interviewees. This is of course expected since speakers in CON differ in their tendencies to talk about other people not present within the speech environment.

6.5.4.1.2 Extrapositional 'it'

The use of *'it'* in its extrapositional (anticipatory) function is different from that of its use as a Third Person in that in the former it has no meaning but to merely signal that the content of the subject is expressed in a later position in the same sentence (Quirk et al 1985:89). As such it has a cataphoric referential function. For instance, ***'it'*** in **Unit 46.2** below refers cataphorically to the whole phrase **'to live here'** :

6.39 (INT I)

KW. 46. 1 and.. **[I]** find **{it}** just..
2 **{it}**'s **(so much cheaper)** to live **[here]**..
3 and **(so much more convenient)**..

(i.e. *to live here is so much cheaper and so much more convenient*).

The use of the pronoun *'it'* in its extrapositional function is common in spoken discourse indicating a more 'focused' position to the substitute for the element being postponed. This is accompanied by a higher pitch movement on the subject complement of *'it'*. The structural manifestations of the elements being focused on after 'it' can be a word, a phrase, a whole clause or even sometimes a series of clauses. Notice, for instance, the use of **{it}** in the following example in which it stands for a series of coordinated clauses :

6.40 (INT VI)

CB. 141.8 and [I] thought {it's} <u>(an ideal opportunity)</u>
9 to go and meet <u>lots of new people</u>..
10 and get {a degree}
11 and.. have **(a good time)** as well..

Sometimes a speaker uses ***'it'*** in its anticipatory function in a succession of units with the span between one and the other being relatively short, as in the following example :

6.41 (INT VIII)

KS. 94. o:::h...
95. 1 e:::m... how important {it} is to get.. to try..
2 and get on with **(the people)** ** **[you]** live with
3 'cos {it}'s very easy to.. make **[your friends]** off **[your course]**..
96. 1 e:::m... and [I] know {it} took **[me]** <u>a good term</u>..
2 {it} was **(the.. beginning of /the Christmas holidays/)**
3 when **(a couple of [us])** had stayed over for **([an extra week])**
4 {that} [I] really got to know **(the girls)**.. ** [I] was living with..
97. 1 and {it}.. {it} really is.. important in.. **(a flat)**
2 when **[you]**'re as in **(as a smaller space)**
3 with **(seven of [you])** [..
JA. 98. [mmhmm

99. 1 but {it}'s e::
2 [I] mean Okay..
3 {it} was **(more difficult for [you])** [..
AM. [*(he he)*

Unlike the Third Person *'it'*, Extrapositional *'it'* is less common in CON than in INT. As Fig. 6.3 (p. 403) shows, the general percentage of the total occurrences for all speakers is 9.1% in INT and only 2.5% in CON. This is also confirmed by the results displayed in the following table :

PARTICIPANTS	INT	CON
HC	9.2	7.9
KW	26.1	1.4
JH	12.6	2.1
BG	10.2	4.1
RH	4.7	3.0
DD	3.1	0.0
RF	9.0	0.0
PM	4.5	0.0
HK	1.3	0.0
KSH	8.4	0.0
AM	7.8	0.0
CB	5.5	1.1
JC	1.8	5.0
GM	6.1	3.3
AB	9.4	0.0
KS	16.3	6.0
MEAN	**8.5**	**2.1**
SD	**6.13**	**2.53**
LB	**5.13**	**0.71**
UB	**11.87**	**3.49**
V	**72.1%**	**120.4%**

Table 6.14 OPTWs of Extrapositional 'it' in INT and CON

Interestingly, seven of the above speakers did not use the extrapositional *'it'* construction in their informal chat but they used it frequently in INT (highly frequently with some of the speakers e.g. RF, KSH, AM). These results, which are highly statistically significant (WSRT $P < 0.001$ level), suggest that this use of anticipatory *'it'* is a feature of formal discourse rather than of the informal type. These are of course conclusions drawn from the limited data analysed for this study, and they need to be verified by examining a larger and more varied body of data of different SES types.

6.5.4.1.3 Repeated Entities

This subcategory of co-referential Evoked entities is important in that it does not involve a replacement of a NP with another item (e.g. pronoun), but it merely involves the use of the same NP again in the same sense, the original one, thus acting as a means to achieve cohesion (cf. Halliday and

Hasan 1976:278). The latter discuss this phenomenon as part of the more general heading of **'REITERATION'** (see Section 6.2.2.1 above). With the exception of 'repetition', nearly all 'REITERATION' processes have been covered in within the subcategories of Inferables (e.g. Inferable from Previous and Subsequent discourse, Section 6.5.3.2 above). Repetition of the same lexical item will be dealt with in this section as it involves, like the rest of TEEs types, a kind of direct 'Evoked' relationship within the same text. In the following example the word **'home'** in **Unit 37.1** has been mentioned several times in the previous text and it is also repeated in **Unit 37.3** :

6.42 (INT V)

KSH. 37. 1 e::m but **[I]** think living away from **{home}** is good..
2 not **(as difficult as) [you]** think **{it}**'s going to be..
3 but **{it}** depends on how independent **[you]**'ve been at **{home}** sort of thing..

On the scale of frequency of occurrence of the Evoked types of entities, the Repetition type comes very high with a general percentage of 28.9% in INT and 26.5% in CON (see Fig. 6.3 p. 403 below). This result, which represents the second in the order of frequency of occurrence after the 3rd Person Pronouns, is high compared to the frequency of the other subcategories of the scale. The high frequency of occurrence of a given repeated item in SES in general has the function of the establishment of topic. Within an interview, the interviewer normally starts the interview with a question about a certain topic. The interviewee's responses and later interviewer questions may contain a repetition of the main items asked about. Sometimes the span between the original occurrence of a word and its repeated occurrence(s) may be relatively long. Notice the use of the word **'pressure'** in the following example, which is introduced by the interviewer for the first time. Notice its frequent recurrence throughout the text in the speech of the two interviewees (7 times within a sequence of 20 units). This phenomenon recurs frequently in CON as well when a

topic under discussion drags on for a period of time. For illustrative purposes, the following example is quoted at length to show the span between one occurrence and the next for the word **'pressure'** :

6.43 (INT VIII)

JA. 57. 1 how much of <u>a pressure</u> to confo::rm..
2 do **[you]** think there is..
3 when **[you]** come to **(university)**..

AB. 58. in what **(way)**..

JA. 59. 1 confo::rm in::..
2 oh **(every way)**..
3 e::: **(like the things)**::
4 **(the:: the::.. a way ** [you] dress)**..
5 e::m **(the way** [you] lead [your life])**..
6 e:: **(moral standards)**..
7 e:: **(what [you] do with [your money])**
8 **(drinking)**..
9 **(drugs)**..
10 whatever..

AB. 60. 1 a lot **(less than when [I] was at {school})**..
2 **({a lot less pressure})** [..
JA. 61. [mmhmm

62.1 e:::m.. at **{school}** **[I]** suppose **{it}**'s **(much more restricted)**..
2 e::m..with **(the number of people)** **{who}** are around..
63. 1 a::nd in a s*..
2 to a sense **[you]** got.. **(much less chance)** to choose **[your friends]**..
3 when **[you]**'re.. sort of **{one}** of.. **(three hundred.. and.. something)**..
4 e:::m whereas where **[you]**'ve got **(several thousand)**..
5 e::m **[you]**'ve got **(a bit more choice)** [..
JA. 64. [mmhmm

65. 1 e::m.. a:nd.. **[you]** tend to::.. make **(friends)** with **(people)**..
2 as soon as **[you]** come in contact with **{them}**
3 **{that}** are **(a lot more similar to [you])** anyway..
4 and so that **{the pressure}** isn't necessarily **[there]**..
5 well
6 not.. from **[my sort of group]**...

JA. 66. **[Kath]**..

KS. 67. 1 e:::m.. **[I]** think **[I]** find that there's **({less pressure})**..
2 to conform in **(some ways)**..
3 **(like in {the way ** [you] dress})**..
68. 1 but maybe in::.. that **[you]** are expected to be in **/the Guild/**..
2 in.. **[I]** mean in **[the first month]** or so anyway
3 at least **[five times a night]**..
4 e:: **five times a week]** [*(he he)*..
JA. 69. [mmhmm

70. 1 e:::m.. and **[I]** e::m.. **[I]**'d find **(that sort of {pressure})** hard
2 because **[I]**'ve never had **{it} [before]**..
71. 1 e:::m... but e:::m **[I]** suppose in **{moral standards}** too..
2 e::m especially in **{the flat}**..
3 where.. **(most of the girls)** at **(one point)** had **a live in boyfriend**..
4 e:::m... **{it}**'s very.. easy to find **[yourself]** under **{pressure}**
5 to at least agree with what **{they}** are doing anyway..
6 **{which} [I]** don't (he--)..

JA. 72. mmhmm..

AB. 73. 1 at **{school}** there was.. **(a similar thort* of sort of thing)**..
2 whereas.. **[you]** couldn't e:::m..
3 **[I]** suppose because **{it}** was **(a public school) (a lot more)** went on..
74. 1 but e:::m... **[you]**:: ...
2 it was a lot **(harder)** to to pull away from **{that}**..
3 and to.. actually disassociate with **[yourself]** with..
4 **(things of a moral standard) {that} [you]** didn't agree with..
5 e:::m whereas at.. **{university} [I]**'ve found **{it} (a lot easier)**..

JA. 75. mmhmm..

KS. 76. 1 **[I]** never had **(that.. sort of.. {pressure})**..
2 in **(that way)** at **{school}** at all..

Following is a calculation of the OPTWs of Repeated items in INT and CON :

PARTICIPANTS	INT	CON
HC	20.6	27.8
KW	20.9	22.6
JH	37.7	24.0
BG	35.1	19.3
RH	44.0	11.9
DD	17.6	17.8
RF	16.5	22.8
PM	17.9	25.5
HK	31.6	13.5
KSH	27.1	34.1
AM	24.4	26.2
CB	12.3	24.0
JC	21.0	35.7
GM	51.0	36.6
AB	20.2	29.7
KS	23.9	24.1
MEAN	**26.4**	**24.7**
SD	**10.75**	**7.14**
LB	**20.49**	**20.77**
UB	**32.31**	**28.63**
V	**40.7%**	**28.9%**

Table 6.15 OPTWs of Repeated Evoked Entities in INT and CON

The table shows a slightly higher recurrence of the phenomenon of Repeated Evoked Entities in INT than in CON with more variability between one speaker and another. This result, however, is not statistically significant on (WSRT $P < 0.5$). The results are perhaps due to the fact that speakers during INT centre their talk around particular topics for a period of time within which certain items related to those topics recur many times (see **e.g. 6.43** above) and the frequent recurrence of such words as **'pressure' 'school', 'university', 'dress', 'flat'** and so on.

6.5.4.1.4 Relative Evoked Entities

These include all the relative pronouns and zero relative pronouns that are used in the conversation of speakers. These are represented by the relative pronouns *'who', 'whom', 'whose', 'which'* and *'that'* and also the other 'wh-' pronouns when they occur as relative pronouns, such as *'when, 'where', 'why'* and so on. Prince has not made any mention of how these pronouns would fit into her taxonomy. In the present classification, they are taken to be co-referential, mostly anaphorically, with the NP they are used to refer to. For instance, in the example :

6.44 (CON I)

KW. 18. 1 **[I]**was talking to **this man**..
2 **{who}** was catching **(the train)** down to **/North Wales/**..
3 **{which}**.. isn't terribly far from **{where} [I]** live really..

the relative pronoun **'who' (Unit 18.2)** refers anaphorically to **'this man'** in the preceding unit, **'which'** refers to **'North Wales'** and **'where'** refers to the place in which the speaker lives. So all these pronouns have reference to 'antecedents'. As such they are co-referential and should be included within any taxonomy of discourse reference. This is confirmed by Quirk et al (1985:365) who state that :

> "Relative pronouns resemble personal pronouns in that they have coreference to an antecedent"

and also by Gutwinski (1976:65), who includes them in his model of analysis for the study of cohesion in English. It has to be mentioned here, however, that Halliday and Hasan (1976) make no mention whatsoever of the co-referential effect of the relative pronouns in their study of cohesion in English, neither do they explain why they do not do so.

Zero relative pronouns have the same co-referential effect as that when the relative pronoun is present and since all instances of zero anaphora are counted as cohesive ties in the same way as the occurrences of anaphoric pronouns (Gutwinski op.cit.:62), hence their inclusion within statistical framework of information structure.

Altogether, relative constructions (i.e. relative pronouns and zero relative pronouns) constitute a percentage of 13.7% of the TEEs in INT and only 7% in CON. The significant difference between INT and CON in the use of this phenomenon is highly suggestive and points to the more complex syntactic nature of INT as compared to CON (see also the results of the syntactic differences between INT and CON in CHAPTER FIVE above). The percentages mentioned above include the use of both relative pronouns and their zero occurrence. The following figure shows the use of each in INT and CON :

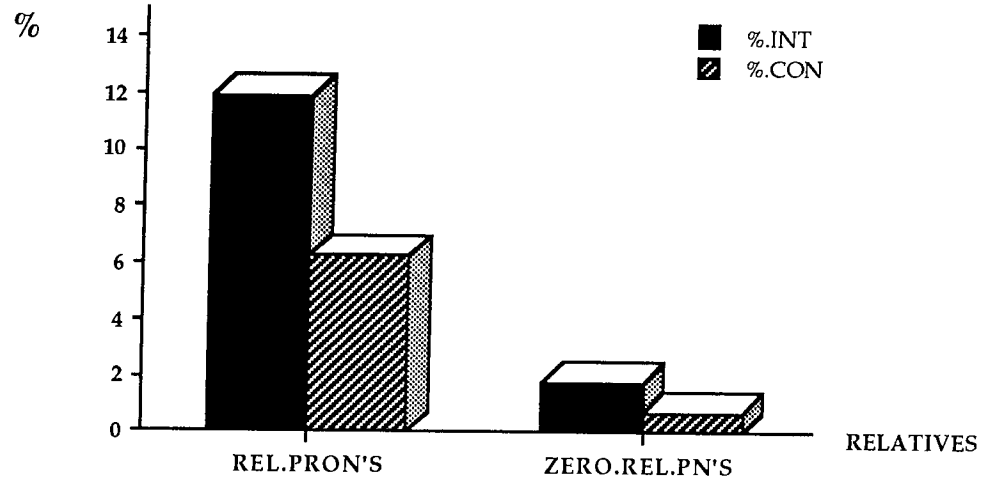


Fig. 6.2 Percentage of Relatives and Zero Relatives in INT and CON

The variability in the use of the Relative and Zero Relative Evoked Entities in OPTWs is displayed in the following table :

PARTICIPANTS	INT	CON
HC	36.6	9.0
KW	10.4	7.0
JH	12.6	8.3
BG	17.6	2.8
RH	9.4	5.9
DD	16.6	8.9
RF	12.0	12.1
PM	6.0	5.1
HK	6.6	3.4
KSH	6.5	10.2
AM	18.5	2.6
CB	6.9	3.4
JC	0.0	8.0
GM	9.7	4.9
AB	25.6	9.4
KS	10.0	4.1
MEAN	**12.8**	**6.6**
SD	**8.79**	**2.94**
LB	**7.96**	**4.98**
UB	**17.64**	**8.22**
V	**68.7%**	**44.5%**

Table 6.16 OPTWs of Relative and Zero Relative Evoked Entities in INT and CON

These results are highly significant (WSRT $P < 0.01$ level). They confirm the previous results about the tendency of speakers to use Relative Evoked Entities and Zero Relative Pronouns significantly more in INT than in CON. Few examples of this subcategory are found in conversation and fewer still of the zero relative type. The reason for this, as has already been mentioned above, is the more complex nature of INT talk as regards that of CON. Also, the tendency seems to be that in INT people use more postmodification and elaboration of NP through the use of restrictive relative clauses than in CON (see CHAPTER FIVE above).

6.5.4.1.5 Textual deictic demonstratives

The various uses of the demonstrative pronouns *'this'*, *'that'*, *'these'*, *'those'* as deictic items have been well covered in linguistic (and other types of) research (see for example, Lakoff 1974; Halliday and Hasan 1976;

Fillmore 1975, 1982; Lyons 1977; Farag 1986). Generally, three types of these uses are distinguished in the literature :

1) their use as indicators of **spatio-temporal deixis**, essentially a form of verbal pointing (i.e. *'this', 'these'* **near**, *'that', 'those'* **far**), as in the following example :

6.45 (CON VIII)

KS. 242. is **[that] ([the door])**..

AB. 243.1 well..
2 **{it}** looks like **{the one}**..

Since this use is exhophoric (i.e. refers to humans and objects in the situation) (cf. Halliday and Hasan op.cit.:57), they have been included under the "Situationally Evoked Category" (see Section 6.5.4.2 below);

2) their use as **'emotional deixis'** items (cf. Lakoff op.cit.:346) which are **"linked to the speaker's emotional involvement in the subject matter of his utterance"** (ibid). In the following example, two such uses of the pronoun ***'this'*** are found in **Unit 9.2** and **Unit 13.1** below :

6.46 (CON I)

KW. 9. 1 e:::m.. **{it}**'s really strange..
2 **[I]** met **this fellow** on **/Euston Station/** [*(he he)*..
JA. 10. [*(he he)* yes..
11. 1 **[I]** got **an interview** at **/Trust House Forte/**..
2 e::.. **{which}** is absolutely awful..
3 **(people)** were just dreadful [..
HC. 12. [yeah
13. 1 **[you]**'d never believe that **{they}**'d actually done **this course**..
2 because **{they}** just seemed to have no **idea** of how to interview **[you]**..

This use can be Brand New as in the two instances of the demonstrative ***'this'*** above (see also Section 6.5.4.1.1) or Inferable (see Section 6.5.4.2 above) if the hearer can infer its meaning from the surrounding discourse

or from any other source (see, for example, the use of ***'this'*** in **Unit 35.1** of **e.g. 6.47** below).

3) The third use of these demonstratives, **discourse deixis** concerns their reference to some portion of the discourse that contains them. This includes those instances of coreferentials with prior (anaphoric) or subsequent (cataphoric) discourse(6). Halliday and Hasan (op.cit.) study them under **"Endophoric Reference"** (7). In the following example, the demonstrative **'that'** in **Unit 36.1** and that in **36.2** refer anaphorically to **'the majority of women students at the university do social sciences'** expressed in **Unit 33.3**. Notice also the use of the demonstrative **'those'** in **Unit 37.1**, which refers to **'social sciences'** in the preceding units :

6.47 (INT II)

BG. 33. 1 oh
2 {it}'s a bit unfortunate..
3 **(the majority of {the women students})** at **{the university}** do **{social sciences}** [..
JA. 34. [mmhmm

35. 1 and there's always **(this dividing line)**..
2 that.. **[we]**'re **(science students)**
3 therefore**[we]**'re **(superior to [you])**..
36. 1 a::nd.. there's **(nothing) [you]** can do about **{that}**
2 because **{that}** is **(the way)** **
3 **(society)** has made **(a science degree)**..
4 **(more valid than..than {social sciences})**..
37. 1 and **(the way)**.. ** **{women}** have always been pushed towards **{those}**..
2 there's **(nothing) [you]** could do about **{it}**
3 until **(the intake of students)**.. comes **(differently)**...

Altogether, the textual deictic demonstratives in their coreferential function have a percentage of occurrence of 11.3% in INT and 11.6% in CON among the other subcategories of TEEs (see Fig. 6.3 below). The differences in the use of these pronouns in INT and CON are displayed in the following table :

PARTICIPANTS	INT	CON
HC	4.6	7.9
KW	5.2	9.9
JH	5.4	12.5
BG	17.5	1.4
RH	18.9	7.4
DD	3.1	3.6
RF	25.5	10.0
PM	4.5	5.1
HK	3.9	3.4
KSH	5.6	0.0
AM	9.7	10.5
CB	9.6	9.1
JC	7.0	13.1
GM	23.1	11.4
AB	6.7	17.2
KS	11.3	23.5
MEAN	**10.1**	**9.1**
SD	**7.20**	**5.97**
LB	**6.14**	**5.82**
UB	**14.06**	**12.38**
V	**71.3%**	**65.6%**

Table 6.17 OPTWs of Textually Evoked Deictic Demonstratives in INT and CON

The results do not show a great difference between the employment of demonstrative pronouns in INT and CON and they are not statistically significant (WSRT $P < 0.5$). There is no systematic difference between the two speech types. 50% of the speakers used these pronouns more in CON than in INT whereas the remaining 50% used them more in the INT sessions. However, there seems to be a great degree of variability between one speaker and another in the use of these demonstratives, as shown by the high level of SD and V values displayed above.

6.5.4.1.6 'Others' Textually Evoked Entities

What remains to be calculated of the TEEs are a number of other features which refer to other entities in the same text but have not been included within any of the subcategories mentioned above. These are mainly represented by the indefinite pronoun *'one'* and its main derivatives when used as substitutes for NPs , mostly anaphorically. These are not very recurrent in the data but they have been noticed to occur frequently in the speech of particular speakers rather than others. They have a general occurrence of just 6 instances in INT and 37 in CON making a percentage of 0.5% in INT to 2.5% in CON. Following is an example taken from CON III, in which the speakers are chatting freely about watches. All subsequent uses of 'one(s)' refer to 'watches' in the preceding discourse :

6. 48 (CON III)

RH. 114.1 {**it**} was..
2 for **[me]**.. {**it**} was **[ages]** before **[I]** had {**one**}..
3 but when **[I]** did get {**one**}..
4 **[I]** didn't have {**it**} very long
5 before **[I]** lost {**it**} [..
JA. [*(he he he)*

6 and {**that**}.. {**that**} has never been forgotten..
7 {**it**}'s.. {**it**}'s always.. e:::m {**a new watch**}..
8 what happened with {**the old one**}.. [
DD. 115.1 [*(he he)*.. yeah
2 yeah..
3 well
4 **[I]** had **(a Timex)** when **[I]** was twelve..
5 or **(some thing like {that)**..
6 and {**that**}was **(a big thing)** [..
RH. 116.1 [yes..
2 mmhmm

117.1 **[I]** never broke {**one**}..
2 then **[my brother]** got {**one**}
3 when {**he**} was eleven..

4 then **[my sister]** got **{one}**
5 when **{she}** was **(nine or something)**
6 and.. you know..
7 so **{it}** goes down **(the family)**..
8 **(younger and younger)** and e:::..

The use of these expressions in the speech of certain speakers rather than others indicates that they can be style specific like other features mentioned in the previous chapters. This needs to be checked against a larger population.

6.5.4.1.7 Textually Evoked Entities in INT and CON : Discussion

In this section a summary of the findings of the previous sections concerning the frequency of occurrence of Textually Evoked Entities is presented and discussed. The tables above show 3rd person pronouns to be the most recurrent in the data followed by the subcategory of 'Repeated' items and then the textual demonstratives. The other subcategories show a greater degree of variability between speakers. The most significant differences between INT and CON are noticeable in the subcategory of 3rd Person which tend to occur more in CON than in INT. With the exception of the OPTWs of Repeated items, which occurs equally frequently in INT and CON, the remaining subcategories occur significantly less in CON than in INT. Again the results do not show a consistent pattern of occurrence in one of the SES type rather than another. The results seem to vary considerably between one person and another and between one SES type and another. The following figure shows the percentages of each of the subcategories in relation to each other. They generally confirm the results of the tables reported above :

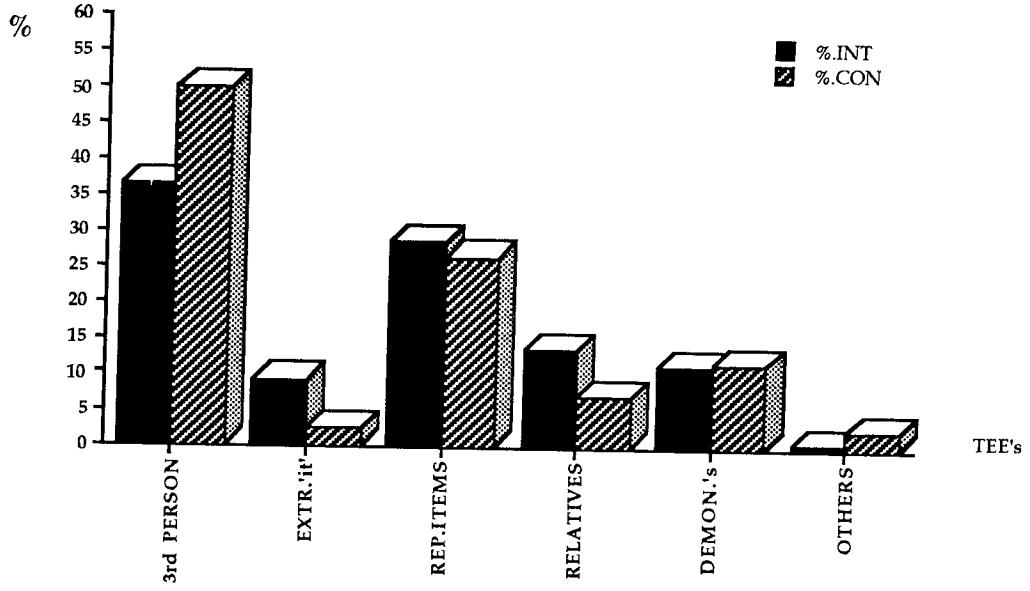


Fig. 6.3 Percentage of Textually Evoked Subcategories in INT and CON

Since the Textually Evoked Categories act as cohesive devices that tie the elements of discourse together (cf. Halliday and Hasan 1976), their high frequency of occurrence points to the high degree of cohesion and unity that characterises any piece of SES when exchanged between two or more speakers. Notice also that the results above tend to confirm the results reported in the previous sections concerning the finding that, the language of INT is more linguistically complex than that of CON. The results above show this finding to be true on the level of information structure as well. Whereas speakers in CON tend to employ more personal pronouns within their speech to refer to other entities in discourse, they tend to use more complex strategies in INT than in CON. For example, they use more of extrapositional *'it'*, more relatives (which is an indication of the great use of qualificatory clauses and phrases), and more co-referential deictic demonstratives. These findings will be touched upon in the final discussion of the results in the few sections that follow.

As far as the factor of educational maturity is concerned, again the PG students gave more significant differences between their INT and CON output than the UGs in the higher use of more complex structures in INT

than in CON. The results go in line with the previous results reported above in the use of the New Entities (Section 6.5.2) and the Inferable Entities (Section 6.5.3 below).

6.5.4.2 Situationally Evoked Entities

The main components of the Situationally Evoked Entities that have been distinguished and studied within the present taxonomy involve a group of personal pronouns, a number of time and place situational deictic adverbials and some other items that can only be interpreted with reference to the salient features of the situational setting within which the talk is taking place, including the participants themselves. The personal pronouns that can be taken to be situationally evoked are those that are classified by Halliday and Hasan (1967:58) under the category of **'SPEECH ROLES'**, i.e. they refer to the speaker(s) and addressee(s) themselves. In other words they are represented by 'First Person' and 'Second Person' pronouns and their derivatives of possessives, determiners and reflexives (see **Table 6.18** below). It is important to stress that these entities are exophorically interpreted (ibid) and as such have no reference to entities within the text per se. This is why Halliday and Hasan (ibid) do not consider them as cohesive elements within the discourse and exclude them from their study of cohesion in English. There are some exceptions though and these will be touched upon in the following sections when each of these subcategories will be considered on its own. The following table adapted from Quirk et al (op.cit.) shows the 1st and 2nd Pronouns in English :

PERSONAL PRONOUNS				POSSESSIVE PRONOUNS	
	SUBJ.	OBJE.	REFLEXIVE PRONOUNS	determin-ative	independent function
1st PERSON					
singular	I	me	myself	my	mine
plural	we	us	ourselves	our	ours
2nd PERSON					
singular	you	you	yourself	your	yours
plural	you	you	yourselves	your	yours

Table 6.18 1st and 2nd Personal,Reflexive and Possessive Pronouns in English

6.5.4.2.1 First Person Pronouns

These include the Subject pronouns *'I', 'we',* Object pronouns *'me', 'us',* Possessive Determiners *'my' 'mine', 'our', 'ours'* and Reflexive pronouns *'myself', 'ourselves'* respectively. The majority of the uses of these pronouns are exophoric referring as such to the speaker or addressee. Notice the use of *'I', 'me'* and *'my'* in the following example :

6.49 (INT I)

HC. 79. 1 e:::m.. well
2 {it} was **(different for [me])**
3 in that **[I]** lived in **a Presbytery**..
80. 1 **[I]** worked in **a parish**
2 while **[I]** was doing..e:::m **[my degree]**..
3 **{which}** e:::m... gave **{it} (a different aspect)**..

Frequently in SES the pronoun *'I'* occurs in 'a chain' with many repetitions of the pronoun within one unit. Notice, for example, the many uses of **'I'** in **Units 170** and **171** of **e.g. 6.50** below. Notice also that not all uses of these pronouns are exophoric. When they occur in 'quoted speech', for example, they are 'endophoric' and refer to 'a third party' other than the speaker and addressee who might not be present in the

speaking situation. The pronoun **'I'**, in **Unit 171.1** of the following example has the same referent as **'he'** in the preceding discourse (quoted speech) :

6.50 (CON I)

KW. 170.1 **[I]** wrote **[my first essay]**
2 and **[I]**.. went along to **(the tutorial)**..
3 **[I]** went into **(the room)**..
4 **[I]** sat down and **{he}** said
5 **(Catholic)**..
6 **[I]** thought **[I]** felt a lot of horns you know blazing on **[my forehead]** *(he he)*..
171.1 **{he}** said **[I]** read **[your essay]**
2 and **[I]**'ve never read **(such.. bigoted rubbish)** in **[all my life]**.. *(he he)*
3 and **[I]** thought..
4 but **[I]** spent **[seven years]** being taught **{this}**
172.1 **[I]** never really thought to argue..
2 'cos **[you]** don't..
3 **[you]** mustn't argue with **{nuns}**..

The pronoun ***'we'*** has many uses besides its being a pronoun referring to the speaker + addressee (see Quirk et al 1985:350, for a review of these uses). In general, it is widely used in SES in its 'membership' sense (cf. Farag 1986:123) referring to the (addresser + they) where ***'they'*** are not necessarily present. In this use it combines features of 'SPEAKER ROLES' and 'OTHER ROLES' (see Halliday and Hasan op.cit.). ***'We'*** has also been noticed to occur frequently by speakers who have some sort of authority, with an assumption of status behind it(8), (sometimes in a chain), as in the following exchange between the interviewer and the Vice-President of the Guild of Students at the university :

6.51 (INT II)

JA. 76. 1 so what would **[you]** try and do..
2 **[I]** mean **[you]**'ve only got **[twelve months]**
3 haven't **[you]** [..
BG. 77. [mmhmm

JA. 78. **{it}**'s.. *(he he)*..

BG. 79. 1 well
2 **[I]** don't think **[we]** can d'..
3 **[we]** can change **(anything)** overnight like **{that}**..
80. 1 **(the only thing)** ** **[we]** can do is..
2 that **[we]** are **{women}**..
3 **[we]** are going on to **{those meetings}**..
81. 1 and **[we]** have just got to show that..
2 that **[we]** are quite capable
3 and to be respected for..what **[we]** are..
4 **[our.. our roles]** in **{the Guild}**..

'We' can also be used in quoted speech, in which case it does not refer to the speakers and addressees in the situation, as in the following example :

6.52 (CON I)

KW. 14. 1 when **[I]** asked what **salary structure** **{they}** had in **(mind)**..
2 **{they}** said
3 o:::h
4 **[we]** haven't really thought about **{it}** yet..

All uses of personal pronouns which are not 'situational' are not included in the calculations which will be presented below. The general results indicate that the Subject pronouns have the highest degree of occurrence within the SEEs in both INT and CON. They have a general percentage of 55% and 52.1% in INT and CON respectively. Of this percentage, the pronoun *'I'* has a greater frequency of occurrence than that of *'we'* in the data. Fig. 6.5 (p. 415) shows the percentage of occurrence of *'I'* *and* *'We'* in relation to the other SEEs pronouns in INT and CON. As the figures show, 49.2% of *'I'* occurs in the INT encounters compared to just 5.8% of *'We'*. In CON, *'I'* has an occurrence of 43.2% compared to 8.9% of *'We'*.

These figures represent general results for the MEAN of occurrences for all speakers. '1st Person Others' include all non-Subject pronouns such as *'me', 'my', 'mine', 'us' 'our', 'ours'* and so on. These have a percentage of 6.1% in INT and 6.9% in CON.

The following table shows the OPTWs of the 1st Person Pronouns occurring as SEEs in INT and CON for each speaker. The calculations include all Subject and Non-Subject 1st Person pronouns in the data :

PARTICIPANTS	INT	CON
HC	66.3	78.3
KW	57.4	49.3
JH	50.2	78.3
BG	76.0	52.5
RH	47.2	54.8
DD	43.4	76.5
RF	48.0	60.6
PM	32.7	66.3
HK	80.3	50.7
KSH	55.1	92.1
AM	100.4	34.8
CB	68.6	42.2
JC	50.8	81.4
GM	71.7	35.0
AB	31.0	53.2
KS	79.1	79.9
MEAN	**59.9**	**61.6**
SD	**18.80**	**17.77**
LB	**49.56**	**51.82**
UB	**70.24**	**71.38**
V	**31.4%**	**28.8%**

Table 6.19 OPTWs of 1st Person Situationally Evoked Entities in INT and CON

The figures above do not show a significant difference in the use of 1st Person Pronouns in INT and CON ; (WSRT $P < 0.3$). However, they show a great degree of variability between one speaker and another, as shown by the results of SD for both SES types.

6.5.4.2.2 Second Person Pronouns

The 2nd Person Pronouns (see **Table 6.18** above) are highly recurrent in SES, especially so the pronoun ***'you'*** both in its generic and specific senses. It seems a feature of English conversation to have speakers using this pronoun to express self-reference or general reference, often by implication to mean **"any individual I would approve of"** (Halliday and Hasan 1976:53). This use sometimes occurs in a continuous chain, as in the example below :

6.53 (INT I)

KW. 5. 1 and when **[you]** are asked to move
2 **[you]**'re given.. absolutely **no notice**..
6. 1 it's as if.. **(the housekeeper)** has no **(idea)**
2 that when **[you]**'re doing **a research project**
3 that **[you]** do actually..
4 'cos.. alter**[your days]**..
5 that **[you]** fit **(everything)** in around **{it}**..
7. 1 only when **[you]**'re given **(a forty-eight hour notice)** to move
2 **{it}**'s difficult to rearrange **(everything)**..

The use of ***'you'*** to refer directly to the addressee is not as frequent as the use of generic ***'you'***. However, it is used frequently in the interviewer's questions, as in **Unit 76. of e.g. 6.51** above.

Generally speaking, ***'you'*** occurs frequently in INT and CON with a percentage of 22.6% in the former and 20.5% in the latter. What is interesting though is the distribution of the generic/non-generic usages within these percentages. The following simple figure is obviously self-explanatory :

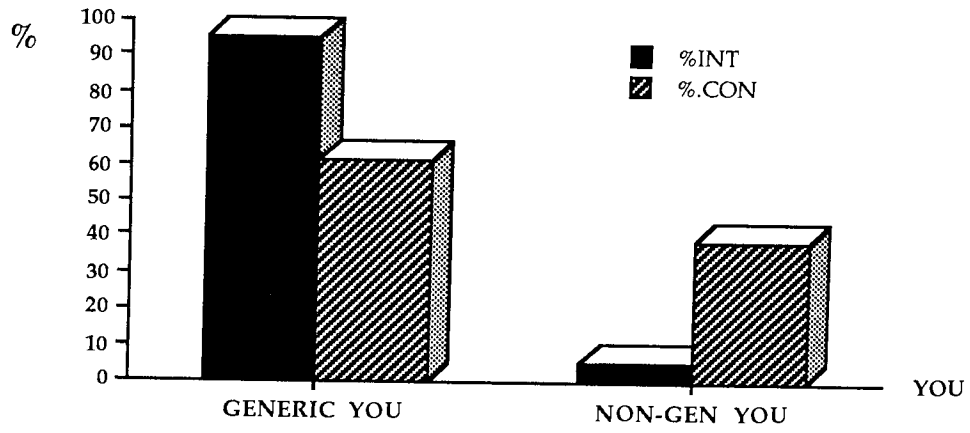


Fig. 6.4 Percentage of Generic and non-Generic 'YOU' in INT and CON

Interestingly, the generic use of ***'you'*** seems to be significantly greater in the language of INT but quite less so in that of CON. This means that a higher use is found in CON than in INT of ***'you'*** which is directed to the addressee. This is perhaps due to the informality of the situation during casual chat and the interactional and cooperative nature of talk. Generic ***'you'***, on the other hand, seems to be a characteristic feature of more formal talk within which direct reference and illustration is drawn on 'other individuals' rather than the addressee present in the situation per se.

The following table shows the OPTWs of the 2nd Person Pronouns in the speech of subjects in INT and CON :

PARTICIPANTS	INT	CON
HC	11.4	10.9
KW	34.8	43.7
JH	16.2	23
BG	7.3	40.1
RH	15.7	20.7
DD	6.2	17.8
RF	42.0	14.3
PM	38.7	40.8
HK	27.6	27.0
KSH	41.1	34.1
AM	53.6	26.2
CB	16.5	14.8
JC	22.8	34.8
GM	13.4	11.4
AB	41.8	14.1
KS	27.6	11.4
MEAN	**26.0**	**24.1**
SD	**14.50**	**11.48**
LB	**18.02**	**17.78**
UB	**33.98**	**30.42**
V	**55.8%**	**47.6%**

Table 6.20 OPTWs of 2nd Person Situationally Evoked Entities in INT and CON

Again, the difference between speakers' performance in INT and CON is not at all significant (WSRT P < 0.4) although 9 of the 16 speakers tend to use these pronouns more in their INT speech. However, the significant differences appear on the individual levels and in the function of the pronouns themselves rather than in their overall OPTWs. The results of Fig. 6.5 (p. 415) confirm this conclusion.

6.5.4.2.3 'Others' Situationally Evoked Entities

The remaining entities of those that have been taken to be SEEs in the present research represent what is usually referred to as **'Time Deixis'**, and **'Place Deixis'** (see Levinson 1983:73) and also some other types of entities that have not been and indeed cannot be classified under any of the subcategories discussed above. Included under the first of these

subcategories are such time adverbials as *'now', 'then'* (in their demonstrative function) *'yesterday', 'today', 'last year'*... etc. Levinson (op.cit.) states that **"like all aspects of deixis, time deixis makes ultimate reference to participant-role"**. As such they have direct relevance to the situation and the time the utterance is being produced. The second of these subcategories, the **'Place deixis'** concerns **"the specification of locations relative to anchorage points in the speech event"** (ibid). These include such place adverbials as *'here', 'there'* (also in their demonstrative function), and objects and locations pointed to in the situation by the use of other demonstratives such as *'this', 'that'* etc. The **'vocatives'** (see Quirk et al for definitional comments and details of types and specifications), which include calling people by their names, appellatives (e.g. *Daddy, Auntie*), titles of respect (e.g. *Sir*), markers of status (e.g. *Father 'priest'*) and so on. All these are related to people in the context of situation in one way or another and are, hence, included within the SEEs. Following are some examples from the data on the use of some of these deictics and vocatives :

6.54 (INT IV)

PM. 101.1 in fact { **people**} {**who**} **[I]**.. *(he he)*..
2 **(the main group of friends) [I]** have **[now]**
3 are at {**the university**}..
4 **[here]**..
5 so.. **[I]** haven't got **(any links)** e::m in **[my old home]**..

6.55 (CON II)

BG. 135.1 **[now] [I]** would like to be with **(some first years) [next year]**..
2 **[I]** think {**it**}.. {**it**} keeps **[you]** in touch..

6.56 (CON I)

KS. 81. 1 **[you]**'re not going to want to hear **{this}**
2 **[John]**..
JA. 82. 1 no.. no..
2 but **[you]** may..
3 **[I]** may not want to hear **{it}**..
4 **{it}** might be..
5 **[I]**'ll be glad as well..

KS. -- -- [–
JA. 83. 1 [Okay
2 is **{that}** clear..

KS. 84. 1 yes
2 **[sir]**..
3 yes..

Counting the general occurrences of the above-mentioned deictic items and vocatives, we notice that Time deixis SEEs have the highest percentage of occurrence within the types examined under this category (i.e. Place deictics 'Others'). Fig. 6.5 gives details of these percentages. The figure shows that the percentages of these subcategories are significantly higher in CON than INT. The following table shows the OPTWs of these features in INT and CON :

PARTICIPANTS	INT	CON
HC	4.6	19.8
KW	15.7	17.9
JH	18.0	16.7
BG	14.6	12.4
RH	11.0	4.4
DD	15.5	8.9
RF	12.0	33.5
PM	8.9	10.2
HK	10.5	13.5
KSH	13.1	3.4
AM	15.6	23.6
CB	12.3	11.4
JC	8.5	15.9
GM	20.7	11.4
AB	6.7	10.2
KS	7.5	26.2
MEAN	**12.2**	**15.0**
SD	**4.37**	**7.96**
LB	**9.80**	**10.62**
UB	**14.60**	**19.38**
V	**35.8%**	**53.07%**

Table 6.21 OPTWs of Time Deixis, Place Deixis and 'Others' SEEs in INT and CON

Eleven of the speakers above show the tendency to use more of these expressions in their CON output than in their INT output but the results are not significant from the statistical point of view (WSRT $P < 0.17$). Since all three categories represent types of location across space and time, their common use in speech is an indication of facilitating the process of listener comprehension and ultimately achieving the goal of communication. They represent features of simple, easy-to-comprehend discourse as they themselves help to provide clues to the listener for understanding such questions as *'who'*, *'when'* and *where'* delivered within the communicational message.

6.5.4.2.4 Situationally Evoked Entities in INT and CON : Discussion

None of the subcategories counted under the SEEs showed any statistical difference between speakers' output in INT and CON. The use of the linguistic elements within each of these categories is equally recurrent in both types of SES. There is a significant difference, however, in the use of 'Generic' and 'Non-Generic' *'you'* in INT and CON. Speakers tend to use the generic type very frequently in their INT output but less so in their CON output. They tend to use the non-generic type more in their CON speech than in their INT speech, however.

The percentages of use of the subcategories of the SEEs are presented in the following figure, to which there has been reference at many points of the discussion. The figure yields clues to the understanding at a glance of the size of each of the subcategories of the SEEs in relation to each other and its recurrence in the two types of SES under investigation.

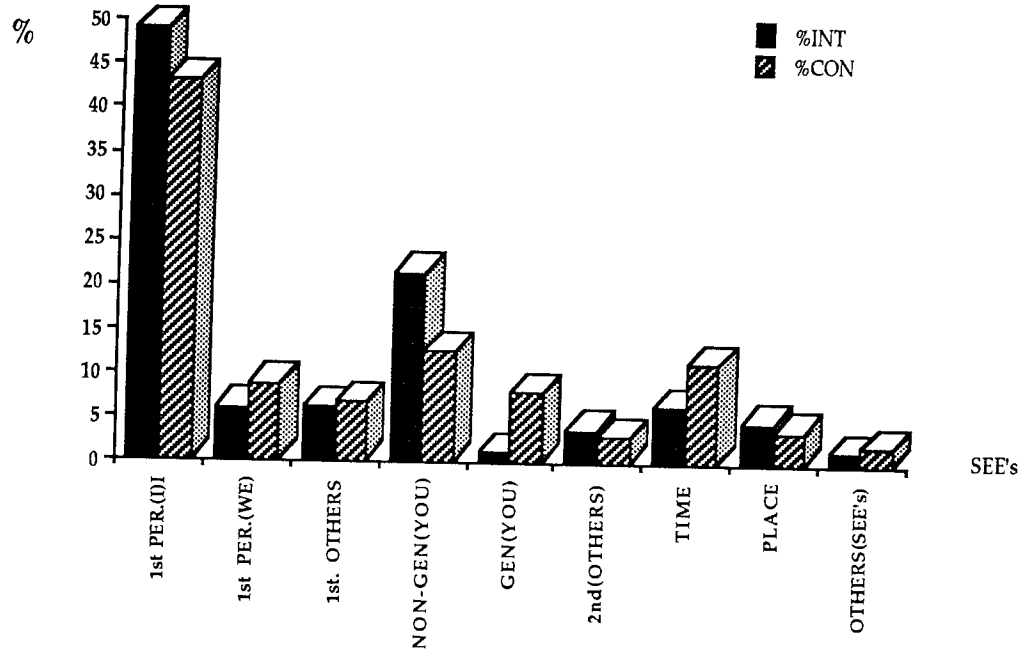


Fig.6.5 Percentage of Situationally Evoked Subcategories in INT and CON

6.6 Information Structure in INT and CON : Summary and general discussion

In this chapter, the status of information in the smaller units of discourse in INT and CON was examined. The structure of information in terms of New, Inferable and Evoked has been statistically investigated with the aim of finding the differences between the two types of SES under examination. The three types of New, Inferable and Evoked entities are seen as the major constituents of IS. They represent the categories of information which speakers generally assume their addressees receive and comprehend during everyday communication.

In this investigation the factor of phonological prominence has been avoided altogether although its importance is realised in any description of information structure within SES. This is of course one of the limitations of this study and indeed one of the problems faced by the present researcher especially in the specification of New and Given information. Although it is realised that the status of information is not always dictated by whether or not an entity has been referred to already within the discourse (cf. Brown and Yule 1983a:157), the categorisation of entities as New or otherwise is almost wholly linguistic depending on relations between entities of discourse (ie *endophoric relations*) and also between entities of discourse and the context of situation including the speakers themselves (ie *exophoric relations*). Accordingly, any entity that cannot be interpretable by the hearer from the co-text (cf. ibid:46) or context of situation or any other source related to the linguistic environment is New and all those that can enter into such relations are Evoked. But this does not mean, however, that Evoked entities cannot be New in the Hallidayan sense. A great number of what is considered to be Inferable or Evoked entities can be New in the sense of Halliday cited above, as they have the necessary conditions for considering them as such. But this is not our concern at the moment. It is only mentioned, however, as it has a lot

to do with the interpretation of the results of the analysis, which will be done in the following few paragraphs.

Two factors have been recognised as important in the comparison of the language of INT and that of CON on the levels of the categories distinguished for the analysis of information structure. The first one is the frequency of these categories and their subcategories in the two types of SES under investigation and the second is their distribution among the units of discourse and within the whole body of the discourse. These factors are taken into consideration in the discussion of the distinction made earlier between the New, Inferable and Evoked entities of discourse, as outlined above. For this purpose, the following figure is presented. This figure gives a clear picture of the size of each category within the AF taxonomy. The values and percentages of this figure are found in Table 6.1 Section 6.5.1 above. The figure acts as a point of departure for the discussion of information structure in SES.

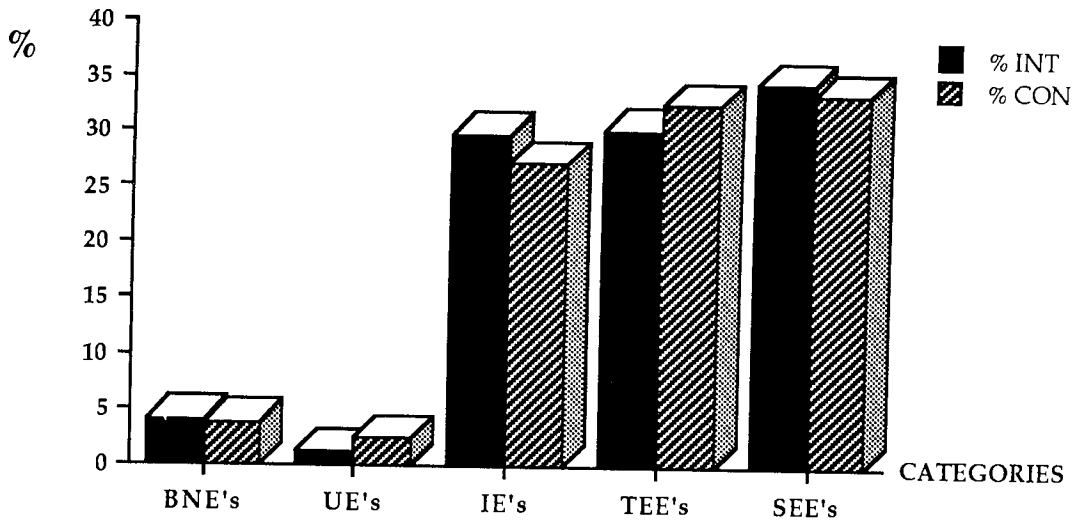


Fig. 6.6 Percentage of 'ASSUMED FAMILIARITY' Categories in INT and CON

The figure above gives the implications required for making the

distinction between the various categories of the taxonomy. First of all, notice the low occurrence of New Entities (BNEs and UEs) in both discourse types in comparison to those that are Inferables or Evoked. This is perhaps due to the centrality of the topics in the INTs and the familiarity of speakers with each other (recalling that each pair represents close friends) and consequently the higher possibility of having possibly common shared knowledge in the CONs. This *'common shared knowledge'* may very likely reduce the possibility of having New elements in their exchange of talk. This also explains the general similarity of the results of the BNEs and UEs in INT and CON. Interestingly, however, it is the distribution of these New elements within the units of discourse that is different between the INT and CON outputs, with a tendency of the speakers to use them earlier in the discourse in INT or within the initial units of new topics in both discourse types. Sometimes, they show randomness of occurrence noticeably more in CON than in INT.

The Evoked Entities of the two types (ie TEEs and SEEs) represent nearly two thirds of the data, and that is not surprising since SES and its overall textual organisation is tightly connected to the contextual and situational roles within which it is conducted. The 'SPEECH ROLES' of the speaker and addressees and the 'OTHER ROLES' that include reference to other parties outside the speech situation (cf. Halliday and Hasan 1976:45), are clear examples of these entities and the relations that hold between them are clearly those that create the best conditions for communication to be achieved. However, in as far as linguistic complexity is concerned, they do not represent a major source of complexity since they are linguistically realised by pronouns and repetitions of NPs as they are used to refer to entities within and outside discourse. There is anyhow no significant difference between the total percentages of their occurrence in INT and CON.

The third of the AF categories, the Inferables, is the most interesting of the three for the purposes here, both for its frequency and distribution in the data. The linguistic expressions and formulae which are used to represent the various subcategories of the Inferables are comparatively variant but highly frequent in the data of INT and CON. However, unlike the other categories of the taxonomy, they give insights to complexity on the linguistic level and complexity on the processing level. Indeed there is some evidence in the results to suggest that there might be a link between the two types of complexities (see below) although this has not been investigated as an aim in the analysis.

From the point of view of linguistic complexity, the language of the INT sessions has been found to have more of the Inferables than that in CON, particularly the Containing Inferables, which are the most linguistically complex of all types of Inferables. They are larger in size and they involve entities within entities, sometimes with many attributes. This finding leads to the conclusion that the language of INT is more linguistically complex than that of CON in the use of linguistic expression representing Inferable entities. This confirms the previous findings concerning the relatively more complex nature of INT over CON in as far as lexical density and information packaging (CHAPTER FOUR) and that of syntactic complexity (CHAPTER FIVE) are concerned.

As suggested above, the results of the Inferable calculations present insights to the presence of complexity on the level of discourse processing as well. The high frequencies obtained in both types of speech suggest that there is a lot of processing load put on the addressee to look for an interpretation for the 'hidden' antecedent referred to in the speech of the addresser. The level of complexity increases if the antecedent referred to cannot be spotted and hence the listener will look for other strategies to achieve comprehension such as 'bridging', 'addition', 'restructuring' (cf. Clark and Haviland 1977:6). During INTs, the results obtained show those

Inferables that present higher processing load on the listeners (e.g. Containing Inferables) occur significantly more in INT than in CON, which suggests that there is more inferencing to be done during INT than during casual CON.

It is worth mentioning here as well that Prince (1981:250) also attributes the fact that the higher complexity of her written data than of the oral narratives to the higher number of Inferables present in the former than in the latter. Put on a scale of linguistic complexity the written text she analysed is highly complex compared to the INTs and CONs analysed in this study on both the linguistic and the processing levels. Yet the INTs show some similar features of complexity, mainly the one related to the packaging of lexical items and information within the units of discourse.

Taken altogether, the results of this chapter point to the conclusion that with certain features there is greater difference between one speaker and another than with other features, as expressed by the SD and V figures of these features in the tables presented above. This seems to suggest that first speakers sometimes show considerable differences in the way they choose linguistic structures to convey information. For example, certain speakers use specific structures frequently which others do not or at least did not use within the same context of situation. Secondly, it confirms the findings of Farag (1986:195) that **"individuals differ considerably in their strategies"** and that each has a style constant which marks his/her own way of speech delivery. This is why more significant results are noticed when the differences between individual speakers are considered than when the total results of all speakers are considered. This is also confirmed by Farag's (ibid) statement that **"grouped results which are treated statistically can <u>obscure</u> variability at the individual level"** (emphasis in original).

6.7 Conclusion

In this chapter, the emphasis has been on finding out through statistical investigation the difference between a) the speaker's output in INT and CON on the level of information structure, constituents and organisation and b) between the speakers themselves. The results suggest that there is some difference between speakers' output in INT and CON in certain aspects of information structures but not in others. However, there seems to be some inter-subject variability in using some of these structures across the 16 subjects. There are also some significant differences between the results yielded by the language of PGs and those yielded by the UGs, which shows that educational maturity can have its influence on the way people express themselves in both types of discourse.

Footnotes on CHAPTER SIX

1. It has to be made clear at this point that although the local/global distinction as proposed by de Beaugrande (1981) is made in relation to the issue of text processing and reading comprehension, it offers along with other helpful distinctions made in the same paper (e.g. **'Processor Contributions, Modularity versus Interaction, Serial versus Parallel Processing, Logical versus Procedural Adequacy** and so on (cf. ibid: 263-7) a number of helpful and illuminating insights into how a model for the analysis of SES is to be constructed and employed. This is why the model adopted here has made a fair balance between most of the distinctions suggested above especially in terms of the following distinctions : **'Bottom-up' versus Top-down, Local versus Global, Modularity versus Interaction, Serial versus Parallel, Openness versus Closedness**. This is not to claim, however, that the Model is exhaustive and constitutes features of perfection but it is just to say in as far as the present analysis and its aims and purposes are concerned, the model seems to be conveniently balanced on the levels and distinctions outlined above.

2. See Halliday and Hasan (1976:38-39) for more details of the three types of referential entities.

3. Although Chafe's (1970, 1974, 1976, 1977) work is normally grouped within linguistic studies, it can more conveniently be grouped under the psycholinguistic approaches as it has a lot to do with psychological issues such as knowledge representation, memory, retrieval of information and so on. As explained in CHAPTER THREE, even his approach of narrative data segmentation into units depends on cognitive rather than linguistic cues (see Section 3.1.3 above).

4. This formula is taken from Prince (1981:245).

5. This example is taken from Prince (1981:233).

6. The endophoric use of the demonstrative pronouns *'this'*, *'that'*, *'these'* and *'those'* is distinct from their exophoric use in which they are used to refer to something within the context of situation (e.g. pick these up), which is normally accompanied by pointing or other body movements. (See Halliday and Hasan 1976:58 for details of this distinction). This latter use is not included in the calculation of this section, as it is not 'textually evoked' or linked to another entity within the text.

7. Notice, however, that Levinson (1983) makes a distinction between discourse deixis and anaphora claiming that they are not mutually exclusive. He states that : **".. in principle the distinction is clear : where a pronoun refers to a linguistic expression (or chunk of discourse) itself, it is discourse-deictic; where a pronoun refers to the same entity as a prior linguistic expression refers to, it is anaphoric" (p.85).**

8. See especially INT II and CON II, in which the speakers are the President and the Vice-President of the Guild of Students at the University.

CHAPTER SEVEN

SUMMARY, CONCLUSIONS AND SUGGESTIONS FOR FUTURE RESEARCH

7.0 Introduction

Two types of talk among young university students that involve the direct exchange of turns have been quantitatively examined in the course of the present thesis with the focus being fully placed on the internal structure of the sequential units of discourse. The first type of talk is that of students being interviewed formally by an interviewer, a chaplain at Aston University. Students were asked questions related to their life experiences, social, religious and educational relations and so on. The second type of talk consists of the informal conversation exchanges that followed the interview encounter, between the pair of friends. This represents the kind of casual, spontaneous, and informal chat that normally goes on between friends. During the first encounters (i.e. INT), the interviewer acted as an 'elicitor of insights'. However, his speech was excluded from the calculations, which gave us a balanced size of data with that of the CON encounters during most of which he was not physically present. So the calculations represent, in fact, the recurrence of features within students' talk only except where otherwise stated.

While the language used in the two types of speech is similar in many respects due to the fact that it has been produced by the same speakers (albeit under two different conditions) there seem to be some differences that characterise the two types of data as two distinct styles of speech. These differences are of two sorts. The first one is related to the overall organisation of discourse in terms of turn size, distribution and content. This issue has not been of concern in this thesis and hence has not been given due consideration. However, a few points that have a relationship

to the second of these issues are worth mentioning (see discussion below). These points have been arrived at by a close observation of the data of INT and CON and by the insights given from their segmentation into smaller units for linguistic analysis (see CHAPTER THREE above). The second set of differences is related to the structure of these turns and the smaller units they consist of in running speech. This structure is manifested in the use of lexis and syntax for the expression of information in communication settings. As such, these differences have been observed to occur on three broad dimensions according to their sources in the structure of units of discourse : Lexis, Syntax and Information Structure. These dimensions constitute the substance for the investigation of linguistic complexity in INT and CON. They have been covered by the statistical method used in the analysis in CHAPTERS FOUR, FIVE and SIX respectively. A summary of the findings will be listed below, with a discussion of theoretical and applied implications. This will be followed by some suggestions for future work based on the conclusions arrived at in this thesis.

7.1 Organisational features of INT and CON

The most noticeable differences in the overall organisation of INT and CON are observed in the turn-taking size, structure and content. As might well be expected, turns in INT are longer than those in CON. They contain longer units and are less elliptical than those in CON. This might partly be due to the lower use of interruptive and self-interruptive talk in INT. Generally, they display specific patterns of occurrence which are determined by their positions within the INT. For example, at the beginning of each INT, students normally respond to the first question in a short unit followed directly by other longer and more complicated ones. Since topics are strictly controlled by the interviewer, it seems that there is less freedom for the interviewees to shift topics and introduce new ones. This is normally done by the interviewer. There are, of course, a few cases

where the interviewees shift topics, but these only occurred after a period of silence when the interviewees had enough time to realise that the interviewer had no further questions to raise. Altogether, turns at INT are more organised, with clearer patterns than those in CON. Other features of conversational discourse organisation such as turn-allocation techniques, closings, openings, sequencing, repair etc. (cf. Sacks et al. 1974) are also influenced by this patterning of turns and are more clearly established and adhered to both by the interviewer and interviewees. But these are not going to be elaborated on here.

As far as the CON talk is concerned, there seems to be a greater variability and inconsistency in the length and organisation of turns. It seems rather difficult to have a prespecified pattern of unit organisation within the turn. Frequently in CON, turns can be very short consisting of just words or phrases. Adjacency pairs (cf Sacks et al. 1974) exhibit continuous dialogic structure with frequent interruptions and simultaneous talk. At the other end, some conversants employ strategies of narrative talk thus using longer turns with multiple and continuous units. Such turns are not infrequent in the CON recordings analysed in this study (see especially **CON II, IV and VII**).

There is significantly more interruption and simultaneous speech in CON than in INT. It is noticeable that interruptive talk increases with the increase of the informality between the conversants. So this phenomenon occurs more when the pair of students (friends) direct their talk to each other but considerably less so when they talk to the interviewer.

7.2 Structural features of INT and CON

Following is a summary of the important results of the analysis undertaken in the thesis concerning the structural features of the spoken output in the data of INT and CON.

(1) The results of the lexical analysis (CHAPTER FOUR) in as far as lexical density (LD) in INT and CON is concerned show that, first of all, both types of talk have relatively higher LDs than those reported by other researchers. The results reported in the chapter show that LD falls wihtin a range of 45% and 55% (see Table 4.6 above). These relatively high LDs in the data in general could well be attributed to the level of education of the students, which may also explain the higher LDs given by the postgraduate pairs of students. The results seem to confirm Stubbs's (1986) speculatory remarks that the reason why his informants have produced higher LDs than those of Ure (1971) is that they are all highly educated academics (Personal Communication), whose speech is **"heavily affected by written language"** (1986:41).

(2) The results show clearly that the language of INT and CON is structurally complex on the level of lexis, unlike what some other researchers seem to imply (see, for example, Brown and Yule 1983a; Halliday 1979, 1985b). As previously mentioned Halliday takes it for granted that spoken discourse types in general and informal conversational discourse in particular is lexically sparse (in comparison to written discourse), **"... the lexical density of written language is likely to be of the order of twice as high as that for speech; and the discrepancy will be greater if other factors such as the relative probability of lexical items are taken into account"** (1985b:80). The results above do not confirm this hypothesis. Although sparsity on the lexical level may be found in speech especially in the extreme informal types of conversational discourse, there is evidence for the conclusion that there is complexity on the level of LD.

This complexity is shown by the relatively high frequencies of LDs in both INT and CON, on the one hand, and by the variability in LD between speakers' speech in INT and CON, on the other (see discussion below). It can also be manifested in the variability of the results of LD in the lower units of speakers' speech as will be explained below.

(3) The results of the functional analysis of the smaller units of speech in terms of their LD and complexity have shown that LD varies in accordance with the functions of these units within running speech i.e. according to the specific functions of these units as informative, narrative, expositive, descriptive, and so on. There seems to be a hierarchical organisation of these units in terms of LD (see Section 4.4 above) when compared in INT and CON. Highly informative units (e.g. General and Personal Informative Units and Expository Units) are those that contain higher LDs, hence higher on the scale than other less informative units (e.g. Elicitive Units). Such units tend to occur fairly frequently in the data analysed here confirming the global results of the analysis summarised in (1) and (2) above.

(4) The other interesting finding as far as LD in types of SES is concerned is related to the difference between INT and CON in this respect. Generally speaking the results show the language of INT to be on the whole more lexically dense than that of CON though in certain cases the difference was not highly statistically significant (e.g. when comparing the MEAN figures of each pair together) (see Table 4.3 and 4.4 above). However, when the speech of each subject is compared in INT and CON (see Table 4.6 above), the results show statistical significance. This finding seems to suggest that formal speech types have higher LDs than informal types (e.g. casual conversation). Such an implication is also suggested by Halliday (1979, 1985b) in his investigation of the differences between writing and speech (see also Stubbs's comments cited in (1) above).

(5) Lexical density in the linguistic output of speakers is directly related to the packaging of information within the units of speech. The more LD a text has the more informationally loaded it is (cf. Halliday 1985b). The results of LD summarised above show that interview speech exchanges and conversational discourse are informationally loaded due to the high LDs found in both types of speech. Since the language of INT seems to have higher LDs than that of CON, it is likely that it is more informative than that of CON. This is expected since the purpose of INT (and indeed many other types of interviews) is to elicit information from the interviewees. This conclusion seems to be confirmed not only by the global results of LD in INT and CON but also by the analysis of the smaller units of both types of speech, as will be explained below.

(6) Differences between one speaker and another concerning the use of lexis have also been depicted. Higher LD results have been noticed in the speech of certain students rather than others, which points to the conclusion that LD can be style specific but this was not consistent with all speakers so that one could make valid conclusions. It seems more plausible, however, to attribute these differences to the level of maturity these students have over others, especially in specific fields (e.g. educational, religious, political), as will be explained in Section 7.3 below.

(7) On the syntactic level, the overall results show the language of both types of SES to be syntactically complex. This is shown by the frequent use and distribution of clause complexes containing embedding, subordination and coordinated subordinate structures. Most of these features have been taken in this research to be features of syntactic complexity in SES types. This finding supports the general line adopted by Halliday (op.cit.) about what he calls **"the intricacy"** of conversational discourse in comparison to written discourse types (see however discussion in (12) below).

(8) Elliptical Units (EUs) have been observed to occur frequently in the data compared to the other types of units. Five types of these EUs have been distinguished in the analysis according to their functions as *short responses, appended units, interrupted or self-interrupted units and others*. It is noticed that the great frequency of EUs might have an influence of simplifying the SC of the spoken output. Short responses, for example, replace longer and more syntactically (and/or possibly lexically) complicated units (see Section 5.5 above).

(9) Other features which have no direct relevance to the syntactic structure per se but have a great role to play as far the SC is concerned (akin to that of the EUs described in (2) above) include *comment clauses, discourse markers and reaction signals*. These are also known to be used frequently in spoken discourse types and this has been confirmed by the analysis in CHAPTER FIVE above. The effect they have at simplifying the complexity of the syntactic units is apparent in the way they are used to break the continuity of these units. Some of the comment clauses and discourse markers are used as parenthetical expressions within the syntactic units (see 11 below).

(10) Comparing the syntax of INT with that of CON shows a different picture altogether. There seems to be some inconsistency in the results obtained. While certain features of English syntax (e.g. Nonfinite Nominal Clauses, 'that' Nominal Clauses, 'Extrapositional' Nominal Clauses, Relative Clauses, (both Restrictive and Non-Restrictive), Appositive Clauses of various types, Adverbial Clauses) occur more in INT than in CON with a noticeably high statistical level of significance, other features, though occurring slightly more in INT than in CON as well, do not show statistical significance in the difference. Some of these features (e.g. Subordinated clauses) are taken to be indices of the syntactic complexity of discourse. This implies that the spoken output in INT is more syntactically complex than that of CON on the level of features mentioned above.

(11) The remaining features of English syntax that have been investigated within the chapter occur less in INT than in CON (see (2) above). Those that occurred significantly more in CON are some coordinatives, Comment Clauses, Elliptical Clauses, and units consisting in the main of reaction signals (e.g. mmhmm, yes, yeah, oh yeah, oh no, etc.) (cf. Quirk et al 1985) and discourse markers (e.g. *oh, well, OK,* , etc.) (cf. Schiffrin 1987). The presence of these features is characteristic of the language of spoken discourse in general and SES in particular. Their frequent use in the conversational data confirms this to a great extent. It also shows conversational discourse to be fragmentary and its complexity, on the whole, to be less systematic, compared to that of the INT data.

(12) The line of the results above seems to contradict the part of Halliday's (op.cit.) comments supported by the findings of Beaman's (1984) analysis that the more informal the conversational discourse is the more syntactically complex and 'intricate' it becomes. The results above show the formal set of data (i.e. INT), according to the indices mentioned above to be more syntactically complex than the informal set (i.e. CON). This seems to suggest that the use of *'complexity'* in this thesis might not be in fact identical with Halliday's use of *'intricacy'* in relation to syntax though they might be similar in certain respects. The indices of syntactic complexity employed in this research are not exactly the same as Halliday's. However, they are similar to those of Beaman who (1984) draws heavily on Halliday's points of view concerning the 'grammatical intricacy' of spoken discourse. The implication is that it might not be correct to compare results of comparisons made between one type of spoken discourse and another with results obtained from spoken and written discourse differentiation. It can also suggest that the types of conversational discourse investigated by Halliday (op.cit.) and Beaman (op.cit.) cannot be generalised to represent all types of SES. I can conclude that it is actually more useful and practical to compare discourse types within one medium (i.e. either spoken or written) than to compare

spoken with written discourse types.

(13) Many interesting conclusions can be drawn from the results and findings of CHAPTER SIX above concerning the complexity of spoken discourse on the level of information structure and organisation. Some of these results confirm those of CHAPTERS FOUR and FIVE respectively. Prince's taxonomy of information status with some necessary modifications has been applied to the data. The taxonomy consists of three main categories which all types of discourse (spoken or written) can be thought to contain. These are NEW, INFERABLE and EVOKED. The general results show that NEW entities have the lowest frequency of occurrence and the EVOKED entities the highest, with the INFERABLE types falling in between. The low occurrence of NEW entities in the data has been referred to the 'centrality' of the topics discussed and the familiarity of most of the students with the interviewer (the Chaplain) and also with each other as each pair was chosen because they were friends. Furthermore, some of the entities that (in other approaches to information structure e.g. Halliday's) might be treated as New because **"they were not previously in the discourse model"** (Prince 1981:252) have been included under the Inferable type following Prince (ibid) since they can be inferred from the situation, context or co-text of discourse (ibid).

The greater use of EVOKED and INFERABLE entities is highly suggestive and needs particular emphasis in this summary.

(14) The considerable degree of occurrence of the INFERABLE category in the data presents a different interpretation of complexity from the one adopted in this thesis. In addition to the conclusion concerning the presence of linguistic complexity due to the high frequency of certain subcategories of the INFERABLES (e.g. COMPARATIVE and CONTAINING INFERABLES in particular, which are mostly represented by compound and complex phrases) (cf. Prince 1981), there is enough

evidence to suggest that there is processing load (caused by the high frequency of the INFERABLES mentioned above) on the part of the listener in interpreting the speaker's messages in conversational discourse. However, since the issue of discourse processing was not the main object of the analysis in the first place in this thesis and since the amount of data investigated is relatively limited in size and type, further analyses dealing with such issues are needed to confirm these claims.

(15) The category of EVOKED entities, both TEXTUALLY and SITUATIONALLY EVOKED, has the highest frequency of occurrence in the data, which is an important finding in itself as it obviously points to the 'explicitness' (cf. Farag 1986:167) of speech in both discourse types. The frequent use of the former type (TEEs) in the data is indicative of the presence of cohesion on the textual level (cf. Halliday and Hasan 1976; de Beaugrande and Dressler 1981). This cohesion is the result of the 'ENDOPHORIC RELATIONS' that hold between the co-referential entities within units of information or what Halliday and Hasan (op.cit.) refer to as 'OTHER ROLES'. The frequent use of the latter type (i.e. SEEs), on the other hand, is indicative of the influence of situational factors (including the participants) in creating the discourse and of the social relationships (cf. de Beaugrande and Dressler 1981:167) that hold between speakers and listeners. In Halliday and Hasan's (op.cit.) terminology, this is done through the use of 'SPEECH ROLES' and 'EXOPHORIC RELATIONS', although they confirm that these entities, unlike the TEEs discussed above, have no cohesive role.

(16) From the point of view of discourse processing on the part of the listener, this high frequency of occurrence is also suggestive in that these various relations have a light processing load which might help the listener in the perception of characters and events in the discourse (cf. Al-Jabr 1987). This stands in opposition to the high frequency of the INFERABLES which tend to impose higher processing loads on the

listeners as they try to interpret speakers' messages by 'inferencing' them from their intended sources or else by other processes such as 'BRIDGING' (cf. Clark 1975), which adds further processing load.

(17) As far as the differences between INT and CON in the employment of the various categories (and their subcategories) mentioned above, the findings generally suggest that there is no significant difference between the two types of speech in the employment of NEW entities. However, the differences appear in the distribution of these NEW entities within the units and topics of discourse. Generally, they have been noticed to occur more within topic-initiation units than otherwise. The differences are also noticed when considering the two types of NEW entities distinguished : BNEs and UEs. The results indicate that while there is no significant difference in the use of BNEs in INT and CON, there is a significant difference in the use of UEs. Speakers in CON used more UEs than in INT. Again, this can be interpreted in relation to the familiarity of the students with each other and their common understanding of the topics discussed and their shared knowledge of the world.

(18) The category of INFERABLES is generally used by speakers significantly more frequently in INT than in CON. There are, however, some differences between the results of the individual subcategories distinguished within the INFERABLE category. Those subcategories that occur more in INT than in CON with a high statistical significance are COMPARATIVE INFERABLES, ADJACENT INFERABLES, and more importantly the CONTAINING INFERABLES. The COMPARATIVE and the CONTAINING types occur nearly twice as much in INT as in CON which is a highly significant result and points to the presence of linguistic complexity as these two types of entity are represented linguistically by complex NPs of the type **"the biggest nuisance of all"** (COMPARATIVE) and **"a number of.. female executive officers"** (CONTAINING).

On the processing level, the above results show the language of INT to be more complex since there is more processing load for the listener. However, this cannot be but a tentative conclusion that need to be verified by other more specialised analyses of information processing methods.

(19) The remaining types of INFERABLE subcategories occur slightly more in the CON encounters but generally the difference is not statistically significant.

(20) In the category of EVOKED entities, there is a slight difference between the two types TEXTUALLY EVOKED and SITUATIONALLY EVOKED in terms of the differences between INT and CON. While the former (TEEs) have been noticed to occur less in the INT encounters, the latter (SEEs) occur more in INT than in CON. The total MEANS of the results do not show a highly statistically significant difference. However, the differences seem to be more significant on the level of the particular subcategories of each type, as will be seen below.

(21) With the exception of the subcategory '3rd Person', which occurs significantly more in the CON data, all other subcategories of TEEs occur more in INT than in CON, some of which have a very high statistical significance. These include the subcategories : 'Extrapositional 'it', 'Repeated Entities', 'Relative and Zero Relative Entities'. The high occurrence of these subcategories is in itself indicative of the presence of linguistic complexity. And their higher use in INT confirms the conclusion made earlier that there is relatively more linguistic complexity in the INT data than in the CON data. Take, for example, the use of 'Relatives' and 'Zero Relatives' in the data. The results show that these have a general OPTWs of 12.8 in INT and 6.6 in CON. This result is due to the more elaborative nature of INT and the tendency of speakers to add more information and qualification to previously mentioned entities. The highly frequent use of Textual Repetition in the data is also significant as it

has a function in establishing the topics being discussed and also helps in creating stability and economy of the text as it entails **"sameness of reference"** (de Beaugrande 1980:135).

(22) The highly significant use of 3rd Person Pronouns in CON more than in INT has been interpreted in terms of the interactional nature of conversational discourse and the tendency of speakers to involve persons other than those present in the speech situation in the talk. In INT however, there seems to be more involvement between the interactants themselves (i.e. between the interviewer and interviewees).

(23) Turning now to the differences in the use of SEEs, although it can be seen that these are the most frequent in the data of all the AF categories, they generally show no statistical difference between speaker's use in INT and CON. One significant difference, however, has been noticed, and that is in the use of Generic and Non-Generic *'you'*. Generic *'you'* is used frequently in INT whereas the opposite is true of the Non-Generic *'you'*. According to the type of Generic meaning it has (cf. Quirk et al. 1985:354), it expresses the speaker's appealing to his/her own and (sometimes) the hearer's experiences of life in general. Indeed, I can refer its greater use in INT to the fact that the interviewees were mainly discussing views about their own personal life at home, at the university or at their social, political or religious places of activities. These topics were far less discussed during the CON encounters.

(24) A final point that can be made in this summary concerns the observations made on the variation of linguistic features discussed above on the inter-individual scale. This has been commented on in relation to 1) the standard deviation and coefficient of variation figures which show the range of possible dispersion and also the differences in the employment of linguistic features and 2) Postgraduate/Undergraduate differences in the use of these features. Generally, the results show some

variation in use between one speaker and another with certain features rather than others. Greater variability has been obtained with features having low occurrences in the data. Other features have been noticed to occur equally well in both INT and CON for certain speakers. These are taken to be style-specific of the speech of those speakers.

(25) Generally speaking, the results obtained from postgraduate students showed comparatively more consistency in linguistic complexity on the three levels of structure examined in the thesis. They also showed more statistical significance in the results they presented. This has been referred to their level of educational maturity which corresponds more than that of the undergraduates to the level of the interviewer.

7.3 Conclusions

Two types of conclusions can be drawn from the results of the present thesis summarised above : general and specific. The general conclusions centre around the outcome of the analysis which is : the linguistic output in both types of speech analysed in this study displays a considerable amount of linguistic complexity on all levels of analysis examined (i.e. lexis, syntax and information structure). Both types have been observed to have lexical densities which establish them as two 'informationally loaded' types of spoken discourse. Both types show complexity on the syntactic level in that they have a great number of intricate structures, involving subordination and complex clause and clause complex structures. The frequent use of such features as ellipsis, comment clauses, discourse markers and reaction signals can be significant in relation to linguistic complexity. Most of these features are used by speakers to interrupt the flow of speech simplifying as such what would otherwise be long and complicated consecutive units containing multi clausal combinations.

The analysis of the information structure of the language of INT and CON confirms in part the linguistic complexity obtained by the lexical and syntactic analyses. Additionally it shows another type of complexity which is found to be prevalent in both INT and CON and that is the one on the information processing level. The types of information delivered in speakers' output in both speech styles contain a comparatively great number of inferable entities that impose some processing load on the listeners. Although comprehension is facilitated by the use of other entities such as the Evoked types (i.e. SEEs and TEEs) and by other linguistic and extralinguistic features which are not our concern here, the listener is still expected to infer via various processing strategies the referents of these entities.

The presence of linguistic complexity in the data and the type of cognitive complexity on the processing level suggests another general (but tentative) conclusion that their might be a link between the two types of complexity. In other words, the high lexical densities in the data and the relatively complex syntactic structure of speech in addition to that of information structure outlined above, can have its effect on the listeners' processing abilities while communication is taking place. But for this conclusion to be valid it needs to be tested against a bigger corpus of conversational data and other types of SES. Therefore, it remains a tentative conclusion.

The above 'general' conclusions can be generalised to include other types of interview and conversation and also other types of SES. This view can be supported by the fact that some other analysts who have done research on the structure of conversational discourse, have reported findings of linguistic complexity on similar lines of analysis (see especially Poole and Field 1976; Halliday 1979, 1985b; Crystal 1980; Beaman 1984).

The specific conclusions that can be drawn from the various results of the analyses conducted above are related to the intra and inter-individual differences in INT and CON. Although in certain linguistic and discoursal features speakers did not show significant differences in their linguistic output in INT and CON, there seems to be a considerable amount of variation with other features, some of which have shown the difference to be highly statistically significant. This conclusion confirms the claims put forward by various researchers (see especially Crystal and Davy 1969) that speakers use different speech styles in different speech situations depending on the purpose of the conversational encounter, the co-participants, the topics discussed and the formality of the speech situation (see discussion below). The differences are depicted on the two levels of linguistic structure, lexis and syntax and also on the level of information structure and organisation within the discourse and the way the information structure is represented in linguistic forms. Features from all these three levels were found to vary between the two styles of speech, with those that are taken to be indices of linguistic complexity to occur more within the INT style.

The questions then that need to be considered now are : why do speakers produce these differences in the first place ? Why is it that their linguistic output is influenced by the relative change of situation ? What are the basic factors that cause these differences ? The answers to these questions are not yet clear and to the best of my knowledge still not widely explored. If we knew the answers to these questions many of the dangling issues related to language variation and the way people use their linguistic forms in various everyday conversational encounters would be resolved; for example, the issue of linguistic complexity in certain types of talk rather than others, its source, type, and interpretation; also the relationship between complexity on the linguistic level and complexity on the level of discourse processing and interpretation. Certain speculatory remarks, however, can be given looking at the data under analysis and at

the results of the analysis per se. The first point concerns the participants themselves, i.e. the presence of the interviewer in the INT situation as regards his absence in the CON situation. Obviously, his presence adds to the formality of the situation and imposes certain constraints on the speakers. Some speakers have been found to adjust their way of speaking to match some of the perceived characteristics of the interview, the presence of the interviewer being by far the most salient one (see CHAPTER FOUR above). Differences have been noticed in the manner in which participants talk to the interviewer as opposed to when they talk with each other as friends. This has been explained in relation to the "ACCOMMODATION THEORY" (see Giles and Clair 1979; Giles 1980), which states that speakers normally adjust their speech styles as a means of displaying attitudes and intentions towards other people. Other constraints might be caused by the speakers' feeling of tension during the INT sessions and their awareness of the fact that their speech is being recorded and monitored. This is related to self-consciousness in INT as regards its absence in the casual CON encounters. Finally the interviewer's control of topic during INTs and freedom of choosing any topic for discussion in CONs can be an important factor influencing the differences in the linguistic output of speakers in the two types of speech encounters. Related to this is the 'seriousness' of these topics when discussed with the interviewer as regards the 'shallowness' of chat often present during casual conversation encounters.

Personal attributes and group attributes can have their role as well in influencing speakers' use of certain features (e.g. LD). This is more clearly observed on the inter-individual level than on the intra-individual one. These factors include maturity, educational level, confidence, age, sex and social class. This is confirmed by the results obtained from comparing, for example, the postgraduate with the undergraduate output in both discourse modes.

Up to now, I have been talking about the general characteristics of INT and CON in comparison to each other and also about the possible differences that might be found between them in terms of their internal linguistic and informational structures. However, some of the features investigated showed more similarities than differences in the two styles of speech. This was also expected since the two types of speech were exchanged by the same speakers except that in the INT encounters the presence of the interviewer and other features of the interview situations imposed certain constraints on the participants as shown above. Crystal and Davy (1969:3) talk about this phenomenon in considerable detail stating that **"Naturally, these varieties have more in common than differentiates them".**

7.4 Some theoretical and applied implications

The results of the analysis conducted above, especially those that are related to the complexity of language of INT and CON, have many practical implications both for theoretical and applied linguistic research. The most important implication on the theoretical level concerns the description and analysis of speech types for the same speakers under different speech situations, i.e. stylistic variations (cf. Coates 1986). The lexical and syntactic structures of speech types can be intricate unlike what some researchers seem to think about them. The results of the preceding chapters have shown the speech of speakers in different speech situations to be complex in terms of lexical and syntactic employment and in the representation of information within the linguistic system. There seems to be an interrelationship between the three levels and this needs to be investigated on different speech types.

The second implication that has been yielded by the above analyses is that the tendency of researchers to describe the complexity of spoken discourse types in relation to written discourse can be misleading and

unfruitful for many reasons most, important of which is that there is no written counterpart to spoken interactional discourse and in most cases the two types are far from being comparable. It is the view of the present researcher that comparing spoken discourse types with each other especially those produced by the same speakers under different situations can really reveal most if not all of the intricacies involved in the spoken output. This is why this research can be taken to be providing proposals and insights towards the description of spoken discourse types. These proposals are not new in the sense that they are known and usually touched upon in the linguistic tradition through writings of such distinguished authors as Quirk et al's comprehensive analysis of English grammar (1972, 1985), Crystal's (1980) and Crystal and Davy's (1969, 1975) analysis of varieties of spoken discourse, Halliday's (1967, 1979, 1985a, 1985b) textual analysis of English grammar. All these linguists and many others besides (see especially the contributions in honour of Randolph Quirk edited by Greenbaum, Leech and Svartvik (1980) lay emphasis on the linguistic description of spoken discourse as a major field of research within the linguistic disciplines. However, the insights suggested by the present analysis shed light on many hidden issues in the complexity of such spoken discourse types as SES.

The implications above lead to a more general one and that is concerned with the need for an overall linguistic theory that takes into consideration spoken discourse types, especially the types that have been referred to here as "SPEECH EXCHANGE SYSTEMS". This is greatly required considering the current advances made in the field of discourse analysis and sociolinguistic research in such areas. As outlined above most of the attempts found in the linguistic literature and in the more recent surge in the study of language variation written discourse into consideration, or in the best cases, touch upon spoken discourse just in passing or in relation to written discourse. This can be helpful in pinpointing some general features of speech but it certainly does not

reflect the actual nature of speech as exchanged by people in different or similar situations. It might give hints to the presence or absence of certain features, but it does not provide an overall picture of the linguistic characteristics of speech nor can it give any explanations for the various processes involved in the exchange of information between speakers.

The implications on the applied level are numerous but only those that are related to the teaching of conversation are worth mentioning here. The various features of SES as yielded by general observation of data or by the actual analyses and findings give insights relevant to the teaching of English conversation to speakers of other languages. Programs and syllabi of teaching English conversational discourse based on artificial materials are clearly inadequate in providing learners with the real characteristics of speech as used by native speakers of English. Comparisons made between naturally occurring informal conversations and more informal types of SES yield many insights as to the true character of conversational discourse and the features that need to be emphasised in the teaching of spontaneous speech. This confirms the implications arrived at by Barkho and Zora (1984) in comparing naturally occurring conversation with artificially prepared extracts for the teaching of conversation. The main conclusion is that there is a need for a more linguistically and culturally based approach to the analysis and teaching of conversation and its main characteristic features. Very little has been done in this area in current applied linguistic research (see however, Brown 1977; Crystal 1980; Brown and Yule 1983b; Barkho and Zora 1984).

7.5 Limitations of the study

The most clearly observed limitation of the study at hand is, of course, the limited amount of data examined for conducting the analysis. Clearly statistical analyses of the type presented in this study call for a greater population to be recorded and analysed. This has been suggested by the standard deviation and coefficient of variation calculations made for each of the features examined. However, this factor is greatly linked to the amount of time available for researchers and also their resources. This factor has been carefully considered when choosing the 16 subjects for conducting the interviews and conversations and the interviewer (see CHAPTER TWO above). However, to make generalisations concerning inter-individual variation (e.g. postgraduate/undergraduate variation, male/female variation and so on requires a greater and more balanced number of participants to be monitored.

447.6 Suggestion for future research

The linguistic study of INT and CON in the present research has revolved around three issues of contention in linguistic research : Lexis, Syntax and Information Structure. The in-depth analysis conducted revealed that the structure of spoken discourse of the types exchanged in communicational encounters can be a vital issue for linguistic research. The findings confirm the need for a comprehensive theory (cf. Farag 1986:200) of English conversational discourse that takes as its primary objective spoken discourse structure and its complexity on the various levels suggested above.

Whereas the investigation of the structure of English SES in this research centred on the three main areas of lexis, syntax and information structure, other areas that might enrich the theory and make it more applicable to spoken discourse types can be incorporated into the analysis.

On top of these areas is the incorporation of prosodic analyses, especially so within the realisation and characterisation of information structure. This should be added not so much as an independent area in its own right (important though such an area might be), nor to the mere distinction between Given and New issues as suggested by Halliday and others, but also to develop an approach to the segmentation of English SES that can take into consideration syntax and semantics in addition to prosody within a functional framework. Such an approach would definitely reveal a lot about the real structure and nature of spoken discourse variation.

Another line of research that the analyses of INT and CON bring into attention for future consideration is the extension of these analyses to include other types of SES controlled on such levels as spontaneity, plannedness, formality, participants, topics etc. The study of language variation along these lines might prove fruitful and insightful both from the theoretical and the more applied perspectives.

The factor of participants in any type of speech situation is the most crucial one. Future research should take into consideration 1) the type of participants whose speech is to be monitored and studied, 2) their age, sex and educational and social backgrounds and 3) the number of informants that are needed for the investigation. The third issue is a very important one and in statistical analyses of the type presented above, there is a need to include as many informants as possible so as to yield valid generalisations. Certainly, the more informants one is able to get the more valid the results and the generalisations will be. With the advances made in computational techniques of research big corpora of data can be analysed and investigated in record times.

The study of linguistic complexity of spoken discourse conducted in the chapters above can also illuminate many interesting issues for a linguistic theory of discourse structure. First of all, variability of linguistic

complexity across spoken discourse types and across individuals might be a helpful area within sociolinguistic research. The influence of linguistic complexity on the currently popular issues of discourse and text processing is interesting for psycholinguists and cognitive psychologists alike. Eventually, such interests might prove useful as well in the attempts of scholars of the latter trends (i.e. cognitivists) to develop the research being carried out within the field of Artificial Intelligence.

One of the most interesting topics that could be pursued in relation to the analyses and findings of the present research is that of the INFERABLE entities within discourse structure. This phenomenon can be studied not just within the linguistic theory proposed above but also in relation to discourse functions and discourse processing. The structure of such entities and its relationship to other entities within the unit of information might reveal a lot about the complexity of discourse on the level of linguistic structure or in relation to language processing and comprehension.

There is certainly a need to exploit the findings and implications of research work conducted on the lines discussed above for pedagogical applications. One of the most important issues (at least within the present researcher's future interests) is the teaching of English conversation according to the insights yielded by the analyses above. This is especially helpful for advanced learners of English, those specialising in English as a career for professional and academic purposes. Such a commendation involves the preparation of teaching materials based on authentic data produced by native speakers with the features related to structure, spontaneity and naturalness focused on and elaborated. It also involves the preparation of relevant teaching methods and techniques based principally on the natural, and unartificial type of data collected for this study. In the opinion of the present researcher, this would be one of the great achievements of the collection and analysis of the data of INT and

CON carried out in this project.

Finally, I can end this discussion by concluding that the language of speech exchange systems, though seemingly structurally untidy and complex, is the most interesting field for conducting linguistic (and other) types of research. Many features and characteristics of SES are still unexplored. There is too much in the language of SES, especially that of natural conversational discourse that merits thorough and incessant investigation. Schiffrin (1988) puts it aptly in simple, but straightforward words, that :

> "Some of the very same qualities that make conversation an important topic for linguistic attention also make conversation a difficult topic for linguistic analysis".

REFERENCES

Aarts, J. and W. Meijs (ed.) (1984) *Corpus Linguistics* . Holland : Rodopi.

Abelson, R.P. (1976) "Script processing in attitude and decision-making", in J. S. Carroll and J. W. Payne (eds.) *Cognition and Social Behaviour.* Hillsdale : Lawrence, Erlbaum, pp. 33-46.

Adelman, C. (ed.) (1981) *Uttering, Muttering* . London : Grant McIntyre Ltd.

Ager, D. E. (ed.) *Written Skills in the Modern Languages degree.* AMLC/CILT, Department of Modern Languages : Aston University.

Aijmer, K. (1979) "The function of tag questions in English" in T. Pettersson (ed.) *Papers from the Fifth Scandinavian Conference of Linguistics* . Frostavallan, Stockholm : Almqvist & Wiksell International, pp. 10-17.

Aijmer, K. (1985) "Just" in S. Backman and G. Kjellmer (eds.) *Papers on Language and Literature.* Sweden : Acta- Universitatis, pp. 1-10.

Aitkenhead, A. M. and J. M. Slack (eds.) (1985) *Issues in Cognitive Modelling,* London : Lawrence Erlbaum, The Open University.

Akinnaso, N. (1981) "The consequences of literacy in pragmatic and theoretical perspectives", Anthropology and Education Quarterly, 12, 163-200.

Akinnaso, N. (1982) "On the difference between spoken and written language", Language and Speech, Vol. 25, 2, 97-125.

Al-Jabr, A. M. (1987) Cohesion in Text Differentiation : A Study of English and Arabic. Unpublished Ph.D. Thesis, Birmingham : Aston University,

Al-Jubouri, A. J. R. (1987) Computer-Aided Categorisation and Quantification of Connectives in English and Arabic (based on newspaper text corpora) . Unpublished Ph.D. Thesis, Birmingham : Aston University.

Al-Shabab, O. S. (1986) Organisational and Textual Structuring of Radio News Discourse in English and Arabic, Unpublished Ph.D. thesis, Birmingham : Aston University.

Allen, D. E. and R. F. Guy (1974) *Conversational Analysis : The Sociology of Talk.* The Hague : Mouton.

Allwood, J. (1985) "On relevance in spoken interaction" in S. Backman and G. Kjellmer (eds.) *Papers on Language and Literature*. Sweden, Acta Universitatis, pp. 18-35.

Anderson, P.V., R. Spiro and W. Montague (eds.) (1977) *Schooling and the Acquisition of Knowledge* . Hillsdale, New Jersey : Lawrence Erlbaum.

Assad, M. S. (1987) The Linguistic Analysis of Dramatic Dialogue Versus Conversation : A Comparative Study of Pinter and Aychbourn. Unpublished MSc. Dissertation, Birmingham : Aston University.

Austin, J. L. (1962) *How To Do Things With Words*. Oxford : Clarendon Press.

Backman, S. and G. Kjellmer (eds.) (1985) *Papers on Language and Literature*. Sweden : Acta Universitatis.

Baka, F. (1989) The Discourse of Biology Lectures : Aspects of its Mode and Text Structure. Unpublished Ph.D. Thesis, Birmingham : Aston University.

Bald, W. D. (1979) "Some functions of Yes-No in conversation", in S. Greenbaum, G. Leech and J. Svartvik (eds.) *Studies in English Linguistics for Randolph Quirk* . London : Longman, pp. 178-191.

Barik, H. C. (1968) "On defining juncture pauses : A note on Boomer's Hesitation and grammatical encoding", Language and Speech 11, pp. 156-159.

Barik, H. C. (1979) "Cross-linguistic study of temporal characteristics of different types of speech materials", Language and Speech 22, pp. 116-126.

Barkho, L and S. S. P. Zora (1984) "An analysis of authentic and non-authentic texts with special reference to the teaching of English conversation to Iraqi university students" Adab Al-Mustansiriyah, 1.10 Baghdad : College of Arts, Al-Mustansiriya University, pp. 101-114.

Bazell, C. E. , J. C. Catford, M. A. K. Halliday and R. H. Robins (eds.) (1966) *In Memory of J.R. Firth* . London : Longman.

Beaman, K. (1984) "Coordination and subordination revisited" in D. Tannen (ed.) *Coherence in Spoken and Written Discourse*. Norwood, New Jersey : Ablex, pp. 45-80.

Beattie, G. W. (1979) "Planning units in spontaneous speech : some evidence from hesitation in speech and speaker gaze direction in conversation", Linguistics 17, pp. 61-78.

Beattie, G. W (1980) "Encoding units in spontaneous speech : some implications for the dynamics of conversation" in H. W. Dechert and M. Raupach (eds.) *Temporal Variables in Speech : Studies in Honour of Frieda Goldman-Eisler*. The Hague : Mouton, pp. 131-144.

Beattie, G. W. (1983) *Talk* . Milton Keynes, Open University Press.

de Beaugrande, R. (1978) "Information, expectation and processing", Poetics, 7, pp. 3-44.

de Beaugrande, R. (1979) "Toward a general theory of creativity", Poetics, 8, pp. 269-306.

de Beaugrande, R. (1980) *Text, Discourse and Process*. London: Longman.

de Beaugrande, R. (1981) "Design criteria for process models of reading", Reading Research Quarterly, 2, pp. 261-315.

de Beaugrande, R. (1984) "The linearity of reading text : Fact, Fiction or Frontier ? in J. Flood (ed.) *Understanding Reading Comprehension*. Delware : International Reading Association, pp. 45-74.

de Beaugrande, R. and R. Dressler (1980) *Introduction to Text Linguistics* . London : Longman.

Bennet, A. (1981) "Interruptions and the interpretation of conversation", Discourse Processes, 4, pp. 171-188.

Benny, M. and E. C. Hughes, (1956) "Of sociology and the interview" Editorial Preface, American Journal of Sociology Vol. 62, pp. 137-142.

Benson, J. D. and W. S. Greaves (1981) "Field of discourse : theory and application" Applied Linguistics, II,1, pp. 43-55.

Benson, J. D. and W. S. Greaves, (eds.) (1985) *Systemic Perspectives on Discourse : elected Theoretical Papers from the Ninth International Systemic Workshop*, Norwood, New Jersey : Ablex.

Bernardo, R. (1979) "The function and content of relative clauses in spontaneous oral narrative" in Proceedings of the Fifth Annual Meeting of the Berkeley Linguistic Society, Feb. 1979, pp. 539-51.

Berry, M. (1975) *An Introduction to Systemic Linguistics 1. Structures and Systems*. London : Batsford.

Berry, M. (1977) *An Introduction to Systemic Linguistics 2. Levels and Links*. London : Batsford.

Berry, M. (1981) "Systemic Linguistics and Discourse Analysis : A multi-layered approach to exchange structure", in M. Coulthard and M. Montgomery (eds.) *Studies in Discourse Analysis*. London: Routledge and Kegan Paul, pp. 120-145.

Bierwisch, M. (1971) *Modern Linguistics : Its Development, Methods and Problems*. Paris, The Hague : Mouton.

Bilton, L. (1987) Coordination and Subordination in Academic Lectures and Textbooks, Unpublished MSc. Dissertation, Aston University, Birmingham.

Blankenship, J. (1962) "A linguistic analysis of oral and written style" Quarterly Journal of Speech, 48, pp. 419-422.

Bloomfield, L. (1933) *Language*. New York : Henry Holt.

Boomer, D. S. (1965)"Hesitation and grammatical encoding", Language and Speech 8, pp. 148-158.

Boomer, D. S. and A. T. Dittman (1962) "Hesitation pauses and juncture pauses in speech", Language and Speech 5, pp. 215-220.

Boomer, D. S. and A. T. Dittman (1964) "Speech rate, filled pause and body movement in interviews" Journal of Nervous and Mental Diseases, 139 (4), pp. 324-327.

Boomer, D. S. and J. D. M. Laver (1968) "Slips of the tongue" British Journal of Disorders of Communication 3, pp. 1-12.

Bowman, E. (1966), *The minor and fragmentary sentences of a corpus of spoken English*. The Hague : Mouton.

Brazil, D. C. (1969) "Kinds of English,spoken, written, literary", Educational Review, 22,1, pp. 78-92.

Brazil, D. C. (1975) "Discourse Intonation", Discourse Analysis Monographs, I, Birmingham, English Language Research, University of Birmingham.

Brazil, D. C. (1978) "Discourse Intonation", **Discourse Analysis Monographs II,** Birmingham, English Language Research, University of Birmingham.

Brazil, D. C. (1981a) "The place of intonation in discourse" in M, Coulthard and M. Montgomery (eds.) ***Studies in Discourse Analysis,*** London : Routledge and Kegan Paul, pp. 146-157.

Brazil, D. C. (1981b) "Discourse analysis as linguistics : a response to Hammersley" in P. French and M. Maclure ***Adult-Child Conversation.*** London : Croom Helm, pp. 59-72.

Brazil, D. C. (1983) "Intonation and discourse : some principles and procedures", **TEXT** 3 (1), pp. 39-70.

Brazil, D. C. (1985) "Phonology : Intonation in Discourse", in T. A. van Dijk, ***Handbook of Discourse Analysis,*** Vol. 2, pp. 57-75.

Brazil, D. C. , M. Coulthard and C. M. Johns (1980) ***Discourse Intonation and Language Teaching.*** London : Longman.

Brenner, M. (1978) "Interviewing : the social phenomenology of a research instrument", in M. Brenner, P. Marsh, and M. Brenner (eds.) ***The Social Contexts of Method.*** London : Croom Helm, pp. 122-139.

Brenner, M. (1981) "Aspects of conversational structure in the research interview" in P. Werth (ed.) **"Conversation and Discourse"**, London : Croom Helm, pp. 19-40.

Brenner, M. P. Marsh and M. Brenner (1978) (eds.) ***The Social Contexts of Method.*** London : Croom Helm.

Broen, P. P. and G. M. Siegel (1969) "Variations in normal speech disfluencies", **Language and Speech** 15, pp. 219-231.

Brown, E. K. (1980) "Grammatical incoherence" in H. W. Dechert and M. Raupach (eds.) ***Temporal Variables in Speech : Studies in Honour of Freida Goldman-Eisler.*** The Hague : Mouton, pp. 28-38.

Brown, E. K and J. E. Miller (1980) ***Syntax : A Linguistic Introduction to Sentence Structure*** . London : Hutchinson.

Brown, G. (1977) ***Listening to Spoken English*** . Hong Kong : Longman.

Brown, G. (1978) "Understanding spoken language" **TESOL** 12,3, pp. 271-283.

Brown, G. (1983) "Prosodic structure and the Given/New distinction", in A. Cutler and D. R. Ladd (eds.) ***Prosody : Model and Measurements.*** Berlin: Springer-Verlag, pp. 66-77.

Brown, G. , K. L. Currie and J. Kenworthy (1980) ***Questions of Intonation.*** London : Croom Helm.

Brown, G. and G. Yule (1983a) ***Discourse Analysis,*** Cambridge : Cambridge University Press.

Brown, G. and G. Yule (1983b) ***Teaching Spoken English.*** Cambridge : Cambridge University Press.

Bruton, G. J., C. N. Candlin and J. H. Leather (1976) "Doctor speech functions in casualty consultations : predictable structures of discourse in a regulated setting" in G. Nickel (ed.) ***Proceedings of the Fourth International Congress of Applied Linguistics.*** Vol. 1, Stuttgart : Hochshul Verlag, pp. 297-312.

Burton, D (1980) ***Dialogue and Discourse.*** London : Routledge and Kegan Paul.

Burton, D. (1981a) "The sociolinguistic analysis of spoken discourse", in P. French and M. Maclure (eds.) ***Adult-Child Conversation.*** London : Croom Helm, pp. 21-46.

Burton, D. (1981b) "Analysing spoken discourse" in M. Coulthard and M. Montgomery (eds.) ***Studies in Discourse Analysis.*** London : Routeldge and Kegan Paul, pp. 61-81.

Butler, C. S. and R. R. K. Hartman (eds.) (1976) ***A Reader : On Language Variety.*** Vol. I Exeter Linguistics Studies, University of Exeter.

Candlin, C. N. (1971) "Sociolinguistics and communicative language teaching", in C. S. Butler and R. R. K. Hartman (eds.) (1976) ***A Reader : On Language Variety.*** Vol. 1 Exeter Linguistic Studies : University of Exeter, pp. 104-110.

Candlin, C. N., Bruton, J. and Leather, J. H. (1974) "Doctor - patient communication skills", **Working Papers 1-4**, Lancaster : University of Lancaster.

Candlin, C. N., J. Bruton and J. H. Leather (1974) "Doctors in Casualty ; applying components of communicative competence to specialist course design" **Work in Progress**, Rep.IV, June 1974.

Carrel, P. L. (1982) "Cohesion is not coherence", TESOL 16,4, pp. 479-488.

Carroll, J. S. and J. W. Payne (eds.) (1976) *Cognition and Social Behaviour*, Hillsdale, New Jersey : Lawrence Erlbaum.

Chafe, W. L. (1970) *Meaning and the Structure of Language*. University of Chicago Press.

Chafe, W. L. (1972) "Discourse structure and human knowledge" in R. O. Freedle and J. B. Carroll (eds.) *Language Comprehension and the Acquisition of Knowledge*. Washington, D. C., V.H. Winston and Sons, pp. 41-69.

Chafe, W. L. (1974) "Language and Consciousness" **Language**, 50, pp. 111-133.

Chafe, W. L. (1976) "Givenness, Contrastiveness, Definiteness, Subjects, Topics, and point of view" in C. N. Li (ed.) *Subject and Topic*. New York Academic Press.

Chafe, W. L. (1977) "Creativity in verbalization and its implications for the nature of stored knowledge", in R. O. Freedle (ed.) *Discourse Production and Comprehension*. Norwood, New Jersey : Ablex, pp. 41-55.

Chafe, W. L. (1979) "The flow of thought and the flow of language", in T. Givon (ed.) *Syntax and Semantics, Vol. 12. Discourse and Syntax.* New York : Academic Press, pp. 159-181.

Chafe, W. L. (ed.) (1980) *The Pear Stories : Cognitive, Cultural and Linguistic Aspects of Narrative Production*. Norwood, New Jersey : Ablex.

Chafe, W. L. (1982) "Integration and involvement in speaking, writing and oral literature" in D. Tannen(ed.) *The Oral/Literate Continuum in Discourse*, Norwood, New Jersey : Ablex, pp. 35:53.

Chafe, W. L. (1984) Mimeo.

Cheshire, J. (1978) "Present tense verbs in Reading English" in P. Trudgill (ed.) *Sociolinguistic Patterns in British English*. London : Edward Arnold, pp. 52-69.

Cheshire, J. (1982) *Variation in an English Dialect : A Sociolinguistic Study*. Cambridge : Cambridge University Press.

Chomsky, N. (1965) *Aspects of the Theory of Syntax*. Mass. MIT Press.

Chomsky, N. (1969) "Should traditional grammar be ended or mended ?", Educational Review, 22,1, pp. 5-17.

Clancy, P. (1972) "Analysis of a conversation", Anthropological Linguistics 14, pp. 78-86.

Clark, H. H. (1975) "Bridging", in R. Schank and B. Nash-Webber (eds.) *Theoretical Issues in Natural Language Processing.* Proceedings of a conference at the Massachusetts Institute of Technology, June 1975, pp.

Clark, H. H. (1978) "Inferring what is meant" in W. J. M. Levelt and G. B. Flores d'Archais (eds.) *Studies in the Perception of Language.* New York : Wiley, pp. 295-323.

Clark, H. H. and E.V. Clark (1977) *Psychology and Language : An Introduction to Psycholinguistics.* New York : Harcourt Brace Jovanovich, Inc.

Clark, H. H. and S. E. Haviland (1974) "Psychological processes as linguistic explanation", in D. Cohen (ed.) *Explaining Linguistic Phenomena.* Washington, D. C.: Hemisphere Publishing, 91-124.

Clark, H. H. and S. E. Haviland (1977) "Comprehension and the Given/New Contract" in R. O. Freedle (ed.) *Discourse Processes : Advances in Research and Theory.* Norwood, New Jersey : Ablex, pp. 1-40.

Clark, H. H. and P. Lucy (1975) "Understanding what is meant from what is said : a study in conversationally conveyed requests", Journal of Verbal Learning and Verbal Behaviour, 14, pp. 56-72.

Clark, H. H. and R. Marshall (1981) "Definite reference and mutual knowledge" in B. Joshi, B. Webber and I. Sag *Elements of Discourse Understanding.* Cambridge : Cambridge University Press, pp. 10-36.

Coates, J. (1986) *Women, Men and Language : A sociolinguistic Account of Sex Differences in Language.* London : Longman.

Coates, J. (1989) "Gossip revisited : language in all-female groups", in J. Coates and D. Cameron, *Women in their Speech Communities.* London : Longman.

Coates, J. and D. Cameron, (1989)*Women in their Speech Communities.* London : Longman.

Cohen D. (ed.) (1974) *Explaining Linguistic Phenomena.* Washington, D.C.: Hemisphere Publishing.

Cole, P. (1978) "On the origins of referential opacity" in P. Cole (ed.) *Syntax and Semantics : Pragmatics,* London : Academic Press, pp. 1-22.

Cole, P. (ed.) (1978) *Syntax and Semantics : Pragmatics.* New York : Academic Press.

Cole, P. (ed.) (1981) *Radical Pragmatics.* New York : Academic Press.

Cole P. and J. L. Morgan (eds.) (1975) *Syntax and Semantics 3 : Speech Acts .* New York : Academic Press.

Coleman, J. (1958) "Relational analysis : The study of social organization with survey methods", Spring 17, pp. 28-36.

Corder, S. P. (1973) *Introducing Applied Linguistics,* Hamondsworth, Middlesex : Penguin.

Coulthard, M. (1975) "Discourse analysis : a short review of the literature", Language and Linguistic Abstracts, 8, pp. 73-89.

Coulthard, M. (1977) *An Introduction to Discourse Analysis* London : Longman.

Coulthard, M. and M. Ashby, (1976) "A linguistic description of doctor-patient interview" in M. Wadsworth and D. Robinson (eds.) *Studies in Everyday Medical Life.* London : Martin Robertson, pp. 1-16.

Coulthard, M and D. Brazil (1979) "Exchange structure" in M. Coulthard and M. Montgomery (eds.) *Studies in Discourse Analysis,* London : Routledge and Kegan Paul, pp. 82-106.

Coulthard, M. and M. Montgomery (1981) *Studies in Discourse Analysis.* London : Routledge and Kegan Paul.

Coulthard, M., M. Hoey and M. Knowles (eds.) (1987) *Discussing Discourse: Sudies Presented to David Brazil on his Retirement.* Discourse Analysis Monographs, Birmingham : University of Birmingham, English Language Research.

Cross, C. P. (1974) (ed) *Interviewing and Communication in Social Work.* London : Routledge and Kegan Paul.

Crystal, D. (1974) *What is linguistics ?* London : Edward Arnold.

Crystal, D. (1975) *The English Tone of Voice.* London : Edward Arnold.

Crystal, D. (1980) "Neglected grammatical factors in conversational English" in S. Greenbaum, G. Leech and J. Svartvik (eds.) *Studies in English Linguistics for Randolph Quirk.* London : Longman, pp. 169-182.

Crystal, D. (1981) *Directions in Applied Linguistics*. London : Academic Press.

Crystal, D. and D. Davy (1969) *Investigating English Style.* London : Longman.

Crystal, D and D. Davy (1975) *Advanced Conversational English.* London : Longman.

Cutler, A. and D. R. Ladd (eds.) (1983) *Prosody : Models and Measurements.* Berlin : Springer-Verlag.

Cutler, A. and M. Pearson (1986) "On the analysis of turn-taking cues" in C. M. Johns-Lewis (ed.) *Intonation in Discourse,* London: Croom Helm, pp. 139-155.

Dahl, O. (1974) "Topic-Comment Structure in a Generative Grammar with a Semantic Base. In F. Danes (ed.) *Papers in Functional Sentence Perspectives*. Prague : Academia, pp. 75-81.

Danes, F. (1974) "Functional Sentence Perspective and the organisation of the text", in F. Danes, (ed.) *Papers in Functional Sentence Perspective.* Prague : Academia, pp. 106-128.

Danes, F. (ed.) (1974) *Papers in Functional Sentence Perspective.* Prague : Academia.

Dechert, H. W. and M. Raupach (eds.) (1980) *Temporal Variables in Speech: Studies in Honour of Frieda Goldman-Eisler* . The Hague : Mouton.

Denzin, N. (1970) *The Research Act in Sociology* . London : Butterworth & Co. Ltd.

Denzin, N. (1972) "Symbolic interactionism and ethnomethodology" in Douglas, J. D. (ed.) *Understanding Everyday Life.* London : Routledge, pp. 259-285.

De Silva S. (ed.) *Aspects of Linguistic Behaviour.* York : University of York.

De Vito, J. (1964) **A quantitative analysis of comprehension factors in samples of oral and written technical discourse of skilled communicators.** Unpublished Ph. D. Thesis : University of Illinois.

De Vito, J. (1965) " Comprehension factors in oral and written discourse of skilled communicators", **Speech Monographs**, 32, pp. 124-128.

De Vito, J. (1966) "Psychogrammatical factors in oral and written discourse by skilled communicators", **Speech Monographs**, 33, pp. 73-76.

De Vito, J. (1967) "Levels of abstraction in spoken and written language", **Journal of Communication**, 17, pp. 354-361.

van Dijk, T. A. (1977) ***Text and Context : Explorations in the Semantics and Pragmatics of Discourse.*** London : Longman.

van Dijk, T. A. (1979) ***Macro-Structures.*** Hillsdale, New Jersey : Lawrence Erlbaum.

van Dijk, T. A. (1980) ***Studies in the Pragmatics of Discourse.*** The Hague : Mouton.

van Dijk, T.A. (1984) ***Prejudice in Discourse.*** Amsterdam : Benjamins.

van Dijk, T. A. (ed.) (1985) ***Handbook of Discourse Analysis. Vol. 1: Disciplines of Discourse.*** London : Academic Press.

van Dijk, T. A. (ed.) (1985) ***Handbook of Discourse Analysis. Vol. 2: Dimensions of Discourse.*** London : Academic Press.

van Dijk, T. A. and W. Kintsch (1983) ***Strategies of Discourse Comprehension.*** London : Academic Press.

van Dijk, T. A. and J. Petofi (eds.) ***Grammars and Descriptions*** . Berlin : de Gruyter.

Donnellan, K. S. (1978) "Speaker reference, descriptions and anaphora" in P. Cole (ed.) **Syntax and Semantics : Pragmatics.** London : Academic Press, pp. 47-68.

Douglas, J. (ed.) (1970) ***Understanding Everyday Life*** . New York : Aldine.

Dressler W. (ed.) (1978) ***Current Trends in Text Linguistics.*** Berlin : Walter de Gruyter.

Drieman, G. (1962) "Differences between written and spoken language", **Acta Psychologica**, 20, pp. 36-57; 78-100.

Duncan, S. D. Jr. (1972) "Some signals and rules for taking speaking turns in conversations" Journal of Personality and Social Psychology 23, pp. 283-292.

Duncan, S. D. Jr. (1973) "Towards a grammar for dyadic conversations", Semiotica 9, pp. 29-46.

Duncan, S. D. Jr. (1974a) "On the structure of speaker-auditor interaction during speaking turns", Language in Society, 3, pp. 161-180.

Duncan, S. D. Jr. (1974b) "Speaker-auditor interaction during speaking turns" Language in Society, 3, pp. 162-180.

Duncan, S. D. Jr. (1974c) "On signalling that it is your turn to speak" Journal of Experimental Social Psycology, 10, pp. 234-247.

Duncan, S. D. Jr. (1975) "Language, paralanguage and body motion in the structure of conversation" in W. C. McCormack and S. A. Wurm (eds.) *Language and Thought* . The Hague : Mouton, pp. 239-265.

Duncan, S. D. Jr. and G. Niederehe (1974) "On signalling that it's your turn to speak", Journal of Experimental Social Psychology, 10, pp. 234-247.

Edelsky, C. (1981) "Who's got the floor", Language in Society, 10, pp. 383-421.

Edmundson, W. (1981) *Spoken Discourse : A Model For Analysis* . London : Longman.

Eeg-Olofsson, M and J. Svartvik (1984) "Four-level tagging of spoken English" in J. Aarts and W. Meijs (eds.) *Corpus Linguistics* . Holland : Rodopi, pp. 53-64.

Ellis, J. (1965) "Linguistic sociology and institutional Linguistics" Linguistics, 19, pp. 5-12.

Ellis, J. (1966) "On contextual meaning", in Bazell et al. *In Memory of J.R. Firth.* London : Longman.

Ellis, J. and J. N. Ure (1969) "Registers" in C. S. Butler and R. R. K. Hartman (eds.) (1976) *A Reader on Language Variety.* Vol. 1, Exeter Linguistics Studies, University of Exeter, pp. 32-40.

Enkvist, N. E. (1978) "Coherence, Pseudo-Coherence and Non Coherence" in J-O. Östman (ed.) *Cohesion and Semantics* . Abo, Finland : Abo Akademi Foundation, pp. 109-128.

Enkvist, N. E. (1978) "Stylistics and Text-Linguistics" in W. Dressler (ed.) *Current Trends in Text Linguistics* . Walter de Gruyter, pp. 174-190.

Enkvist, N.E. (ed.) (1982) *Impromtu Speech : A Symposium* . Abo : Abo Akademi.

Evett, L., F. Knowles and B. Williams (1985) Mechanical Determination of the "Breathgroup" in English : The cognitive-linguistic unit in natural language processing. Final report to the British Library R & D Department ; Cambridge : Cambridge Language Research Unit.

Farag, S. M. (1983) "Discourse analysis of spoken guided tours : a critique of an ESP project" MALS, 9 (1984), pp. 65-86.

Farag, S. M. (1986) A Linguistic Analysis of Spoken and Written Narrative Discourse. Unpublished Ph.D Thesis, Birmingham : Aston University.

Farrugia, A. (1988) An Examination of Lexical Density in Newspaper Editorials. Unpublished MSc. Dissertation, Birmingham : Aston University.

Fasold, R. W. and R. W. Shuy (eds.) (1977) *Studies in Language Variation : Semantics, Syntax, Phonology, Pragmatics, Social Situation, Ethnographic Approaches.* Washington : Georgetown University Press.

Fawcett, R.P (1980) *Cognitive Linguistics and Social Interaction : Towards and Integrated Model of a Systemic Functional Grammar and the other Components of a Communicating Mind,* Heidelberg, Julius Groos Verlag and Exeter University.

Ferguson, N. (1975) Interruptions : Speaker Switch non-Fluency in Spontaneous Conversation. Unpublished Ph.D. Thesis, Edinburgh : Edinburgh University Press.

Ferguson, Ch. A. (1983) "Sports announcer talk : syntactic aspects of register variation" Language in Society, 12, pp. 153-172.

Fillmore, Ch. J. (1973) "May we come in ?" SEMIOTICA, IX, pp. 97-116.

Fillmore, Ch. J. (1981) "Pragmatics and the description of discourse" in P. Cole *Radical Pragmatics* . New York : Academic Press, pp. 143-166.

Fine, J. and R. O. Freedle (eds.) (1983) *Developmental Issues in Discourse* . Norwood, New Jersey : Ablex.

Firbas, J. (1964) "On defining the Theme in Functional Sentence Analysis" TRAVAUX LINGUISTIQUE DE PRAGUE, I, pp. 267-380.

Firbas, J. (1966) "Non-Thematic Subject in Contemporary English" TRAVAUX LINGISTIQUE DE PRAGUE, 2, pp. 239-256.

Firbas, J. (1974) "Some aspects of the Czechoslovak Approach to the problem of Functional Sentence Perspective", in F. Danes , *Papers in Functional Sentence Perspectives.* Prague : Academia, pp. 11-37.

Firbas, J. (1975) "On 'Existence/Appearance on the Scene" in Functional Sentence Perspective, Prague Studies in English, Vol. XVI, pp. 47-70.

Firth, J. R. (1935) "Modes of Meaning", in *Essays and Studies.* 1951, reprinted in *Papers in Linguistics.* 1935-1951, London, Oxford University Press, pp. 190-215.

Fishman, J. A. (1969) "The uses of sociolinguistics", in G. E. Perren and J. L. M. Trim (ed.) (1971), *Applications of Linguistics* . Cambridge : Cambridge University Press, pp. 19-40.

Flood, J. (ed.) (1984) *Understanding Reading Comprehension : Cognition, Language, and the Structure of Prose* . Newark, Delware : International Reading Association.

Fodor, J. A. and M. Garret (1967) "Some syntactic determinants of sentential complexity" Perception and Psychophysics, 2, 7, pp. 289-296.

Francis, G. and S. Hunston (1987) "Analysing everyday conversation", in M. Coulthard, M. Hoey and M. Knowles (eds.) (1987) *Discussing Discourse : Studies Presented to David Brazil on his retirement.* Discourse Analysis Monographs, Birmingham : University of Birmingham, English Language Research, pp. 107-148.

Franck, D. (1979) "Speech act and conversational move", Journal of Pragmatics 3, pp. 461-466.

Franck, D (1981) "Seven sins of pragmatics: Theses about speech act theory, conversation analysis and Rhetoric" in H. Parret, M. Sbisa and J. Verschueren (eds.) *Possibilities and Limitations of Pragmatics* . Amsterdam : Benjamins, pp. 225-236.

Fraser, B. (1983) "The domain of pragmatics" in J.C.Richards and R.W. Schmidt Language and Communication, London : Longman, pp. 29-59.

Freedle, R. O. (ed.) (1971) *Discourse Production and Comprehension* . Norwood , New Jersey : Ablex.

Freedle, R. O. (ed.) (1977)*Discourse Processes : Advances in Research and Theory*. Norwood, New Jersey : Ablex. Goldman-Eisler,F. (1968) *Psycholinguistics : Experiments in Spontaneous Speech.* London : Academic Press.

Freedle, R. O. (ed.) (1979) *New Directions in Discourse Processing* . Norwood, New Jersey : Ablex.

French, P. and Local, J. (1986) "Prosodic features and the management of Interruptions" in C. M. Johns-Lewis *Intonation in Discourse* . London : Croom Helm, pp. 157-177.

French, P. and M. Maclure (eds.) (1981) *Adult-Child Conversation* . London : Croom Helm.

Frederiksen, C. "Semantic processing units in understanding text", in R. Freedle (ed.) *Discourse Production and Comprehension.* Norwood, New Jersey: Ablex, pp. 57-86.

Garfinkel, H. (1967) *Studies in Ethnomethodology* . Englewood Cliff, New Jersey : Prentice Hall.

Garfinkel, H. (1972) "Remarks on ethnomethodology" in J. J. Gumperz and D. Hymes, *Directions in Sociolinguistics* . New York : Holt, Rinehart and Winston, pp. 301-324.

Garfinkel, H. and H. Sacks (1970) "On formal structures of practical actions in J. C. McKinney and E. Tiryakian (eds.) *Theoretical Sociology* NewYork : Appleton, pp. 337:366.

Giglioli, P. P. (ed.) (1972) *Language and Social Context : Selected Readings.* Harmondsworth, Middlesex : Penguin.

Giles, H. (1980) "Accommodation Theory : Some new directions", in S. De Silva (ed.) *Aspects of Linguistic Behaviour.* York : University of York pp. 105-125.

Giles, H. and R. St Clair, (1979) *Language and Social Psychology* . Oxford : Basil Blackwell.

Givon, T. (1979) "From discourse to syntax : grammar as a proceeding strategy" in T. Givon (ed.) *Syntax and Semantics :, Vol. 12 Discourse and Syntax* . New York : Academic Press, pp. 81-112.

Givon, T. (ed.) (1979) *Syntax and Semantics : Vol. 12, Discourse and Syntax.* New York : Academic Press.

Goffman, E. (1976) "Replies and responses" **Language in Society** 5, pp. 257-313.

Goffman, E. (1979) "Response cries" **Language**, 54, pp. 787-815.

Goffman, E. (1981) *Forms of Talk* . Philadelphia, University of Pennsylvania Press.

Goldman-Eisler, F. (1954a) "A study of individual differences and of interaction in the behaviour of some aspects of language in interviews" **Journal of Mental Science** 100, pp. 177-194.

Goldman-Eisler, F. (1954b) "On the variability of the speed of talking and on its relation to the length of utterances in conversations", **British Journal of Psychology** XLV, pp. 94-107.

Goldman-Eisler, F. (1956) "The dominants of the rate of speech output and their mutual relations", **Journal of Psychosom Research** 1, pp. 137-143.

Goldman-Eisler, F. (1958) "Speech analysis and mental processes" **Language and Speech** 1, pp. 59-75.

Goldman-Eisler,F. (1968) *Psycholinguistics : Experiments in Spontaneous Speech.* London : Academic Press.

Goldman-Eisler, F. (1972) "Pauses, clauses and sentences", **Language and Speech**, 15, 103-113.

Golub, L. (1969) "Linguistic Structures in Students' oral and written discourse", **Research in the Teaching of English**, 3, pp. 70-85.

Goodenough, D. R. and S. L. Weiner (1979) "The role of conversational moves in the management of topical transitions", **Discourse Processes** Vol. I pp. 395-403.

Goody, J. (1977) *The Domestication of the Savage Mind* . Cambridge : Cambridge University Press.

Gorden, R. L. (1969) *Interviewing, Strategy, Techniques and Tactics* . Illinois : The Dorsey Press.

Gordon, D. and G. Lakoff (1975) "Conversational postulates" in P. Cole and J. L. Morgan (eds.) *Syntax and Semantics 3 : Speech Acts* New York : Academic Press, pp. 83-106. (Reprint of Gordon and Lakoff 1971).

Greatbatch, D. (1988) "A turn-taking system for British news interviews", Language in Society, 17, pp. 401-430.

Greenbaum, S., G. Leech and J. Svartvik (1980) (eds.) *Studies in English Linguistics for Randolph Quirk* . London : Longman.

Gregory, M. (1967) "Aspects of varieties differentiation", Journal of Linguistics, 3, pp. 177-198.

Gregory, M. and S. Carroll (1976) *Language and Situation : Language Varieties and their Social Contexts.* London : Routledge and Kegan Paul.

Grice, H. P. (1957) "Meaning" Philosophical Review, 66 , pp. 377-388.

Grice, H. P. (1975) "Logic and Conversation" in P. Cole and J. L. Morgan (eds.) *Syntax and Semantics 3 : Speech Acts* . New York : Academic Press, pp. 41-58.

Grice, J. (1981) "Presupposition and conversational implicature" in P. Cole (ed.) *Radical Pragmatics* . New York : Academic Press, pp. 183-198.

Grimes, J. E. (1975) *The Thread of Discourse* . The Hague : Mouton.

Grimes, J. E. (1978) "Narrative studies in in oral texts" in W. Dressler (ed.) *Current Trends in Text Linguistics* . Walter de Gruyter, pp. 123-132.

Gumperz, J. J. (1982) *Discourse Strategies.* Cambridge : Cambridge University Press.

Gumperz, J. J. and D. Hymes (eds.) *Directions in Sociolinguistics* . New York: Holt, Rinehart and Winston.

Gumperz, J. J., H. Kaltman and M. C. O'Connor (1984) "Cohesion in spoken and written discourse : Ethnic style and the transition to literacy" in D. Tannen *Coherence in Spoken and Written Discourse.* Norwood, New Jersey : Ablex , pp. 3-19.

Gutwinski, W. (1976) *Cohesion in Literary Texts* , The Hague : Mouton.

Halle, M. J. Bresnan and G. A. Miller (eds.) (1978) *Linguistic Theory and Psychological Reality* Mass.: The MIT Press.

Halliday, M. A. K. (1961) "Categories of the theory of grammar" WORD, 17,3, pp. 241-292.

Halliday, M. A. K. (1966) "Lexis as a linguistic level" in C.B. Bazell et al. (eds.) *In Memory of J.R. Firth*. London : Longman, pp. 148-162.

Halliday, M. A. K. (1967) "Notes on transitivity and theme in English, Part 1" Journal of Linguistics, 3, 1, pp. 37-88.

Halliday, M. A. K. (1967) "Notes in Transitivity and Theme in English", Part 2, Journal of Linguistics, 3, 2, pp. 177-274.

Halliday, M. A. K. (1970) "Language structure and Language function", in J. Lyons, (ed.) *New Horizon in Linguistics* Harmondsworth, Middlesex : Penguin, pp. 140:165.

Halliday, M. A. K. (1973) *Explorations in the Functions of Language* . London : Edward Arnold.

Halliday, M. A. K. (1976) "Early language learning : a sociolinguistic approach", in W. C. McCormack and S. A. Wurm (eds.) (1977) *Language and Thought-Anthroplogical Issues* . The Hague : Mouton, pp. 99-124.

Halliday, M. A. K. (1977) "Text as a semantic choice in social contexts" in T. van Dijk and J. Petofi (eds.) *Grammars and Descriptions* . Berlin : de Gruyter, pp. 176-225.

Halliday, M. A. K. (1979) "Development of texture in child language" in T. Myers (ed.) *The Development of Conversation and Discourse* . Edinburgh : Edinburgh University Press, pp. 72-87.

Halliday, M. A. K. (1985a) *An Introduction to Functional Grammar.* London : Edward Arnold.

Hallidy M. A. K. (1985b) *Spoken and Written Language* . Victoria: Deakin University Press.

Halliday. M. A. K. and R. Hasan (1976) *Cohesion in English* . London : Longman.

Halliday, M. A. K., A. McIntosh and P. Strevens (1964) *The Linguistic Sciences and Language Teaching* . London : Longman.

Hankamer, J. and I. Sag (1976) "Deep and surface anaphora" Linguistic Inquiry 7, 3. pp. 391-426.

Harper, W. M. (1965) *Statistics.* Plymouth : Macdonald and Evans LTD.

Harris, J., D. Little, and D. Singleton (1985) Proceedings of the First Symposium on Hiberno-English at Trinity College. Dublin.

Harris, Z. S. (1952) "Discourse analysis" Language, 28, pp. 1-30.

Hasan, R. (1976) "Socialization and cross-cultural education" International Journal of the Sociology of Language, 8, pp. 7-25.

Hasan, R. (1978) "Text in the systemic-functional model" in W. Dressler *Current Trends in Text Linguistics.* Berlin : Walter de Gruyter, pp. 228-246.

Hasan, R. (1984) "Coherence and cohesive harmony" in J. Flood (ed.) *Understanding Reading Comprehension : Cognition, Language and the Structure of Prose* . New York : International reading Association, pp. 181-219.

Hasan, R. (1985) "Meaning, text and context : Fifty years after Malinowski" in J. D. Benson and W. S Greaves (eds.) *Systemic Perspectives on Discourse: Selected Theoretical Papers from the Ninth International Systemic Workshop,* Norwood, New Jersey : Ablex.

Hassan, A. (1989) Variation in Spoken Discourse in and Beyond the English Foreign Language Classroom . Unpublished Ph.D Thesis, Birmingham : Aston University.

Hatim, B. (1981) Discourse Analysis Models : Towards an Applied Linguistics Definition of Text Variation . Unpublished Ph.D. Thesis, Exeter : University of Exeter.

Haviland, S. E. and H. H Clark (1974) "What's new ? acquiring new information as a process in comprehension" Journal of Verbal Language and Verbal Behaviour, 13, pp. 512-521.

Hawkins, P. R. (1971) "The syntactic location of hesitation pauses" , Language and Speech 14.1, pp. 277-288.

Hawkins, J. A. (1978) *Definiteness and Indefiniteness* . Atlantic Highlands, New Jersey : Humanities Press.

Hays, D. G. (1971) "Applied computational linguistics" in G. E. Perren and J. L. M. Trim (eds.) (1971) *Applications of Linguistics.* Cambridge : Cambridge University Press, pp. 65-84.

Hobbs, J. R. (1978a) Why Is Discourse Coherent ? Technical Note, 176, Menlo Park : Stanford Research Centre.

Hobbs J. R. (1978b) "Coherence and coreference" Cognitive Science 3 (1), pp. 67-90.

Hobbs, J. R. and J. J. Robinson (1979) "Why ask ?" **Discourse Processes**, 2, pp. 311-318.

Hoey, M. (1983) *On the Surface of Discourse* . London : George Allen and Unwin.

Holec, H. (1977) "You did say oral interactive discourse" **Applied Linguistics**, 1,3, pp. 189-199.

Horowitz, M. and J. Newman (1964) "Spoken and Written expression : an experimental analysis" , **Journal of Abnormal Social Psychology** . 68. pp. 640-647.

Hudson, R. A. (1980) *Sociolinguistics* . Cambridge : Cambridge University Press.

Hughes, A. and P. Trudgill (eds.) (1979) *English Accents and Dialects : An Introduction to Social and Regional Varieties of British English* . London : Edward Arnold.

Hunt, K. W. (1971) "Teaching syntactic maturity" in G .E. Parren and J. L. M. Trim (eds.) (1971) *Applications of Linguistics.* Cambridge : Cambridge University Press, pp. 287-300.

Hymes, D. (1972a) "Models of interaction of language and social life" in J. J. Gumperz and D. Hymes (eds.) *Directions in Sociolinguistics : The Ethnography of Communication.* New York : Holt, Rinehart & Winston, pp. 35-71.

Hymes, D. (1972b) "Toward an ethnography of communication : the analysis of communicative events" in J.J Gumperz and D. Hymes (eds.) *Directions in Sociolinguistics : The Ethnography of Communicative events.* New York : Holt, Rinehart & Winston, pp. 21-43.

Hymes, D. (1974) *Foundations in Sociolinguistics : An Ethnographic Approach* . Philadelphia : University of Pennsylvania Press.

James, C. (1979) "Macrolinguistic contrastive analysis", **Mimeo**, UCNW, Bangor.

James, C. (1980) *Contrastive Analysis* . London : Longman.

Jefferson, G. (1972) "Side sequences" in D. Sudnow (ed.) *Studies in Social Interaction* . New York : Free Press, 294-338.

Jefferson, G. (1973) "A case of precision timing in ordinary conversation", **Semiotica**, 9,1 pp. 47-96.

Jefferson, G. (1974) "Error Correction as an interactional resource", Language in Society 3, pp. 181-200.

Jefferson, G. and J. Schenkein (1978) "Some sequential negotiations in conversation : Unexpanded and expanded versions of projected action sequences" in J. Schenkein (ed.) *Studies in the Organization of Conversational Interaction* New York : Academic Press, pp. 155-172.

Johns-Lewis, C. M. (1972) Slips of the Tongue in Purtogese. Unpublished M. Litt. Dissertation, Edinburgh : University of Edinburgh.

Johns-Lewis, C. M. (1986) "Prosodic differentiations of discourse modes" in C. M. Johns-Lewis (ed.) (1986) *Intonation and Discourse* . London : Croom Helm, pp. 199-216.

Johns-Lewis, C. M. (1986) (ed.) *Intonation and Discourse* London : Croom Helm.

Johns-Lewis, C. M. (1987) "Linguistic differentiation of spoken and written language : implications for language teaching and learning" in D. E. Ager (ed.) *Written Skills in the Modern Languages degree* . AMLC/CILT, Department of Modern Languages : Aston University, pp. 10-19.

Johns-Lewis, C. M. and J. Skelton (1987) "Variation in discourse", Paper presented at conference on 'Variation in Second Language Acquisition', English Language Institute, Ann Arbor : Michigan University.

Joshi, B., B, Webber and I. Sag (1981) *Elements of Discourse Understanding.* Cambridge : Cambridge University Press.

Just, M. A. and P. A. Carpenter (1977) *Cognitive Processes in Comprehension.* Hillsdale, New Jersey : Lawrence Erlbaum.

Kalin, R. and D. Rayko (1980) "The social significance of speech in the job interview" in R. N. St. Clair (ed.) *The Social Psychological Contexts of Language* . Hillsdale, New Jersey : Lawrence Erlbaum, pp. 39:41.

Kaplan, R. B. (1976) "Some remarks on the scope of applied linguistics" in G. Nickle (ed.) "Proceedings of the 4th International Congress of Applied Linguistics". Stuttgart : Hochschulverlag, pp. 77-80.

Karttunen, L. (1971) "Discourse Referents" in J.D. McCawley (ed.) *Syntax and Semantics, Vol.7, Notes from the Linguistic Underground* . New York : Academic Press, pp. 364-382.

Keenan, E. O. (1970) "The universality of conversational postulates", Language in Society, 5, pp. 67-80.

Keenan, E. O. and T. Bennet (eds) (1977) *Discourse across Time and Space.* Southern California Occasional Papers in Linguistics, 5. Los Angeles : University of Southern California.

Kempson, R. M. (1977) *Semantic Theory.* Cambridge : Cambridge University Press.

Kendon, A. (1967) "Some functions of gaze direction in social interaction", Acta Psychologica 26, pp. 22-63.

Khan, R. L. and Cannel, C. F. (1957) *The Dynamics of Interviewing* . New York : John Wiley.

Kintsch, W. (1977a) *Memory and Cognition.* New York : Wiley.

Kintsch, W. (1977b) "On comparing stories", in M. A. Just and P. Carpenter (eds.) *Cognitive Processes in Comprehension.* Hillsdale, New Jersey: Lawrence Erlbaum, pp. 33-62.

Kreiman, J. (1982) "Perception of sentence and paragraph boundaries in natural conversation", Journal of Phonetics 10, 2, pp. 163-175.

Kroll, B. (1977) "Combining ideas in written and spoken English: a look at subordination and coordination" in E. O. Keenan and T. Bennet (eds) *Discourse across Time and Space.* Southern California Occasional Papers in Linguistics, 5. Los Angeles : University of Southern California, pp. 69-108.

La berg, D. and S. J. Samuels (eds.) (1977) *Basic Processes in Reading : Perception and Completion* . Hillsdale, New Jersey : Lawrence Erlbaum.

Labov, W. (1970) "The study of language in its social context" Stasium Generale, 23, pp. 66-84.

Labov, W. (1972a) *Sociolinguistic Patterns* Philadelphia : University of Pennsylvania Press.

Labov, W. (1972b) "The transformation of experience in narrative syntax" in W. Labov, *Sociolinguistic Patterns* . Philadelphia : University of Pennsylvania Press, pp. 354-397.

Labov, W. (1972c) "Some principles of linguistic methodology", Language in Society, 1,1, pp. 97-120.

Labov, W. (1972d) "Rules for ritual insults" in D. Sudnow (ed.) *Studies in Social Interaction* New York : Free Press, pp. 120-169.

Labov, W. and Fanshel, D. (1977) *Therapeutic Discourse : Psychotherapy as Conversation* . New York : Academic Press.

Lakoff, G. (1976) "Pronouns and reference" in J. D. McCawley (ed.) *Syntax and Semantics : Vol.7, Notes from the Linguistic Underground* New York: Academic Press, pp. 275-337.

Lakoff, R. (1974) "Remarks on this and that " **10th Regional meeting of Chicago Linguistic Society**. April (1974) pp. 345-356.

Lakoff, R. (1979) "Expository writing and the oral dyad as points on a communicative continuum", Mimeo.

Lakoff, R. (1982) "Some of my favourite writers are literate : the mingling of oral and literate strategies in written communication" in D. Tannen *Spoken and Written Language : Exploring Orality and Literacy* . Norwood, New Jersey : Ablex, pp. 239-260.

Lang, G. (1984) **A linguistic analysis of the radio news interview.** Unpublished project, Department of Modern Languages, University of Aston.

Levelt, W. J. M. and G. B. Flores d'Archais (eds.) (1978) *Studies in the Perception of Language*. New York : Wiley.

Levinson, S. C. (1979) "Activity types and language", **Linguistics** 17,5/6, pp. 356-399.

Levinson, S. C. (1980) "Speech act theory : the state of the art" , **Language and Linguistics Teaching Abstracts** 13,1, pp. 5-24.

Levinson, S. C. (1981a) "The essential inadequacies of speech act models of dialogue" in H. Parret, M. Sbisa and J. Verschueren (eds.) *Possibilities and Limitations of Pragmatics.* Amsterdam : Benjamins, pp. 473-492.

Levinson, S. C. (1981b) "Some pre-observation in the modelling of the dialogue" **Discourse Processes**, 4,2, pp. 93-110.

Levinson, S. C. (1983) *Pragmatics* Cambridge : Cambridge University Press.

Li, Charles, N. (ed.) (1976) *Subject and Topic* . New York : Academic Press.

Lieberman, P. (1980) "The innate central aspect of intonation" in L. R. Waugh and C. H. van Schooneveld (eds.) *The Melody of Language.* Baltimore : University Park Press, pp. 187-199.

Linde, Ch. (1977) "Information structuring in discourse" in R. W. Fasold and R. W. Shuy (eds.) *Studies in Language Variation.* Washington: Georgetown University Press, pp. 226-236.

Linde, Ch. (1983) "A framework for formal models of discourse, what can we model and why ?" **TEXT** 3 (3), pp. 271-276.

Linde, Ch. and W. Labov (1975) "Spatial networks on a site for the study of language and thought" **Language** 51,4, pp. 924-939.

Longacre, R. W. (1983) *The Grammar of Discourse* . New York : Plenum Press.

Longacre, R. and S. Levinson (1978) "Field analysis of discourse" in W. M Dressler (ed.) *Current Trends in Text Linguistics* . Walter de Gruyter. pp. 103-121.

Lyons, J. (1968) *Introduction to Theoretical Linguistics* . Cambridge : Cambridge University Press.

Lyons, J. (ed.) (1970) *New Horizons in Linguistics* . Harmondsworth, Middlesex : Penguin.

Lyons, J. (1975) "Deixis as the source of reference" in E. L. Keenan (ed.) *Formal Semantics of Natural Language* . Cambridge : Cambridge University Press, pp. 61-83.

Lyons, J. (1977a) *Semantics* . Vol. 1 & 2. Cambridge : Cambridge University Press.

Lyons, J. (1977b) "Deixis and anaphora" in T. Myers (ed.) (1979) *The Development of Conversation and Discourse* Edinburgh, Edinburgh University Press, pp. 88-104.

Maccoby, E. E. and N. Maccoby(1954) "The interview : a tool of social science", in G. Lindzey (ed.) *Handbook of Social Psychology* . Reading, Mass., : Addison-Wesley, Vol. 1, pp. 449-487.

Maclay, H. and C. E. Osgood (1959) "Hesitation phenomena in spontaneous English speech", **Word** 15, pp. 19-44.

Mahl, G. F. (1954) "Disturbances and silences in the patient's speech in psychotherapy", **Journal of Social Psychology** 53, pp. 1-14.

Malinowski, B. (1923) "The problem of meaning in primitive languages", Supplement 1 to Ogden and Richards, *Meaning of Meaning*, pp. 296-346.

Martin, J. R. (1983) "The development of register" in J. Fine and R. O. Freedle (eds.) ***Developmental Issues in Discourse*** . Norwood, New Jersey : Ablex, pp. 1-38.

Martin, J. R. (1984) "On the analysis of exposition" in R. Hasan (ed.) ***Discourse on Discourse***. Sydney Applied Linguistic Association of Australia, Publication 7, pp.

Mathesius, V. (1939) "On So-called Functional Sentence Perspective", Slova a slovensnort, Vol. 7, pp. 177-174.

Mathesius, V. (1941) "The basic function of word order in Czech", Slova a slovensnort, Vol. 7, pp. 169-180.

Matthews, P. H. (1981) ***Syntax*** . Cambridge : Cambridge University Press.

Mayer, R. E. (1982) "Instructional variables in text processing" in A. Flammer and W. Kintsch (eds.) ***Discourse Processing*** . North Holland Pub. Company, pp. 443-461.

McCawley, J. D. (ed.) (1976)***Syntax and Semantics, Vol.7, Notes from the Linguistic Underground*** . New York : Academic Press.

McCormack W. C. and S. A. Wurm (eds.) (1977) ***Language and Thought : Anthropological Issues*** . The Hague : Mouton.

McGregor, G. (1982) "Intonation and meaning in conversation" Language in Communication, 2,2 pp. 123-131.

McGregor, G. (1983) "Listeners' comments on conversation" Language and Communication, 3,3, pp. 271-304.

McGregor, G. (1984) "Conversation and communication" Language and Communication, 4, pp. 71-83.

McGregor, G. (1985) "Utterance interpretation and the role of the analyst", Language and Speech, 28,1, pp. 1-28.

McKinney, J. C. and E. Tiryakian (eds.) (1970)***Theoretical Sociology*** . New York : Appleton.

McKnight, A. (1976) Large Group Conversation : Problems for Discourse Analysis. Unpublished M. A. Thesis, Lancaster : University of Lancaster.

McTear, M. (1977) "Starting to talk ; how preschool children initiate conversational exchanges", Unpublished mimeo, School of Communication Studies : Northern Ireland Polytechnic.

Medis, R. (1975) *Statistical Book for Non-Statisticians.* London : McGraw Hill.

Meijs, W. and J. Aarts (1986) *Corpus Linguistics II.* Amsterdam : Rodopi.

Merrit, M. (1976) "On Questions following questions in service encounters", Language in Society, 5, pp.

Merton, R. K., M. Fiske and P.L. Kendall (1956) *The Focus Interview.* Glencoe, Illinois : Free Press.

Meyer, B. F. (1975) *The Organisation of Prose and its Effects on Memory* . Amsterdam : North Holland.

Meyer, B. F. (1977) "What is remembered from Prose : A function of passage structure. In R. Freedle, *Discourse Production and Comprehension*. Norwood, New Jersey: Ablex.

Milroy, L. (1980) *Language and Social Network* . Oxford : Basil Blackwell.

Milroy, L. (1984) "Comprehension and context : successful communication and communication breakdown" in P. Trudgill (ed.) *Applied Sociolinguistics.* London : Academic Press, pp. 7-32.

Milroy, L. (1985) "Questions and Answers ; an analysis of the discourse structure of interviews" in J. Harris, D. Little, and D. Singleton, Proceedings of the First Symposium on Hiberno-English at Trinity College. Dublin. pp. 49-73.

Milroy, L. (1987) *Observing and analysing Natural Language* . Oxford Basel Blackwell.

Minsky, M. (1975) "A framework of representing knowledge", in P. Winston (ed.) *The Psychology of Computer Vision.* New York : McGraw Hill, pp. 211-268.

Montgomery, M. M. (1976) "The Structure of Monologue", English Language Research, Birmingham : Birmingham University.

Morgan, J. L. (1979) "Toward a rational model of discourse comprehension", American Journal of Computational Linguistics, Fische 79, pp. 109-114.

Murray, S. O. (1985) "Toward a model of members' methods for recognizing interruptions", **Language in Society** 14, pp. 31-40.

Myers, T. (1979) "Verbal and nonverbal interactivity" in T. Myers (ed.) ***The Development of Conversation and Discourse.*** Edinburgh : Edinburgh University Press, pp. 1-44.

Myers, T. (ed.) (1979) ***The Development of Conversation and Discourse*** . Edinburgh : Edinburgh University Press.

Naucler, K. (1982) "Spoken and written language : relationship between linguistic means of expression" in **Working Papers**, 24, Linguistics and Phonetics, Sweden, Lund University, pp. 49-58.

Norman D. A. and D.E. Rumelhart (1975) ***Explorations in Cognition*** . San Francisco : Freeman.

Nwogu, K. (1989) **Discourse Variation in Medical Texts : Schema, Theme, and Cohesion in Professional and Journalistic Accounts.** Unpublished Ph.D Thesis, Birmingham : Aston University.

Nwogu, K. and T. Bloor (1989) "Thematic progression patterns in professional and popularised medical texts", a paper presented at the 16th Systemic Congress at Helsinki, Finland, 12-15 June (1989).

Ochs, E. (1979) "Planned and unplanned discourse" in T. Givon (ed.) ***Syntax and Semantics, 12 : Discourse and Syntax*** . New York : Academic Press, pp. 51:80.

O'Donnell, R. C. (1974) "Syntactic differences between speech and writing", **American Speech,** 49, pp. 102-110.

Opitz, K. (1983) "On dictionaries for special registers", in R. R. K. Hartmann, ***Lexicography : Principles and Practices*** . London : Academic Press, pp. 163-180.

Orestrȍm, B. (1983) ***Turn-Taking in English Conversation*** . Lund, CWK Gleerup.

Ȍstman, O. (ed.) ***Cohesion and Semantics*** . Abo, Finland : Abo Akademi Foundation.

Palmer, F. R. (1971) ***Grammar*** . Harmondsworth, Middlesex : Penguin.

Palmer, F. R. (1976) ***Semantics*** . Cambridge : Cambridge university Press.

Parret, H., M. Sbisa and J. Verschueren (eds.) (1981) *Possibilities and Limitations of Pragmatics : Proceedings of the Conference on Pragmatics at Urbino* July 8-14 (1979) Amsterdam : Benjamins.

Pawley, A. and F. H. Syder (1975) "Sentence formation in spontaneous speech", New Zealand Speech Therapists' Journal 30/2, pp. 2-11.

Pawley, A. and F. H. Syder (1983) "Two puzzles for linguistic theory, native like selection and native like fluency" in J. C. Richards and R. W. Schmidt *Language and Communication* . London : Longman, pp. 191-226.

Pawley, A. and F. H. Syder (1983) "Natural selection in syntax" Journal of Pragmatics, 7, pp. 551-579.

Pearce, R. D. (1973) The Structure of Discourse in Broadcast Interview. Unpublished M. A. Thesis, Birmingham : University of Birmingham.

Perren G. E. and J. L. M. Trim (eds.) (1971) Applications of Linguistics : Selected Papers of the Second International Congress of Applied Linguistics. Cambridge , 1969, : Cambridge University Press.

Polanyi, L. (1978) "False starts can be true", Berkeley Linguistics Society 4, pp. 628-639.

Polanyi, L. and J. Scha (1983) "The syntax of discourse" TEXT 3(3), pp. 201-209.

Poole, M. E . and T. W. Field (1976) "A comparison of oral and written code elaboration", Language and Speech, 19, pp. 304-312.

Prince, E. F. (1978) "Discourse analysis in the framework of Zellig Harris" in W. U. Dressler (ed.) *Current Trends in Text Linguistics* . Berlin : Walter de Gruyter, pp. 191-211.

Prince, E. F. (1981) "Toward a taxonomy of given-new information" in P. Cole (ed.) *Radical Pragmatics* . New York : Academic Press, pp. 223-254.

Psathas, G. (ed.) (1979) *Everyday Language : Studies in Ethnomethodology* . New York : Irvington.

Psathas, G. (1986) "The organization of direction in interaction" WORD, 37,1-2, pp. 83-91.

Quirk, R., S. Greenbaum, G. Leech and J. Svartvik (1972) *A Grammar of Contemporary English.* London : Longman.

Quirk, R., S. Greenbaum, G. Leech and J. Svartvik (1985) *A Comprehensive Grammar of the English Language* . London : Longman.

Richards, J. C. (1980) "Conversation" TESOL, XIV,4, pp. 413-432.

Richards, J. C. and R. W. Schmidt (1983) "Conversation analysis" in J. C. Richards and R. W. Schmidt (eds.) *Language and Communication* . London : Longman, pp 117-154.

Rimmer, Sh. E. (1984) "The variable nature of narrative" in Midland Association for Linguistic Studies Journal, 9, pp. 1-24.

Rimmer, Sh. E. (1988) *Sociolinguistic Variability in Oral Narrative.* Unpublished Ph.D. Thesis , Birmingham : Aston University.

Robins, R. H. *General Linguistics : an Introductory Survey* . London : Longman.

Robson, C. (1983) *Experiment, Design and Statistics in Psychology.* London: Penguin Books Ltd.

Rumelhart, D. E. (1975) "Notes for a schema in stories" in D. Bobrow and A. Collins, *Representation and Understanding : studies in cognitive Science* . New York : Academic Press, pp. 372-272.

Rumelhart, D. E. (1977) "Understanding and summarizing brief stories", in D. Laberge and S. J. Samuels (eds.) *Basic Processes in Reading : Perception and Comprehension* . Hillsdale, New Jersey : Lawrence Erlbaum, pp. 265-304.

Rumelhart, D. E. (1984) "Understanding understanding" in J.Flood (ed.) *Understanding Reading Comprehension : Cognition, Language, and the Structure of Prose* Newark, Delware : International Reading Association, pp. 1-21.

Rumelhart, D. E., P. H. Lindsay and D. A. Norman (1972) "A process for long-term memory" in E. Tulving and W. Donaldson (eds.) *Organization of Memory* . New York : Academic Press, pp. 198-246.

Rumelhart, D. E. and A. Ortony (1977) The representation of knowledge in memory", in P.V. Anderson, R. Spiro and W. Montague (eds.) *Schooling and the Acquisition of knowledge* . Hillsdale, New Jersey: Lawrence Erlbaum, pp. 98-135.

Rumelhart, D. E. and D. A. Norman (1987) "Representation of knowledge", in A.M. Aitkenhead, and J.M. Slack *Issues in Cognitive Modelling,* London : Lawrence Erlbaum, The Open University, pp. 13-81.

Rutherford, H. C. (1965) *Syntax Analysis by Computer Based on Word Order and Segmentation.* Cambridge : Cambridge Language Research Unit.

Sacks, H. (1972) "An initial investigation of the usability of conversational data for doing Sociology" in D. Sudnow (ed.) *Studies in Social Interaction.* New York : Free Press, pp. 31-74.

Sacks, H., E. A. Schegloff and G. Jefferson (1974) "A simplest systematics for the organization of turn-taking in conversation", **Language** 50,4, pp. 696-735.

Sacks, H. and E. A. Schegloff (1979) "Two preferences in the organization of reference to persons in conversation and their interaction" in G. Psathas, *Everyday Language : Studies in Ethnomethodology.* New York : Irvington, pp. 15-21.

Sadock, J. M. (1974) *Toward a Linguistic Theory of Speech Acts* New York : Academic Press.

Sag, I. and J. Hankamer (1977) "Syntactically vs pragmatically controlled anaphora" in R. W. Fasold and R. W. Shuy (eds.) *Studies in Language Variation* Washington : Georgetown University Press, pp. 120-135.

St. Clair, R. N. (1980) "The Context of Language" in R. N. St.Clair and H. Giles (eds.) *The Social and Psychological Contexts of Language.* New Jersey: Lawrence Erlbaum, pp. 11-36.

St. Clair, R. N. and H. Giles (1980) "Introduction" of R. N. St. Clair and H. Giles *The Social and Psychological Contexts of Language.* New Jersey : Lawrence Erlbaum pp. 1-11.

St. Clair, R. N. and H. Giles (1980) (eds.) *The Social Psychological Contexts of Language* . New Jersey : Lawrence Erlbaum.

Sanford A. J. and S. C. Garrod (1981) *Understanding Written Language* . Chichester : Wiley.

de Saussure, F. (1916) *Cours de linguistique general* . Lausanne : Payot. Translated (1959) as *Course in General Linguistics* by W. Baskin, London : Peter Owen LTD.

Schank, R. C. (1975) "The structure of episodes in memory" in D. Bobrow and A. Collins (eds.) *Representation and Understanding.* New York : Academic, pp. 237-272.

Schank, R. C. and R. Abelson (1977) *Scripts, Plans, Goals and Understanding* . Hillsdale, New Jersey: Lawrence Erlbaum.

Schank, R. C. and B. Nash-Webber (eds.) (1975)*Theoretical Issues in Natural Language Processing.* Proceedings of a conference at the Massachusetts Institute of Technology, June 1975.

Schegloff, E. A. (1968) "Sequencing in conversational openings", American Anthropologist, LXX:61. pp. 346-380.

Schegloff, E. A.(1971) "Notes on a conversational practice : formulating place" in P. P. Giglioli (ed.) *Language and Social Contexts.* Hammondsworth, Middlesex : Penguin, pp. 95:135.

Schegloff, E. A.(1972) "Notes on conversational practice : formulating place" in D. Sudnow (ed.) *Studies in Social Interaction* . New York : The Free Press.

Schegloff, E. A.(1978) "On some questions and ambiguities in conversation" in W. M. Dressler (ed.) *Current Trends in Text Linguistics* . Berlin : Walter de Gruyter, pp. 81-102.

Schegloff, E. A. (1979) "The relevance of repair to syntax-for-conversation" in T. Givon, *Syntax and Semantics 12 : Discourse and Syntax* . New York : Academic Press, pp. 261-288.

Schegloff, E. A. and H. Sacks (1973) "Opening up closings", Semiotica 7, 4, pp. 289-327.

Schenkein, J. (1978) "Identity negotiations in conversation" in J. Schenkein (ed.) *Studies in the Conversational Interaction.* New York : Academic Press, pp. 57-78.

Schenkein, J. (1978) (ed.) *Studies in the Conversational Interaction.* New York : Academic Press.

Schiffrin, D. (1987) *Discourse Markers* . Cambridge : Cambridge University Press.

Schiffrin, D. (1988) "Conversation analysis" in F.J. Newmeyer (ed.) *Linguistics : The Cambridge Survey Vol. II. Linguistic Theory : Extensions and Implications.* Cambridge : Cambridge University Press, pp. 251-276.

Schmidt, S. J. (1978) "Some problems of communicative text theories" in W. U. Dressler (ed.) *Current Trends in Text Linguistics* . Berlin : Walter de Gruyter, pp. 47-60.

Searle, J. (1969) *Speech Acts* Cambridge : Cambridge University Press.

Searle, J. (1972) "What is a speech act ?" in P. P. Giglioli (ed.) *Language and Social Context* . Harmondsworth, Middlesex : Penguin, pp. 136-154.

Searle J. (1975) "Indirect speech acts" in P. Cole and J. L. Morgan (eds.) *Syntax and Semantics 3 : Speech Acts* New York : Academic Press, pp. 59-82.

Seuren, P. (1975) "Referential constraints on lexical items" in E. L. Keenan (ed.) *Formal Semantics of Natural Language* . Cambridge : Cambridge University Press, pp. 84-89.

Silverman, D. (1971) "Interview talk : bringing off a research instrument", Sociology 7, pp. 31-48.

Sinclair, J. McH. (1966a) "Linguistics in colleges of education", Dudley Journal of Education, pp. 17-25.

Sinclair, J. McH. (1966b) "Beginning the study of lexis" in C.B. Bazell et al (eds.) *In Memory of J.R. Firth.* London : Longman, pp. 410-430.

Sinclair, J. McH. (1979) "Discourse in relation to language structure and semiotics" in S. Greenbaum, G. Leech and J. Svartvik (eds.) *Studies in English linguistics for Randolph Quirk* . London : Longman, pp. 110-125.

Sinclair, J. McH. (1984) "Naturalness in language" in J. Aarts and W. Meijs (eds.) *Corpus Linguistics* . Holland : Rodopi, pp. 203-210.

Sinclair, J. McH. and R. M. Coulthard (1975) *Towards and Analysis of Discourse* . London : Oxford University Press.

Sperber, D. and D. Wilson (1986) *Relevance : Communication and Cognition* . Oxford : Basil Blackwell.

Speier, Matthew (1972) "Some conversational problems for interactional analysis" in D. Sudnow (ed.) *Studies in Social Interaction* . New York : The Free Press, pp. 397-427.

Stenning, K. (1978) "Anaphora as an approach to Pragmatics" in M. Halle, J. Bresnan and G. A. Miller (eds.) *Linguistic Theory and Psychological Reality* Mass.: The MIT Press, pp. 162-200.

Stenström, A-B. (1982) "Feedback" in N. E. Enkvist (ed.) *Impromtu Speech: A Symposium* . Abo : Abo Akademi, pp. 319-340.

Stenström, A-B. (1984a) *Questions and Responses in English Conversation.* Lund : CWK Gleerup.

Stenström, A-B. (1984b) "Discourse tags" in J. Aarts and W. Meijs (eds.) *Corpus Linguistics.* Holland : Rodopi, pp. 65-82.

Stenström, A-B. (1986a) "What does really really do ? : Strategies in speech and writing", in G. Tottie and I. Backlund (eds.) *English in Speech and Writing : A Symposium.* Uppsala : Acta Universitatis Upsaliensis, pp. 149-204.

Stenström, A-B. (1986b) "A Study of pauses as demarcators in discourse and syntax" in W. Meijs and J. Aarts (eds.) *Corpus Linguistics II.* Amsterdam : Rodopi.

Stenström, A-B. (1987) "Carry-on Signals in English conversation", in W. Meijs (ed.) *Linguistics and Beyond* . Amsterdam : Rodopi, pp. 87-120.

Streeck, J. (1980) "Speech acts in interaction ; a critique of Searle" Discourse Processes 3, pp. 133-158.

Stubbs, M. (1980) *Language and Literacy.* London : Routledge and Kegan Paul.

Stubbs, M. (1983) *Discourse Analysis ; The Sociolinguistic Analysis of Natural Language.* Oxford : Basil Blackwell Ltd.

Stubbs, M. (1984) "Applied discourse analysis and educational linguistics" in P. Trudgill (ed.) *Applied Sociolinguistics* London : Academic Press, pp. 203-244.

Stubbs, M. (1986) "Lexical Density : A computational technique and some findings", in M. Coulthard (ed.) *Talking about Text* . Birmingham : University of Birmingham, English Language Research, pp. 27-48.

Sudnow, D. (ed.) (1972) *Studies in Social Interaction* . New York : The Free Press.

Svartvik, J. (1979) "Well in conversation" in S. Greenbaum, G. Leech, and J. Svartvik (eds.) *Studies in English Linguistics for Randolph Quirk.* London Longman, pp. 167-178.

Svartvik, J. (1982) "The segmentation of impromptu speech" in N. E. Enkvist (ed.) *Impromtu Speech : A Symposium* . Abo : Abo Akademi, pp. 131-146.

Svartvik, J. and R. Quirk (eds.) *A Corpus of English Conversation* . Lund Studies in English 58. Lund : Liber/Gleerup.

Svartvik, J. and A-B. Stenström (1985) "Words, words, words : the rest is silence ?" in S. Backman and G. Kjellmer (eds.) *Papers on Language and Literature* Sweden : Acta Universitatis, pp. 342-353.

Swales, J. (1981) Aspects of Article Introductions, Research Reports No. 1., Birmingham : Aston University, Language Studies Unit.

Tannen, D. (1982) "The oral-literate continuum in discourse" in D. Tannen (ed.) *Spoken and Written Language : Exploring Orality and Literacy.* Norwood, New Jersey : Ablex, pp. 1-17.

Tannen, D. (ed.) (1982) *Spoken and Written Language : Exploring Orality and Literacy.* Norwood, New Jersey : Ablex.

Tannen, D. (1984a) *Conversational Style : Analyzing Talk among Friends* , Norwood, New Jersey : Ablex.

Tannen D. (ed.) (1984b) *Coherence in Spoken and Written Discourse* . Norwood, New Jersey : Ablex.

Taylor, B. P. and N. Wolfson (1978) "Breaking down the free conversation myth" TESOL, 12,1, pp. 31-39.

Thavenius C. (1982) *Referential Expressions in English Conversation* . Lund Studies in English (eds.) C. Schaar and J. Svartvik 64, CWK : Gleerup.

Tottie, G. and I. Backlund (eds.) (1986) *English in Speech and Writing : A Symposium.* Uppsala : Acta Universitatis Upsaliensis.

Trudgill, P. (1974) *Sociolinguistics : An Introduction to Language and Society* . Harmondsworth, Middlesex : Penguin.

Trudgill, P. (ed.) (1978) *Sociolinguistic Patterns in British English.* London : Edward Ernold.

Trudgill, P. (ed.) (1984) *Applied Sociolinguistics* . London : Academic Press.

Turner, R. (1972) "Some formal properties of therapy talk" in D. Sudnow (ed.) Sudnow, D. (ed.) (1972) *Studies in Social Interaction* . New York : The Free Press, pp. 367-397.

Ure, J. (1971) "Lexical density and register differentiation" in G. E. Parren and J. L. M. Trim (eds.) *Applications of Linguistics.* Cambridge : Cambridge University Press, pp. 443-452.

Ure, J. (1982) "Introduction : approaches to the study of register range" International Journal of the Sociology of Language, 35, pp. 5-23.

Tulving, E. and W. Donaldson (eds.) (1972) *Organization of Memory.* London : Academic Press.

Vachek, J. (1975) (ed.) *Valem Mathesius : A Functional Analysis of Present Day English on a General Linguistic Basis*. The Hague : Mouton.

Wadsworth, M. and D. Robinson (eds.) *Studies in Everyday Medical Life.* London : Martin Robertson.

Wardaugh, D. (1985) *How Conversation Works*. Oxford : Basil Blackwell.

Waugh L. R. and C. H. van Schooneveld (eds.) *The Melody of Language.* Baltimore : University Park Press.

Wayne, D. (1971) Hesitation Phenomena in the Spontaneous Speech of Non-Native Speakers of English, Unpublished Ph.D. Thesis, Illinois : University of Illinois.

Webber, B. L. (1978a) A Formal Approach to Discourse Anaphora. Report No. 3761, Cambridge, Mass.: Bot, Beranek & Newman.

Webber, B. L. (1978b) "Jumping ahead of the speaker : On recognition from indefinite descriptions", Paper presented at the Sloan Workshop on Indefinite Reference. University of Massachusetts : Amcherst.

Webber, B. L. (1981) "Discourse model synthesis : Preliminaries to reference", in A. K. Joshi et al. (eds.) *Elements of Discourse Understanding.* Cambridge : Cambridge University Press, pp. 283-299.

Widdowson, H. G. (1973) An Applied Linguistic Approach to Discourse Analysis. Unpublished Ph.D. Thesis, Edinburgh, University of Edinburgh.

Widdowson, H. G.(1975) *Stylistics and The Teaching of Literature.* London: Longman.

Widdowson, H. G. (1979) *Explorations in Applied Linguistics.* Oxford : Oxford University Press.

Widdowson, H. G. (1984) *Explorations in Applied Linguistics 2* . Oxford : Oxford University Press.

Winter, E. O. (1982) *Towards a Contextual Grammar of English.* London : George Allen and Unwin.

Wolfram, W. and T. D. Wolfram (1977) "How come you asked how come?" in R. W. Fasold and R. W. Shuy (eds.) *Studies in Language Variation.* Washington : Georgetown University Press, pp. 237-254.

Wolfson, N. (1976) "Speech events and natural speech : some implications for sociolinguistic methodology" Language in Society 5, 189-209.

Wolfson, N. (1979) "The conversational historical present alternative" Language, 55,1, pp. 168-182.

Wolfson, N. (1983) "Rules of speaking" in J. C. Richards and R. W. Schmidt (eds.) *Language and Communication.* London : Longman. pp. 61-87.

Wood, A., P. Fletcher and A. Hughes (1986) *Statistics in Language Studies.* Cambridge : Cambridge University Press.

Wootton, A. J. (1981) "Conversation analysis" in P. French, and M. Maclure (eds.) *Adult Child Conversation.* London : Croom Helm, pp. 99-110.

Yngve,V. H. (1970) "On getting a word in edgewise" in Papers from the sixth regional meeting of the Chicago Linguistic Society. Chicago : Chicago Linguistic Society, pp. 567-578.

Yule, G. (1979) "Pragmatically controlled anaphora" Lingua, 49, pp. 127-135.

Yule, G. (1981) "New, current and displaced entity reference", Lingua, 55, pp. 41-52.

Zora, S. S. P (1982) "Notes on the predictive power of contrastive analysis", Adab Al-Rafidayn, Vol. 15, Mosul : College of Arts, Mosul University, Iraq, pp. 59-69.

Zora, S. S. P (1986) Pausing and Syntactic Complexity in Interview and Conversation . Unpublished Report, Birmingham : Aston University.

Zora, S. S. P and C. M. Johns-Lewis (1987) "Lexical density in interview and conversation", to appear in **York Papers in Linguistics,** 14, York : University of York.

Zora, S. S. P. and C. M. Johns-Lewis (1988) "Lexical Variability in conversational discourse", paper presented at BAAL 20 th Meeting held at Exeter, Exeter University, September 1988.

Zora, S. S. P. (Forthcoming) "Linguistic complexity in male and female conversational discourse : A case study in all male, all female and mixed groups".