

**Bangor University**

## **DOCTOR OF PHILOSOPHY**

### **A stylistic analysis of administrative English through a qualitative and quantitative investigation of government information leaflets**

Ward, R. J.

*Award date:*  
1988

*Awarding institution:*  
Bangor University

[Link to publication](#)

#### **General rights**

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal ?

#### **Take down policy**

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

A STYLISTIC ANALYSIS  
OF  
ADMINISTRATIVE ENGLISH  
THROUGH A QUALITATIVE  
AND QUANTITATIVE INVESTIGATION  
OF GOVERNMENT INFORMATION LEAFLETS

BY

R. J. WARD B. A.

In fulfilment of the requirements of the

DEGREE OF PHILOSOPHIAE DOCTOR

Department of Linguistics

Universty College of North Wales Bangor

1988

## ACKNOWLEDGEMENTS

This thesis could not have been submitted without the valuable help of my tutor Professor Alan Thomas of U.C.N.W. who supervised my work. I would also like to express my thanks to Mr. Philip Schofield of U.C.N.W. who provided unstinting help and guidance not only on the statistical work of Part 2 but also in printing the finished text.

## Abstract

Previous work in stylistics has concentrated a great deal on theory to the exclusion of practical investigation of styles. The practical investigations deal with Newspapers or Advertising copy. No thorough stylistic analysis has been made of Administrative English.

A qualitative analysis of Government Information leaflets, reveals that they are stylistically distinct at all linguistic levels, but that two different types of text emerge, closely linked to the means by which the reader is addressed either personally as 'you' (P.A.) or impersonally as for example in 'the claimant' (I.A.).

A subsequent quantitative analysis of a selection of the most prominent stylistic features of Government Information Leaflets and their comparison with the leaflets published by financial institutions reveals that whilst most of the variables chosen are stylistic, there is little evidence to assume a single Administrative variety. Checks on the relationship between supposed style categories and the individual texts assigned to them are shown by a Cluster Analysis to be very accurate.

Patterning of variables is revealed around 2 stylistic dimensions: Status and Modality. I.A. leaflets are distinguished from P.A. largely by Status variables. The leaflets of Financial Institutions group with P.A. texts. All three of these styles are grouped together by Modality Variables.



CONTENTS

	PAGES
<u>PART 1 - QUALITATIVE STUDY</u>	1-234
CH. 1 INTRODUCTION.....	1-60
1.1. A General Statement of the Problem.....	1
1.2. The Concept of Style.....	2-19
1.2.1. General.....	2
1.2.2. Subjective and Objective Views of Style....	2- 3
1.2.3. Style as an Addition.....	4- 5
1.2.4. Style as Meaning.....	5- 7
1.2.5. Style as Deviance from a Norm.....	7- 9
1.2.6. Style as Choice.....	9- 11
1.2.7. Style as Individualism, as a Sociological Phenomenon, and Style as Personality.....	12-14
1.2.8. Style as a Probabilistic Concept.....	15-16

1.2.9. Style as a Property of Texts.....	17
1.2.10 A Working Definition of Style.....	17-19
1.3. The Situation Concept - General.....	19-30
1.3.1. Crystal and Davy's (1969) Concept of 'Situation'.....	21-28
1.3.2. The Concept of Situation and Administrative English.....	28-30
1.4. Style Variables - General.....	30-60
1.4.1. Graphetics and Graphology.....	33-34
1.4.2. Grammar - Syntax.....	34-38
1.4.3. Morphology.....	39-46
1.4.4. Semantics.....	46-53
1.4.4.1. Reference.....	47-49
1.4.4.2. Strategic Semantics.....	49-50

1.4.4.3. The Higher Level Semantic Component.....	50-53
1.4.5. Measures of Periodicity and Complexity.....	53-58
1.4.6. Style Variables and AdE.....	58-60
CH.2 METHOD.....	61-72
2.1. General.....	61
2.2. Design and Procedure.....	61-64
2.3. Texts.....	64-65
2.4. Reference Works.....	65-72
CH.3 THE NOUN PHRASE.....	73-116
3.1. General.....	73
3.2. Noun Phrase Premodification in English.....	73-81
3.2.1. Predeterminer.....	74-75
3.2.2. Determiner.....	75-76

3.2.3. Postdeterminer.....	76
3.2.4. Adjective.....	76-78
3.2.5. Participle.....	78
3.2.6. -s genitive.....	78-79
3.2.7. Noun.....	79-80
3.2.8. Adverbs.....	80-81
3.2.8. Sentences.....	81
3.3. Premodification and GIL.....	81-96
3.3.1. Predeterminer.....	82-83
3.3.2. Determiner.....	83-84
3.3.3. Postdeterminer.....	84
3.3.4. Adjective.....	84-89
3.3.5. Participle.....	89-91

3.3.6. -s genitive.....	91-93
3.3.7. Noun.....	93-95
3.3.8. Remaining Categories.....	95
3.3.9. Summary.....	95-96
3.4. Noun Phrase Postmodification in English.....	96-103
3.4.1. Single Word Postmodifiers.....	97
3.4.2. Postmodification by Prepositional Phrase...	97-98
3.4.3. Clausal Postmodifiers.....	98-101
3.4.4. Multiple Postmodification.....	101-103
3.5. Postmodification in GIL.....	103-114
3.5.1. General.....	103-105
3.5.2. Word Postmodifiers.....	105
3.5.3. Prepositional Phrase Postmodifiers.....	105-107

3.5.4. Clausal Postmodifiers.....	107-111
3.5.5. Multiple Postmodification.....	111-113
3.5.6. Summary.....	113-114
CH.4 The Verb Phrase.....	117-171
4.1. General.....	117
4.2. The Verb Phrase in English.....	117-142
4.2.1. Tense in English.....	117-124
4.2.2. Aspect in English.....	124-129
4.2.3. Mood in English.....	129-142
4.3. The Verb Phrase in GIL Texts.....	142-
4.3.1. Tense in GIL.....	142-151
4.3.2. Aspect in GIL.....	151-152
4.3.3. Mood in GIL.....	153-169

4.4. Summary.....	169-171
CH.5 The Clause.....	172-206
5.1. General.....	172-173
5.2. The Clause in English.....	173-182
5.2.1. Subject.....	173-175
5.2.2. Verb.....	175
5.2.3. Object.....	175-176
5.2.4. Complement.....	176-177
5.2.5. Adverbial.....	177-178
5.2.6. Vocative.....	178
5.2.5. Linking and Apposition.....	178-179
5.2.8. Passive and Active Clauses.....	179-182
5.3. The Clause in GIL Texts.....	182-206

5.3.1. Subject.....	182-186
5.3.2. Verb.....	186-187
5.3.3. Object.....	187-189
5.3.4. Complement.....	189-190
5.3.5. Adverbial.....	191-195
5.3.6. Vocative.....	195
5.3.7. Linking and Apposition.....	195-196
5.3.8. Passive and Active Clauses.....	196-204
5.4. Summary.....	204-206
CH.6 The Sentence .....	206-225
6.1. General.....	206
6.2. The Sentence in GIL Texts.....	206-214
6.2.1. Major and Minor Sentences.....	209-210



6.2.2. Simple and Complex Sentences.....	210-211
6.2.3. Compound and Mixed Sentences.....	211-212
6.2.4. Adverbial.....	212-214
6.3. The Sentence in GIL.....	212-223
6.3.1. Major Minor Sentences in GIL.....	214-215
6.3.2. Simple/complex Sentences in GIL.....	215-217
6.3.3. Adverbial Clauses in GIL.....	217-222
6.3.4. Compound and Major Sentences in GIL.....	222-223
6.4. Summary.....	223-225
CH.7 Summary.....	226-
7.1. General.....	226-234
7.2. Summary of Observations on the NP in GIL.....	226-227
7.3. Summary of Observations on the VP in GIL.....	227-229

7.4. Summary of Observations on the Clause in GIL...230-232

7.5. Summary of Observations on the Sentence in GIL 232-234

PART 2 - QUANTITATIVE STUDY

CH.1 INTRODUCTION.....235-284

1.1. Rationale for Conducting Qualitative  
Stylistic Research.....235

1.1. General Statement of the Problem.....236-237

1.2 Further Notes on 'Difference and Similarity'237

1.3. Format for Reviewing Previous Studies.....237-238

1.4. Some Basic Concepts associated with  
Statistical Research.....238-241

1.4.1. Independent and Dependent Variables..238-239

1.4.2. Significance Testing and Levels of  
Significance.....239-241

1.4.3. Hypotheses, Null Hypotheses and Alternative Hypotheses.....	241
1.5. MARKWORTH AND BAKER (1980) - One Way ANOVA - Bi-Variate Study (comparing I.V. and D.V.).....	241-244
1.6. SANDELL (1977) - Pearson Correlation Coefficients - Bi-variate Study Comparing 2 D.V.'s.....	244-250
1.7. MARKWORTH AND BAKER (1980) - Discriminant Analysis - Multivariate Comparisons Taking Account of the I.V.....	250-256
1.8. BIBER (1985) - Factor Analysis - Comparing more than 2 Dependent Variables.....	256-263
1.9. McENTEGART and LE PAGE (1982) - Cluster Analysis Grouping Observations into 'Natural' Clusters..	263-267
1.10. General Summary of the Information Provided by the Statistical Techniques Reviewed.....	267-268
1.11. Basic Requirements of all the Research Tech- niques Described Above.....	268-274

1.11.1. Experimental v Observational Studies.....	268-270
1.11.2. Random Sampling.....	270-271
1.11.3. Underlying Population Distributions.....	271-274
1.12. Implications of the Foregoing for the Present Study.....	274
1.13. Choice of Statistical Techniques in the Present Study.....	274-277
1.14. Design of the Present Study.....	277-279
1.15. Hypotheses.....	279-284
1.15.1. Formal Hypotheses about Normality.....	282-284
CH.2 METHOD.....	284-304
2.1. Design.....	285-298
2.1.1. General Theoretical Aspects.....	285
2.1.2. Theoretical Aspects Specific to the Present Study.....	285-286

2.1.3. Practical Aspects.....	286-288
2.1.4. The Independent Variable.....	288-291
2.1.5. The Dependent Variable.....	291-298
2.2. The Texts.....	298-302
2.2.1. Sampling.....	298-302
2.3. Procedure.....	302-304
CH.3 RESULTS.....	305-380
3.1. The Kolmogorov-Smirnov Test.....	305-
3.1.1. Between Groups.....	305-307
3.1.2. Pooled Groups.....	307-309
3.2. Kruskal-Wallis Test.....	309-341
3.2.1. Pairs Tests - General.....	315
3.2.2. Pairs Test - Group 1 (P.A.) with Group 2 (I.A.).....	315-317

3.2.3. Pairs Test - Group 1 (P.A.) with Group 3	
(BBSIL).....	318-322
3.2.4. Pairs Test - Group 1 (P.A.) with Group 4	
(DMA).....	322-327
3.2.5. Pairs Test - Group 2 (I.A.) with Group 3	
(BBSIL).....	327-331
3.2.6. Pairs Test - Group 2 (I.A.) with Group 4	
(DMA).....	332-341
3.3. Spearman Correlation Co-efficients.....	341-348
3.4. Factor Analysis.....	348-356
3.5. Manual Analysis of Rank Order Based Distance	
Coefficients.....	356-360
3.6. Cluster Analysis.....	360-369
3.6.1. Initial Interpretation of Single Linkage	
Dendrogram.....	365
3.6.2. Initial Interpretation of Ward's Method	
Dendrogram.....	365-369

3.7. Factor Score Plot.....	369-372
3.8. Discriminant Function Analysis.....	372-380
CH.4 DISCUSSION.....	381-423
4.1. Summary of the Results & Evaluation of the Techniques.....	381-384
4.2. Comparison with Earlier Studies.....	384-417
4.2.1. Preliminary.....	384-386
4.2.2. Carroll's (1969) Study.....	386-390
4.2.3. Moerk's (1973) Study.....	391-394
4.2.4. Sandell's (1977) Study.....	394-399
4.2.5. Markworth and Baker's (1980) study.....	399-405
4.2.6. CLUSTER ANALYSES: a) McEntegart and Le Page's (1982) study & (b) Bailey (19 ..)	405-409
4.2.7. Biber's (1985) study.....	409-417

4.3. Theoretical Significance and Practical Implications of the Findings.....	417-423
CH.5 CONCLUSION.....	424-443
5.1. Final Summary of Findings.....	424-428
5.2. Shortcomings of the Present Study.....	429-440
5.2.1. Shortcomings of Part 1.....	430-433
5.2.2. General Design Type.....	433-434
5.2.3. Shortcomings in the Choice of Variables..	434-436
5.2.4. Shortcomings in the Choice of Texts.....	436-438
5.2.5. Shortcomings in the Use of Statistical Techniques.....	439-440
5.3. Suggestions for Further Research.....	440-443
BIBLIOGRAPHY.....	445-456



PART 1 - CHAPTER 1 - INTRODUCTION

1. A General Statement of the Problem

Assuming that the situation in which an utterance is used can condition its form or style; is there a distinct style of language associated with Government administration? This is the general problem which this thesis sets out to solve.

Before however any work can be undertaken on tackling this problem, 4 preliminary tasks need to be undertaken. These are listed below:

1. A working definition of the word 'Style'.
2. A definition of the word 'situation' in its stylistic context and an analysis of what types of situation exist.
3. A discussion of what linguistic variables might be used to distinguish styles, and in considering 2 & 3:
4. An examination of previous research methods and their strengths and weaknesses.

These tasks will be dealt with in the order above.

## 1.2. The Concept of Style

### 1.2.1 General

One of the most notable facts about the style concept, even if this word is only applied to language (as is being done here), is the large number of different definitions it has been given by different theorists. Most of these definitions represent different aspects of style rather than competing theories, and any worker wishing to do practical analysis in the field of stylistic studies cannot really afford to ignore them. A review of the major definitions of the style concept will therefore now be made

### 1.2.2 Subjective and Objective Views of Style

A distinction needs first to be made between those theories of style which are objectively verifiable and those which are merely subjective and impressionistic. Enkvist (1964) p.11 sees the recognition of this dichotomy as an important first step in separating differing views of style.

An example of an impressionistic theory of style is that found in Murry (1922) who says (p13):

"A true style must, therefore, be unique, if we understand by the phrase 'a true style' a completely adequate expression in

language of a writer's mode of feeling"

Views like that of Murry which equate style with expressiveness can never be open to empirical linguistic investigation as 'expressiveness' would prove impossible to define objectively in Linguistic terms. Subjective views of style are not pursued further in this thesis because they could not be fruitfully researched. Objective theories of style fall into six major types according to Enkvist (ibid) p.12. These are:

1. Style as an Addition.
2. Style as Choice.
3. Style as Deviation from a norm
4. Style as Individualism
5. Style as a Sociological phenomenon.
6. Style as a property of texts

Sandell (1977) in his review of stylistic theories (Ch.2) also includes:

7. Style as Personality
8. Style as Meaning
9. Style as Statistical Probability

A review of these differing views of style will now be made, not necessarily following the order above.

### 1.2.3 Style as an Addition

One of the simplest ways to view style is as an optional addition to language, not essential, but something which increases the expressiveness of the language used or makes it more pleasing. A advocate of this view is Riffaterre (1959):

"Style is understood as an emphasis (expressive or aesthetic) added to the information conveyed by the linguistic structure, without alteration of meaning."

The problem with this type of theory as Enkvist (1964) p.12 makes clear is that only certain texts will possess style whilst others that are not expressive or aesthetically pleasing in some way are devoid of stylistic content. This view is of no use in a study like the present one which sets out to examine language in which expressiveness is not an important quality.

The second argument against a view of style as an addition is that it would be impossible for the linguist to capture such subjective notions as beauty or expressiveness in a linguistic description: These terms are not properties of the language used in a text but rather an interpretation of the reader's reaction to that language. Such terms are measurable by the 'Semantic Differential' technique, see Osgood (1964) where subjects are asked to rate language use on varying scales such as 'Good/Bad'. Such a technique however is

psychological rather than linguistic and outside the scope of this study.

#### 1.2.4 Style as Meaning

Riffaterre (above) was quoted as saying that style is non-semantic, ie. it does not alter meaning. Most views of style reviewed here adopt this stance. One theorist who does not however is Beardsley (1959):

"Where there is no difference in meaning at all, or else a gross difference, we do not say that there is a difference in style; where the difference in meaning is relatively subtle and is present along with some basic similarity on the primary level, we call the difference in meaning a difference in style."

On the surface Beardsley seems to be saying that Stylistics is no more than a sub-discipline of semantics: If style is meaning then the Stylistician has not got anything to study. In reality what Beardsley (ibid) seems to be concerned with is what Leech (1974) calls 'Stylistic meaning', the difference between:

"(1) They chucked a stone at the cops and did a bunk with the loot.

(2) After casting a stone at the police, they absconded with the money."

Leech (ibid) p.17

There is a great difference between this sort of meaning and the actual sense of a word, (what Leech refers to as 'Conceptual Meaning' (ibid) p.10., and Turner (1973) p.27 as 'Denotative Meaning'). Conceptual meaning is an essential factor in language structure in a way that stylistic meaning is not. Sense or Conceptual meaning rests on a system of contrasting relations in the same way that phonology does and describes the differences between items like 'boy' and 'girl'. Without this sort of logical system language would simply not work. The difference however between 'horse' and 'steed' for example has to do with factors other than logical relations, as both words may clearly refer to exactly the same animal.

If stylistics were simply to become part of the study of meaning as Beardsley's remarks suggest, then a very complex theory of meaning would be needed. A far more realistic view is to appeal to the notion of context in defining style as Enkvist (1964) p.20 suggests. Words like 'horse' and 'steed' above are likely to be used in very different situations. According to Enkvist (ibid) p.35 it is the fact that these sort of variants are used in different contexts that results in apparent differences of meaning: Meaning has 'rubbed off'

onto these words from the context in which they are usually used. Turner (1973) p.28 adopts exactly the same position.

#### 1.2.5 Style as Deviance from a Norm

The most problematic view of style is perhaps that which views it as deviance from a norm. A writer who adopts this view of style is Ullman (1964) when he comments on:

"...the powerful stylistic implications of Racine's departures from the norm."

Ullman's point seems to be that language only has style if it is markedly different from what is to be expected. There are three serious objections to this view. The first is that under this definition only a very limited number of texts indeed would display style, since most language and particularly that of government administrators, studied here, is not deviant in Ullman's (ibid) sense of the word.

The second criticism is that what is deviant language when compared with the norm is the norm for the group that uses that style: Thus poetry might be said to have style because it is 'deviant', but poetic language is the norm for a poet writing poetry; any use of 'normal language' could be considered 'deviant'.

The third and most fundamental problem is that of indentifying the 'norm'. If the norm is some sort of 'average' text then this would have to include all possible texts in the English language, including any which one might want to class as deviant. This is clearly very unsatisfactory as Enkvist (1964) p.23 Note 3 makes clear.

Enkvist (ibid) p.24 finds it acceptable to consider the norm as only part of the language e.g. "third leaders in the Times during 1960", and to set individual texts in such contexts. This seems a tenable view but only if the stylistician is interested in seeing how a particular Times leader in 1960 relates to the remainder of the set of leaders, ie. to make individual text analyses. This definition of style does not cover the work of the stylistician wishing to describe the norm itself. Enkvist (ibid) anticipates this problem and suggests that it might be possible to keep the notion of 'norm' without the concept of deviation, and concentrate on defining the different norms.

This latter suggestion of Enkvist's (ibid) is helpful but the use of the word 'norm' no longer seems very appropriate when robbed of the idea of 'deviation'. Turner (1973) p.16 uses the word 'scheme' to describe a very similar concept. Turner compares (ibid) the work of the grammarian and the stylistician by showing that whilst the grammarian is often only interested in describing one particular 'idealised' scheme, the stylistician is interested in comparing schemes or studying the variation possible within a given scheme.



It would be premature to make a decision at this point concerning the suitability of either Enkvist's (ibid) or Turner's views to the work of this study. The issue of norms and of deviance will be taken up again later in the formulation of a working theory of style.

#### 1.2.6 Style as Choice

The notion of style as choice is summed up by Ullman (1964) in the following:

"Intimately connected with expressiveness is another key-concept of stylistics, the idea of choice: the possibility of choosing between two or more alternatives-'stylistic variants', as they have been called - which mean the same thing but are not put in the same way."

The most obvious objection to such a definition is that choice is something that goes on in the brain of the speaker or writer and is not evident as such in the resulting text or utterance. All that the would be researcher has to work on is the finished text i.e. the final result of the choice This provides no evidence of what the nature of the actual choice was, since by definition a choice must be between two or more items.

A second argument against this approach is one that has been used

already when discussing style as meaning. Linguistic choice according to Enkvist (1964) p.17 can be of three possible types: grammatical (eg. the choice of a plural verb with a plural subject), non-stylistic (eg. the choice between 'boy' or 'girl' in the utterance 'The baby is .... '), and stylistic (eg. the choice between 'bloke', 'chap' and 'fellow' in 'He's a decent enough .....').

Grammatical choice is not choice in the sense used by Ullman (above) since a plural subject conditions the selection of a plural verb if the utterance is to be grammatically acceptable. The two remaining types of choice are those which relate to the concepts of denotative and connotative meaning respectively (considered earlier under 1.2.4). This view of Style as choice suffers consequently from the same drawbacks as the theory of Style as meaning: It is very difficult to say whether a group of linguistic items (eg. geezer, buffer, character, bloke, fellow, chap, man) may all refer to the same person, (ie. have the same semantic referent) or not. Individual speakers are likely to differ in their opinions making the clearcut distinction between denotative and connotative meaning a fiction. The Stylistician is faced with the impossible task in this event of trying to untangle his area of work from that of the semanticist.

Perhaps the most fundamental criticism of style as choice is that it places too much reliance on the conscious effort of the speaker or writer in making choices. Very seldom it seems do language users

deliberately choose linguistic features for stylistic effect, except perhaps when telling a joke and assuming the voice of someone else, or in literary language. In many everyday language situations there is very little choice, as Crystal and Davy (1969) p.5 point out:

"Clearly there are appropriate linguistic 'manners' for the different types of situation in which language is used, which we are expected (and usually want) to show."

The 'choice' here is between speaking appropriately or not so. The negative element in this choice is rarely used, to be linguistically 'unmannered' generally results in unfavourable reactions among other participants in any language situation.

It is important therefore not confuse the the ability to choose which is an overt ability, with a sense of appropriateness which is covert. Style as choice would appear at best to account for only a small part of style; that which involves the choice of language for special effect. It is in no way a 'definition of style'.

### 1.2.7 Style as Individualism, as a Sociological Phenomenon, and Style as Personality

An example of a view of style as the language of an individual is summed up by Williams (1970) p.2:

"May there not be "fingerprints" in writing, of which the author, and most of his critics, are quite unconscious, but which could be discovered by some new approach..."

Williams devotes the rest of his book to the study of exactly such a method. This definition of style assumes that style is the result of individual language variation. On its own such a view is entirely unsatisfactory, since it ignores the fact that the purpose of some language styles is to ensure anonymity, as in the impersonal language of the modern scientific report. Williams' definition would likewise exclude the printed varieties of administrative English.

The idea that each individual just has one unique style is challenged by Labov (1970) who through practical research techniques shows that one individual may wield a variety of different styles, though the style inventory of each individual may not be identical, and the features which mark each style not at identical frequencies for each language user:

"As far as we can see, there are no single-style speakers. Some

informants show a wider range of style shifting than others, but every speaker we have encountered shows a shift of some linguistic variables as the social context and topic change."

Labov's statement gives a definition of style which is not only not confined to individual speakers, but is tied to sociological factors. Elicitation of different styles from Labov's informants was done by the use of a variety of contexts intended to cover a gradation from casual to formal speaking styles.

Turner (1973) p.24 describes the difference between the two views, represented here by Labov and Williams as one of explanation and description. The work of researchers like Williams (ibid) has generally been to use statistical techniques to describe the important features making up an individual's language, no attempt is made to explain why the language of an individual possesses these features. Labov's approach is above all explanatory in that he seeks to explain the use of different styles through relating them to contexts.

In reality the explanatory approach is not an alternative to the descriptive one because any explanation of style must have a style to explain. Labov's research had also to include descriptive work on variables e.g. the incidence of post vocalic 'r' in individual's speech, but the important point was that descriptions were made of speech recorded in predetermined contexts (ie. some hypothesis about

the contexts determining the frequency of post vocalic 'r' had to be made first.

A theory very similar to that of style as individualism is style as personality. An example of this view is given by Hogben (1978):

"Language is an extremely diversified behaviour. Nevertheless individuals may use language in highly individual ways independently of the topics communicated. Psychotherapeutic clinicians think that linguistic style conveys a person's character."

What Hogben seems to have in mind here is not individual character but personality types under which a number of individuals may be grouped. As such this view forms a half way stage between Style as individualism and style as a social phenomenon, individuals being classified according to character or personality type. The rationale behind this is simple, the assumption is made that whilst all individuals vary, some are more alike than others and distinctive groupings can be found. However, such a view of style takes the researcher out of linguistics into psychology and personality theory. This approach is beyond the scope of this study and not entirely relevant since one would not expect personality type to affect Administrative English any more than Individuality.

### 1.2.8 Style as a Probabilistic Concept

One of the main proponents of this view is Dolezel (1969) who goes as far as to propose a theory of style based on probability:

"Style is a probabilistic concept. A probabilistic concept has 2 fundamental features: (1) In a probabilistic "world" the occurrence of phenomenon A is unequivocally predetermined by the existence of condition X."

Style's second fundamental feature according to Dolezel (ibid) is that whilst the probability of occurrence of a particular feature for a given population, (in the case of style this would be the total no. of texts belonging to a particular style) is stable the frequency of the same feature in a samples from the same population may vary randomly. For a given set of texts which the researcher believes to be drawn from this population, (and therefore belong to the same style) this means that the variation must be within the bounds of what can reasonably be expected to occur by chance (eg. using a probability against this of 5 in 100)

It seems rather too much to suppose that a probabilistic concept of style can support an actual theory. In no way for example can a view of style in terms of probability make explanatory statements as did Labov (1970) (see above), it can only provide the evidence needed to describe whether two texts might belong to the same style or not.

With respect to statistical techniques Turner (1973) p.26 has the following to say:

"It is precise to say that a man had a main meat dish in 85.71% of his dinners and that in 14.29% of his dinners the main dish was fish - it is more explanatory to point out that he ate fish on Fridays."

Style as a probabilistic concept is therefore only a useful tool to the stylistician in terms of enabling him to make explanatory statements, and even this tool can be dispensed with according to Sandell (1977) p.8:

"There is no genuine difference between an impression of style generated during reading and based on some unconscious counter of style indicators, and the impression of style gained from the inspection of a frequency table, as far as methods of style analysis are concerned that is."

In spite of this Sandell (1977) is full of statistical work, and there is some justification for this. Statistical measures whilst clearly not a substitute for the interpretive abilities of the stylistician, have the potential to provide valuable additional support to the findings of the stylistician.



### 1.2.9 Style as a Property of Texts

Enkvist (1964) p.26 completes his review of concepts of style in favouring the view of Hill (1958), (re-quoted here because of the non-availability of the original source):

'Professor Hill has defined stylistics as concerning:

"all those relations among linguistic entities which are statable, or may be statable, in terms of wider spans than those which fall within the limits of the sentence"...

This is a very simple view of style which does not really commit itself to any of the theories so far discussed (except perhaps tacitly to the probabilistic view of style). It enshrines the principle that style cannot be found in a single sentence, only in texts. For example a single sentence may contain an example of a passive verb, but only if it were to be discovered that a great many other sentences in the text from which the sample was taken were also passive, could it be said that the use of the passive was stylistic.

### 1.2.10 A Working Definition of Style.

In choosing a working definition of style suitable for the analysis of Administrative Language, concepts of style such as Personality and Individualism are of little use, since they cannot hope to cover

use of language by a specific group of people (e.g. Administrators). Style as an Addition is not pertinent to this study either, as one would not expect Administrative language to be expressive. What is needed is a definition of Style which acknowledges the social role of language, and at the same time the principle that style is a textual property which involves the manifestation of consistent variation within some sort of system of schemes (Turner, *ibid*) or norms (Enkvist, *ibid*).

The reason for this choice of requirements is that Administrative language is language produced by a clearly defined social group, and that in order to make any useful statements about the nature of this language, a corpus of texts representative of Administrative language needs to be located. The selection of the term 'Administrative Language' (AdE from now on) presumes that there must be other sorts of 'language' within English from which AdE differs.

A definition of style which seems to fit these requirements is that of Crystal (1980) p147, who describes Style as:

'...systematic, situationally distinctive, intra-language variation. By 'situation' I am referring to a sub-set of non-linguistic variables (such as occupation, status, purpose, regional or class background) which a native speaker can intuitively identify as accounting for a particular selection of features in a given (spoken or written) text.'

If the concept of Administrative English is now considered in this light it is obvious that there is at least one situation involved, that of occupation. But may there not be other situation types types involved as well? Status or Modality for example? Before any analysis is made of administrative texts it is essential to be certain along what situational dimensions the language of AdE itself might vary. If these varying dimensions are not properly accounted for any would be investigator runs the risk of missing important factors in the make up of an administrative style. It is to this problem of 'situation' that attention is turned in the next section.

#### 1.3.0 The Situation Concept - General

In section 2.8 the concept of situation was noted as being central to the definition of Style. A great deal of theoretical work was written on the concept of style and situation in the 1960's, though the term often used to describe stylistic variation was 'register' and not 'style' as is used here. Examples of this work are Catford (1965), Gregory (1966), Halliday et al (1964), and Strang (1962). All of these works with the exception of Gregory (1966) are general works with chapters on language variation.

Though much theoretical work was published on style very little in the way of practical analysis was conducted, (the one exception to this rule seems to be Leech (1966)). Halliday et al (1964), p.90,

admit this problem but yet feel that enough evidence exists for a three way categorisation of Register. Halliday et al (ibid) are careful to distinguish language according to use from language according to user, (ie. the distinction between Dialectal/Idiolectal language variation on the one hand and language use conditioned by situation on the other). Register according to Halliday et al (ibid), can be accounted for by 3 different situational conditions: the profession or activity being undertaken by the participants in the language event - field, the hierarchical and status relations between participants in the language event - Tenor, and the medium of expression, (speech or writing) - Mode.

The other works quoted; Gregory (1966), Strang (1962), Catford (1965), all employ very similar schemes to that of Halliday et al (ibid), and do not add to this tripartite division of the language situation. The terminology however is confusing and hard to follow since different authors use one and the same term to mean completely different things. 'Register' for example is used by Catford with a restricted meaning similar to that of Halliday et al's (1964) 'Field'. Strang (1962) makes a more extended use of the same term to cover status relations as well as the activity being engaged in. Gregory (1966) does not define the term in relation to his scheme but prefers to refer to use the term 'diatypic variation' instead of the global term 'register' which Halliday et al (1964) use. Gregory (1966) still uses the term 'register' and it is by no means clear what this refers to.

Because of these terminological insufficiencies, the lack of practical support and also the fact that three situational subcategories seem unlikely to account for all stylistic variation, (the aims and purposes of language use appear to be at least one missing feature) none of the works mentioned above really provides a sound theoretical basis for this research. A serious attempt has however been made to provide a uniform framework of stylistic analysis backed up by practical work in Crystal and Davy (1969). Though this work is no more than an introduction to Stylistic analysis it does provide theory which would appear to cover the major elements of the situation concept. It is therefore worth devoting some time to the examination of Crystal and Davy's discussion of the situation concept, and this is done below.

### 3.1 Crystal and Davy's (1969) Concept of 'Situation'.

One major difference between Crystal and Davy (1969) and earlier works is their attention to the nature of the relationship which exists between language and situation. A discussion of exactly how this relationship works is an essential preliminary to any stylistic research. Crystal and Davy (1969) p.62 are very careful to point out that a particular situation rarely guarantees, or conditions the occurrence of specific linguistic features in a one to one way, or even in a way that can be confidently predicted, as Dolezel (1969) seemed to imply above (p.15) when he proposed 'unequivocal' conditioning of style by situation:

"Style is a probabilistic concept. A probabilistic concept has 2 fundamental features: (1) In a probabilistic "world" the occurrence of phenomenon A is unequivocally predetermined by the existence of condition X."

Further evidence that this is not so is provided by Labov (see quotation from Labov on pages 11-12). There are two implications of this evidence for practical stylistic study; one is that the selection of a particular non-linguistic situation by the researcher (eg. washing the dishes) does not mean that the language used in that situation will necessarily be different stylistically from the language of conversation for example. The second implication would seem to be that any sort of statistical analysis in terms of probability would be doomed to failure.

The first of these implications has to be accepted: Reality could be infinitely subdivided into situations like 'washing up carried out by two teenage children on Sunday 25th August 1985', ie. one situation to every language event. To be useful to the researcher a theory of style must allow the grouping of language events into categories where the linguistic similarities between them far outweigh the differences. If situation conditions language style then some of the minutely divided situations like those above will have to be classified together as one super-situation where a great many common features are shared. How finely the situations are defined in an individual piece of research will ultimately depend on

whether two proposed situational categories appear to give rise to detectable stylistic variation. Finding one's way in this area is a matter of trial and error and personal intuition, as Enkvist (1964) points out.

The second implication is not an altogether valid one to draw: One to one relationships between situation and language are not a necessity for the quantitative measure of style, it is sufficient to prove that the variation in the frequency of a particular linguistic feature between texts from situation A and texts from situation B is not likely to be the result of chance factors in order to disprove the null hypothesis that there is no significant difference between the two groups.

This does not strictly allow for the setting up of a cause and effect relationship between the situation and the language style concerned, since as the situation and language event are merely observed and not set up and controlled by the linguist (as would be the case in a true experiment), other conditioning factors correlatable with the situation but not directly observed may be responsible for the differences measured. This is a necessary condition of observational research (see Plutchik (1983)). Crystal and Davy (1969) pp.64-83, unlike Halliday et al (1964) use a rather more elaborate description of situation, setting the major factors affecting style in the context of total language variation; both the relatively permanent features like Dialect, Temporal provenance, and idiolect, and the fundamental language distinctions

between speech/writing and monologue/dialogue.

The inclusion of these (for the stylistician) less interesting features are important background material for any stylistic analysis. The important point here is that not every situationally distinct aspect of language variation is a manifestation of style; some variations are the result of dialect, some the result of individuality etc. If one imagines a stylistician trying to investigate the language of farmers for example as opposed to doctors, he would be interested in language conditioned by occupation and not the various dialect varieties in which different informants would undoubtedly speak.

The three situational categories (Crystal and Davy (1969) use the term 'Dimensions of Situational Constraint') of major interest to the stylistician are those of Province, Status and Modality. Province covers the type of 'occupational or professional activity being engaged in' (ibid) p.71.

At first this category would seem to cause few problems, occupation and profession being particularly easy to describe and objectively assess. In the case of administration this includes all those employed in administrative work. However this is a very broad field, and it might be wise to separate public administration (ie. gov't admin') from private sector (ie. financial and commercial administration). It might then be possible to separate local from central government administration and banking and finance from



general commercial administration. Central Government Administration could be broken down by department and financial into separate companies, and so on.

Considering what has already been said in this section about Situation and its effect on Style, it is important to be wary of defining Province too finely. Intuitive selection is the only possible basis for fixing on the optimum level of analysis and is unavoidable as a first stage in stylistic research as Crystal (1971) p.152 observes. He points out that the only possible way of avoiding a subjective approach is to take data for analysis from the whole field. In the case of administration this would be a task too large and impractical for a small piece of research such as the present study. A decision on what to take as the exact province for study will be made at the end of section 3 along with decisions relating to the other stylistic dimensions yet to be discussed.

'Status' is a dimension which Crystal and Davy (1969) explain as accounting for:

"..variations in the relative social standing of the participants in any act of communication."

pp.73-74

However the authors find that clear definition of the nature of Status itself is elusive, (p.74) and that attempts to isolate scales

with discrete divisions (like the five point scale attempted for formality relations in Joos (1959)) are premature. Crystal and Davy (ibid ) do however note that status is a complex concept covering notions such as:

"..formality, informality, respect, politeness, deference, intimacy, kinship relations, business relations, and hierarchic relations in general."

Administrative English is often criticised because its language tries to put social distance between the administrator and the general public. A good example of such criticism is that made by Cutts and Maher (1980), when they assert:

"Much official writing seems deliberately pompous, archaic and frightening, as if to impress on the reader that he is bureaucracy's servant....High-sounding language has developed the power to hold people in awe. If they can't understand it, they feel somehow to blame. They say, "These officials must know what they are talking about, who are we to argue?"

At the same time there is some pressure on government departments to rid their language from these aspects when it is intended for public consumption. Pressure groups like the 'Plain English Campaign' have been lobbying the government for the reform of administrative language, and some government departments have themselves been keen

to issue handbooks and instructions to their staff with injunctions like the following:

"Public notices should be informal in style. Therefore for the majority of notices use the second person. If it is conversational and personal, readers are more likely to respond positively."

Forms and Notices Unit (HM Customs and Excise) (1983), sect 4.2

In an area as vast as that of government administration it seems unlikely that a change to more informal English will occur very quickly. The Government White Paper on official language Cmnd.8504 lends support to this view. The area of administrative English which involves written texts for public consumption, (basically forms and information leaflets) seems a promising area for the study of Status differences in that two different styles: an older formal one and a newer informal one (not necessarily having a discrete break between the two) might possibly exist.

'Modality', Crystal and Davy's (1969) third and final situational dimension, is concerned with the purpose of an utterance (p74 ff). It is a dimension independent of Status and Province, and tends to determine the physical form of the utterance. Crystal and Davy's example of this is the language of 'correspondence' being variously conveyed through the form of letters, postcards, notes, telegrams or

memos.

Clearly a province as wide as Administrative English could have equally large ranges of Modalities in both its spoken and written forms. For example written AdE could take the form of forms, notices, information leaflets, explanatory instructions accompanying forms, official warnings and reminders, written replies to enquiries and so on. Some selection needs to be made from among these modalities in order to make a manageable research topic.

### 1.3.2 The Concept of Situation and Administrative English.

It is now necessary to come to some decision concerning the sort of divisions of Administrative language likely to lead to fruitful study. This decision may of course need to be revised if the types of language chosen do not display stylistic independence.

Status seems intuitively to be an important area to investigate, and the most useful area to investigate it seems to be in language intended for the general public, because this is the area in which changes are reported to be taking place, (see page 27)

When Province differences e.g. the difference between private and public administration are considered, the private sector is something of an unknown quantity in the extent to which Status differences are present; critics of AdE concentrating more on the

output of Public administrative bodies (though not exclusively). Local Government has disadvantages for data gathering since just one or a selection of local authorities might constitute an unrepresentative sample, and to collect material from every local authority would be a very large task. Central Government with only a relatively small number of departments dealing direct with the general public, seems the best choice for research.

The main modalities through which written Central Government AdE is conveyed to the general public appear to be:

- 1) Questionnaire forms
- 2) Information leaflets
- 3) Written replies to individual enquiries
- 4) Written requests for information to individuals
- 5) Pre-printed notices/warnings/reminders.
- 6) Individually written warnings/reminders.

'Written' AdE includes material that is both individually composed (and usually typed (3,4 & 6)), and printed material not aimed at any particular individual. It can be seen that 3, 4 & 6 are the individually composed counterparts of 2, 1 & 5 respectively.

There are two main reasons for not using individually composed material as data. The first is that an extra situational dimension is introduced into the material, that of individuality, itself not

of particular interest to this study. The second reason is that individually composed texts are likely to be considered as confidential and therefore difficult to obtain. It is also possible that in being faced with a request for such material, individual government departments may be tempted to sift the correspondence in some way therefore invalidating the sample.

Of the printed types of AdE, only 2 above is likely to be readily available as many forms are mailed directly to individual members of the public in the course of an application to a particular department whereas information leaflets are freely available to members of the public in post offices and public libraries etc. Information leaflets then published by central government departments seem to be an appropriate area of study.

Having now discussed the notion of situation, the way in which situation concerns AdE, attention needs to be paid to the individual units of stylistic variation; the different linguistic features which may operate as style marking variables. The next section of this chapter is devoted to this problem.

#### 1.4.0 Style Variables - General

Virtually the whole of linguistic structure is available as a source of potential stylistic variation. Traditionally the main levels of linguistic structure have been classified under the broad headings

of; phonetics/phonology, grammar & lexis, and semantics. Grammar has been traditionally divided into 'morphology' the study of word typology and structure, and 'syntax'; the study of the way in which words can be structured into larger units, namely; the phrase, the clause and the sentence. Some authors eg. Halliday and Hasan (1976) have gone beyond the limits of the sentence and endeavoured to examine the ways that sentences may be linked to form texts.

Lexis is the study of vocabulary and unlike the study of morphology and syntax where choices involve closed sets (eg. only nouns may take plural inflections), lexical choices involve the selection from often very large sets, (Halliday et al, (1964) pp.21-22), for example the selection of words which could fill the blank space in:

He sat down on the .....

is a huge set including such diverse items as: grass, wall, chair, box, floor, bed etc. However the distinction between open and closed sets is not completely clear as Halliday et al (1964) p.22 point out. For example the number of prepositions which could replace 'on' in the example above are a relatively small set. This has lead those who need to set up practical schemes for stylistic analysis like Crystal and Davy (1969) p.58 to consider lexis alongside grammar and not as a level on its own. Within the pattern of lexical choices the stylistician might expect to find vocabulary differences between texts written in different styles and particular collocations of

words. For example Turner (1973), pp.173-174 notes that only in the language of stock market reports can tin 'suffer' and steel 'sag'

Because modern linguists have concerned themselves very much with spoken rather than written language, the level of linguistic structure which deals with the raw materials of communication, and their systematisation in individual languages; (phonetics and phonology) has been orientated towards the raw materials of speech, ie. sound. The companion areas of study relating to writing; graphetics, and graphology have been little investigated, but these areas too are areas of potential stylistic variation.

Very often, individual linguistic theories place all the levels so far mentioned in a strictly hierarchical arrangement, this is not held to be of great importance to stylistic study by Crystal & Davy (1969) p.15. On purely practical grounds the stylistician is interested in studying the levels of linguistic structure independently in order to discover stylistic effects. The position of a particular level is not therefore of great importance to him. This is not to say however that it is valueless to discuss the hierarchical nature of linguistic levels.

Although all the levels so far mentioned are potentially interesting to the stylistician some of them have been far more widely used than others, in particular lexis, along with variables that are very hard to fit into the traditionally defined grammatical categories. These



are variables like measures of periodicity, and of text difficulty. Sometimes the preference seems to have good reasoning behind it; differences in vocabulary can be expected to vary greatly between different texts. But also it may be that some variables are chosen because of the ease with which they can be measured, eg. reading tests, or because they are particularly amenable to statistical treatment, eg. noun adjective ratios, rather than necessarily being very good measures of stylistic variation. These criticisms are discussed in full later.

The different linguistic levels, the ways in which they have been used in stylistic research, and their possibilities for yielding reliable style variables in AdE are discussed below. Because it has been already established that this study will deal only with texts in the written medium there will be no discussion of phonetics or phonology.

#### 1.4.1 Graphetics and graphology

There are very few examples of researchers finding graphetic or graphological style markers. One of the reasons for this is perhaps that the study of writing systems is a very neglected area of language study anyway, (as Crystal and Davy (1969) p.16 observe). Crystal and Davy do make reference to graphologically defined style markers though, for example in their discussion of the language of legal documents, p.198:

"Thus legal English, at least from a graphetic and graphological point of view, has something in common with other varieties in which clear logical sequence is essential. As in technical handbooks, collections of instructions, and much Civil Service language, blocks of print are arranged so as to reveal the sections into which the content is organised, the relationships between them, and their relative importance."

Leech (1966) in his study of the language of advertising is another author who uses graphological variables to distinguish style, p.177, though he prefers the term 'orthography':

"Breaking the orthographic rules is a popular practice in trade names, where the purpose is to provide the product with a distinctive written symbolisation (which may also do duty as its trade-mark). In the names Brylcreem and Rice Krispies, the elements -creem and Krisp- are clearly derived, by 'misspelling', from the English words cream and crisp."

Whilst the latter type of style variable is unlikely to occur in AdE, the former, mentioned by Crystal and Davy (ibid) is perhaps more likely. Variables relating to the writing system may therefore have some small part to play in distinguishing AdE as a style.

#### 1.4.2 Grammar - Syntax

Syntactic categories are frequently observed to function as style variables in works like Crystal and Davy (1969), Leech (1966), and Sandell (1977) which investigate the effect of situational factors like province, status and modality on language style. Examples of such style variables are too numerous to mention, they cover all aspects of syntax from the noun phrase, through the clause to the sentence and the way in which sentences are linked. Leech (1966) devotes 3 separate chapters in his study (chapters 12, 13 and 14) to clause and group (noun and verb phrase) structure.

In analysing noun phrase structure, the main potentialities for stylistic differentiation are in the modification of the head noun. For example, is there a great deal of premodification or none at all, is postmodification preferred instead? Are postmodifying structures themselves postmodified? How often are postmodifiers clauses? Leech (1966) p.125 for example discovered that noun phrases in advertising English carry heavy and often complex premodification sequences like;

"They go for its fresh, wake up tang, its thick satisfying texture."

Verb phrase structure offers fewer possibilities for stylistic variation, the main possibilities for differentiation being in the use of tense, aspect, passive marking and the use of modal verbs.

Leech (1966) p.121 found that the verbal groups used in advertising English tended to be simple and marked mainly for present tense forms or imperative. Chiu (1973) p.65-66, studying verb phrases in Canadian administrative writing found that up to 16% of verb phrases were complex and almost half of the marked verb phrases were marked for passive voice. Markworth and Baker (1980), p.240, found that whilst noun phrase structure does not vary greatly between fiction and non-fiction styles verb phrase structure does; fictional styles make heavy use of the past tense and perfective aspect, but passive verbs, a characteristic of non-fictional styles are absent in fiction.

Clause structure like that of noun phrases is full of different possibilities; clauses must contain a subject and verb but may or may not contain a complement or an adverbial. Either subject or complement can be a simple or a complex noun phrase, the clause itself may be dependent on another clause structure or may function on its own and be coextensive with the sentence. Dependent clauses may be finite or non-finite, may function at either subject, complement or adverbial position in a sentence and so on.

Crystal and Davy (1969) make frequent mention of clause variables in their practical stylistic analyses; Legal documents make frequent use of conditional and concessive adverbial clauses p.203, and formal Religious English is marked by deviations from standard clause order, (p161).

Sentence structure offers possibilities for stylistic variation in the different sentence types which can be selected by a language user. Sentences may be simple, complex, co-ordinated or mixes of complex and co-ordinated structures. The individual elements of sentence structure; subject, verb, complement, adverbial can all be realised by different elements of structure, and sentences may vary in mood according to whether they are declarative, imperative or interrogative.

Some language styles like Advertising using only simple sentences show little of stylistic interest beyond clause structure, simple sentences being by definition composed of only one clause. Leech (1966) reports this as being true in his study, p.120. Crystal and Davy (1969) p.203 find on the other hand that sentences in legal documents very often have adverbials and tend therefore to be complex. Markworth and Baker (1970), p.244, report that 'Mood' tends to distinguish Newspaper Reportage and Popular Journals: Popular journals using far more imperative sentences and far fewer declarative sentences than News Reportage.

Intersentence linkage may take a variety of forms. Sentences may be simple, complex, co-ordinated or mixes of use of devices like anaphora (back reference), ellipsis, lexical repetition, conjunctions and so on. Halliday and Hasan (1976) have developed a system for scoring texts according to the types of linkage they use (they refer to individual instances of linkage as 'ties') and the

number of ties in a text. Whilst Halliday and Hasan (1976) do show textual analyses carried out with their system they do not go to the extent of claiming that the differences present are due to style.

Crystal and Davy (1969) who do cite intersentence linkage as a possible source of stylistic variation, pp.43-44, find interesting differences between different text types: legal documents, they report, tend to have very long self contained sentences and therefore little inter sentence linkage, p.201. In examining the English of two newspaper articles both on the same topic, the same authors find that inter sentence linkage appears to distinguish different newspapers.

In conclusion, syntactic structure carries a great many possibilities for stylistic differentiation, from phrase structure up to sentence formation. Any one language variety may be characterised at several different levels of syntactic structure, as the frequent reference to the language of legal documents shows above. Any stylistic analysis cannot therefore afford to concentrate solely on one particular aspect of syntactic structure (as does Chiu, see above) but must investigate all the syntactic possibilities for stylistic variation. The amount of potential for stylistic variation, particularly in clause and noun phrase structure means that an examination of the stylistic variation at the syntactic level in any style will be a large undertaking.

### 1.4.3 Morphology

Whilst Morphology deals both with word structure and with typology, by far the greatest amount of work in investigating styles has concentrated on the latter. Some small reference is made to word formation processes, eg. Leech (1969) p.141 commenting on the role of the prefix 'super-' in Advertising English.

Attention has perhaps been focused on word typology because different word classes are a relatively easy thing to count, but this often leads to results of somewhat uncertain significance, as is explained in more detail below

Studies investigating individuality or singularity (see definitions in sect 1.2.7 above) often make great use of word typology measures like the ratio of nouns to verbs, or of adjectives to nouns. An Example of this approach is the work of Williams (1970) p.143-144. He compares 1000 word samples from the works of Shakespeare, Macaulay and Osbert Sitwell and comes up with the following data:

	USAGES		
	Nouns	Adjectives	Adjectives per noun
Shakespeare	224	115	0.51
Macaulay	223	133	0.57
Sitwell	214	138	0.64

This small sample appears to show that Sitwell uses rather more adjectival qualification of nouns than either Shakespeare or Macaulay. There are several problems with this method; one is that with such short pieces of text the number of adjectives could be related to the content of the passage and have nothing whatsoever to do with the dimension of individuality, a second important question is that of whether the relatively small differences in the number of adjectives seen above are really a manifestation of different styles?

Counting the occurrences of different word classes eg. noun and verb is really just another way of assessing syntactic structures, in the case mentioned above this was adjectival modification in the noun phrase. A far more sensitive way of investigating such phenomena (though far more time consuming) would be to count the frequencies of different types of adjectival modification along varying parameters; whether premodification or postmodification was involved, whether the modifier was an adjective a noun adjunct or a gerund, how many modifiers were involved and so on.

Measures like the noun adjective ratio (presented on the previous page) are easy to perform and give an impressive looking index of style variability. However this one figure is a generalisation taken from an amount of individual data which are all lost in its calculation. To make this point clear imagine the case of two



styles, one of which always uses a constant amount of adjectival modification in each noun phrase, and another displaying the same noun adjective ratio, but which intermittently uses heavy modification alternated with no modification at all. Potential stylistic variation is masked by using such indexes as the noun adjective ratio based merely on word classes.

#### 1.4.4 Lexis

The ways in which stylisticians have investigated the differentiation of styles by lexical style variables fall into two broad categories:

- 1) Qualitative study where comment is made on unusual words, neologisms, archaisms, technical terms, and so on.
- 2) Quantitative study where vocabulary items are counted in some way

An example of the first type of approach is found in Leech (1966), (this author also uses a quantitative method too). Leech makes particular comment on the use of neologisms (p.178) which are observed by him to be frequent in advertising: Examples are 'peelability' (in reference to an orange), 'temptational', and 'Ricles are twicles as nicicles'.

There appears to be no formal theory behind the selection of items worthy of comment in Leech (*ibid*), the author simply notes what he feels intuitively to be particularly redolent of an advertising style. Unusual words, technical terms, and archaisms are likely to be revealed by such a treatment, although all these categories, including neologisms could be researched on a more formal basis with a good dictionary as a standard reference as to whether a word was archaic, technical, etc. This does not however seem to have been tried in stylistic work.

Quantitative approaches to lexical style appear to divide into two main types; those which simply look at how often individual lexical items are used (usually only single words are studied), and those that go beyond simple descriptions about word frequency to statements about the richness and dispersion of the vocabulary used in a given text.

The first type of study mentioned takes the form of drawing up a word frequency list, and it is so frequently used in stylistic work that a thorough consideration of its use needs to be made here. Leech (1966) pp.151-155, is an example of how this first type of approach can be carried out on a small scale. Leech draws up frequency lists of verbs and adjectives used in a sample of television advertising, and ranks the 20 most frequent in order of descending frequency. Leech appears to have some reservations about using this technique for examining Advertising English, because of

the fact that Advertising English is often very unusual and innovatory in its choice of words.

Leech restricts his analysis of frequency lists to general comments on what the data appears to reveal about advertising English, eg. the complete absence of adjectives with negative overtones. What Leech could have done but does not do, is to compare his list with other large lists eg. Thorndike and Lorge (1964) in order to give an idea of what stylistic differences are apparent in vocabulary inventory and word frequency. One study which does this is a much larger one (140,000 word sample) of verb phrases in Canadian Administrative Language (Chiu, 1973). Chiu shows that the 15 most frequent verbs (excluding BE and HAVE) in her sample come quite low on the word list compiled by Kucera and Francis (1967) (1,000,000 word sample from 15 different 'genres')

In a critique of word counts Engels (1968) shows that large 'general' word counts like Kucera and Francis (see above) fail to account for all of the words in samples of non-specialised journalistic material. Engels does this by taking 1000 word blocks from newspaper articles and finding out how many of the 1000 words are accounted for by the first 1000 words in West's (1953) list. The result varies between 68 and 78%. Successively larger chunks of West's list fail to account for many more of the words; a residue of words being left.

Lyne (1983) argues that these residues are often very indicative of the style of the passage from which they are taken and proposes a different type of frequency list determined on the criterion of what he terms 'registral value'. Assigning registral values to words in a sample is a statistical technique which demotes those items which would tend to rank high in any word count (generally function words) and promote those particularly indicative of the style under study (often indicative of Province).

The present state of the art as regards word counts and their use in stylistic study is therefore quite sophisticated. There is no doubt that significant results can be obtained by using this method, but in spite of the sophistication involved the results produced by Chiu (ibid) for example are rather predictable and anyone who has a reasonable familiarity with Administrative correspondence might expect verbs like attach, enclose, receive and provide (all very high on Chiu's list) to figure highly in any sample taken from this style. Moreover, these same verbs might well be expected in commercial and business correspondence as well.

This seems rather to indicate that vocabulary in Chiu's study was not a reliable measure of province but of the modality of formal letters. In spite of their widespread use in stylistic study therefore word counts need to be used with great care.

The method used to investigate the total vocabulary inventory used in a text generally involves counting the total number of word

types, the number of tokens, (number of different words) and comparing the tokens with the number of total types and also with their frequency in descending order. This is often displayed in the form of a line graph. From the ratio of tokens to total types an idea of the richness and diversity of the vocabulary used in the text can be obtained: for example a text which uses a small set of words repeatedly shows a very small total vocabulary whilst one with a large set of total words will involve very few repeats and be more diverse.

This method of investigation is criticised by Posner (1963), p.12ff on several grounds. The main problems being that it is too dependent on sample size and also that it ignores potentially large differences in the word inventories of different texts. The problem of dependency on sample size has been to a certain extent overcome by using alternative statistical methods (eg. that of Yule (1944) reported in Posner (ibid)), but the second problem is fundamental, and limits the application of this sort of measure to stylistic study.

Work on the richness and diversity of different styles does not appear to have been carried out outside the literary field, (Yule for example did his work on Macaulay and Bunyan) and therefore is something of an unknown quantity in the area of style defined by situation. It seems likely however that even if used it would mainly serve to confirm fairly obvious effects as with the use of word counts.

In relation to lexical differentiation of styles it can only be concluded that a great deal more work needs to be done. Much of the work of real interest to the stylistician does not seem to have been tackled at all, nor is there a theoretical framework in existence for this. Particularly of interest to the stylistician would be the investigation of the incidence of particular word collocations in different varieties, and the way in which different varieties use near synonyms.

Reference to synonyms takes the researcher into the hazy area of Semantics, an area which Leech (1966) p. 109 observes to have the potential for the most interesting stylistic observations of all if only the methods of analysis were sufficiently developed. The next section is devoted to a discussion of this subject.

#### 1.4.4 Semantics

Semantics is very similar to stylistics in that as Leech (1974) p. ix observes it suffers from a multiplicity of different theoretical approaches which are by no means clearly defined in relation to one another. Like stylistics semantics too appears to suffer from a great lack of practical analysis. This fact is an unfortunate one for the stylistician who like Leech (1966), quoted in the previous paragraph, would often like to explore the rich potential of semantics to reveal stylistic contrasts.

Investigative work on styles has therefore been largely concerned

with making notes on general points of semantic interest which are felt intuitively to mark the Styles under study. This work has tended to be concentrated on such easily accessible topics as particularity of reference. More adventurous treatments are those like Leech (1966) on 'strategic semantics', endeavouring to explain how language varieties like 'propaganda' convey their meaning, and also an appeal to 'a higher semantic component' to distinguish styles in Hasan (1973). The review of these various approaches will be dealt with in the order in which they have just been presented.

#### 1.4.4.1 Reference.

Reference, also referred to as conceptual, cognitive, denotative meaning and sense, is according to Leech (1974) essential to language functioning in a way in which other types of meaning are not (see earlier discussion of 'Style as meaning', pp.5-7). Because it is not an optional language element the role of reference in making style contrasts might be expected to be small, but both Crystal and Davy (1969) and Leech (1966) have shown that the particularity and precision or vagueness of reference in a style can be of interest.

Leech, for example (ibid) ch.18 shows how particularity of reference, through brand names, to an advertisers product, and through deixis to the potential consumer plays an important part in Advertising English. Also, universal reference as in:

'Everyone loves Hartleys Jam'

and vagueness of reference as in:

'You'll have fresher breath...'

in a toothpaste commercial also demonstrate, according to Leech (ibid) the advertiser's use of the system of reference to communicate his message.

Crystal and Davy (1969) p.211 make use of reference when referring to the language of legal documents. Legal Documents represent a variety in which great care is taken that reference is exact and precise in order to avoid potential ambiguities and possible loopholes. The way in which language is used in legal documents is therefore diametrically opposed to that in advertising. In Advertising the writer often relies on ambiguity to sell his product, eg. with a new toothpaste your breath may be fresher than if you had use other manufacturers' products or just fresher than before you brushed your teeth.

The main problem in describing reference features would seem to be the difficulty of quantifying such terms as 'precise' and 'vague' in such a way as to make meaningful stylistic comparisons possible.



#### 1.4.4.2 'Strategic Semantics'

Reference, according to Leech (1974) is essentially concerned with the informational function of language (p.47). Meaning which goes beyond simple reference to associative meaning (see earlier discussion under 'Style as meaning', pp.5-7) is more concerned with the expressive function (putting over the speaker/writer's viewpoint) and the directive function of language (influencing the opinions of the reader/hearer).

Leech (1974) ch.4 shows how the manipulation of associative meaning may characterise certain language styles, particularly propaganda statements and advertising copy. Both these varieties according to Leech (ibid) make use of the favourable and unfavourable connotations of words to extol the virtues of their own point of view and to denigrate the points of view of others. Propagandists, according to Leech (ibid) often employ 'conceptual engineering' in their language; building up arguments on the basis of affective meaning relations rather than logical conceptual relations.

Like reference discussed above, strategic semantics suffers from the lack of any clear descriptive framework in the literature. Strategic semantic features would be difficult to identify in a clear and quantifiable way because of this, and even if identified, difficult to compare in different styles.

#### 1.4.4.3 The 'Higher Level Semantic Component'

Up to this point the discussion of the relevance of semantics to stylistic analysis has concentrated on the semantics of individual words. A criticism of this might be that as style is essentially a textual property (and not discernible in individual words when removed from context) within the working definition adopted for this thesis (pp.17-19), a semantic view which takes into account the 'semantic structure' of text might be a more fruitful basis for stylistic research.

Some authors have even gone as far as to say that it is with the semantic structure of texts that stylisticians should be working rather than with syntactic structure. An advocate of such a view is Hasan (1973):

"It has been too readily assumed that the easiest and most valid form of describing the linguistic characteristics of registers is to state the frequency or likelihood of individual patterns or of their combinations. I would suggest that it might be advantageous to specify the characteristics of given registers by reference to some high level semantic component."

Hasan (1973) p.273

Hasan's (1970) 'high level semantic component' involves classification of utterances according to their function, e.g.

'denigration', and 'stressing desirability'. These two functions could be used to characterise the language of buying and selling according to Hasan (p.274). Hasan contrasts this mode of approach with the type which in this introduction has been represented by Crystal and Davy (1969) and Leech (1966), among others, and claims:

"This mode of approach (Hasan's) seems to be perhaps more helpful for stating the characteristics of registers; it would certainly appear to be somewhat more productive than the counting of subordinate clauses or whatever."

Hasan (197 ) p.279

This type of semantic structure approach has been tried on the language of Advertising, and set out in Vestergard and Schroder (1985). The framework of analysis used by the authors is an eclectic one taken from the work of many disciplines. Models of Cohesion/coherence and information structure are taken from the work of linguists like M.A.K. Halliday, the 'actantial model' from the work of A.J. Greimas (this describes 'role' structure within a text in terms of subject/object, helper/opponent, giver/receiver), and 'schema theory' borrowed from psychology. Together, these models are used to explain the persuasive function of Advertising Language rather than why it might be conceived as belonging to a particular language style.

The English of Advertising is an attractive field to analyse in terms of how it goes about its business of persuading, because as

Leech (1966) acknowledges in his introduction this is a highly emotive subject. The way in which the English of Administration sets out to regulate people's everyday lives is also an equally emotive subject and one about much has already been written, eg. Wason (1984):

"It might be an interesting psychological exercise to penetrate the mental processes of the writers of official leaflets (as we once contemplated doing), but the dominance of bureaucratic obscurantism would remain untouched because it is motivated (in a very broad sense) by political interest. Lucidity is not the prime consideration of those who wield power, as even a socialist Minister of the Crown confessed to us in a casual remark. In such cases control is truly exerted through the written word: rules are made to bind people."

The point remains though, that however interesting and topical it might be to investigate how a particular language style like the Language of Advertising or Administration fulfills the aims of its writers, such work is fairly clearly outside the scope of mainstream linguistics, and appears to have more in common with psychological research.

It seems right then within this thesis to take the rather more conservative position of investigating whether a particular form of

the English language has developed to meet the particular needs of one situation: that of Government administration, rather than to evaluate how the aims and purposes of government administration are fulfilled in the language it uses. Because of this the use of semantic style variables in the present study will be fairly small, and will be restricted to the sort of analysis assumed under the heading of reference above. There will not be a separate chapter devoted to either Reference or Semantics.

#### 4.5 Measures of Periodicity and Complexity

Now follows a discussion of a group of possible stylistic variables that do not fit easily under any of the headings so far discussed. This is because whilst most of them do measure the frequency of various language attributes, usually the length of various linguistic items, none of these are centrally linguistic if linguistics is considered as the study of language structure. Furthermore claims are made of some of these measures that what they are actually measuring is 'complexity', though as will become apparent later this 'complexity' is not the 'complexity' of linguistic terminology.

Measures of periodicity are often included by linguists in style studies as part of an inventory of style variables. A good example of this is Walker Gibson (1966) who counts the average length of dependent clauses separating subject and main verb in sentences, in 3 different styles which he classifies subjectively into: The

Sweet talker, the Tough talker, and the Stuffy Talker. These styles are hard to classify in general terms but appear to correspond to light reading matter, heavier, more serious reading matter, and official language. Walker Gibson's (ibid) findings, p. 160 are that subordinate clause length in this position varies with style: (Tough -24 wds, Sweet - 36 wds, Stuffy - 182 wds).

On its own such information tells us very little, it is only by interpretation of this data that Walker Gibson (ibid) can arrive at any conclusions as to the significance of his results. In counting the length of dependent clauses coming between subject and verb Walker Gibson claims that he is investigating 'self embedding' structures. What the author appears to mean by this (p.164n) is the interrupting of the normal unmarked subject verb sequence with an adverbial clause, eg:

'The vegetables, once they have been peeled and diced, should be added to the broth'

His hypothesis appears to be on this issue that the longer the interrupting subordinating structure, the greater the tax on the reader's memory, (ibid, p.156). Walker Gibson however does not offer any psycholinguistic support for his view.

This study is a good example of the dangers of reading more into measures of periodicity than is necessarily there. Firstly not all dependent clauses between subject and main verb function as in the

example above, a dependent clause in this position may merely be a postmodifier in the nominal group of which the subject is a part. Secondly there is no psycholinguistic support offered for a link between length in this instance and reading difficulty. This is a problem which occurs again in other measures of periodicity.

When a measure of periodicity is used without interpretation, as is sentence length in Wallace (1981), there hardly seems to be any point at all in its use, except perhaps as input to a test of statistical significance. Wallace uses sentence length as a possible variable to distinguish between sports and news stories in the same newspapers. Not surprisingly Wallace fails to get significant results.

It does seem that this sort of use of a periodicity variable is hardly stylistic. It is sometimes true that sentence length may differentiate different texts, (the measure does work on part of Wallace's data), but it is by no means clear what meaning sentence length has in terms of linguistic structure. It could be a function of linguistic complexity in any one or several of the numerous structural items which go to make up a sentence.

The term 'complexity' is used in at least two ways in linguistics, the first use is simply to distinguish linguistic structures where there is a possibility of more than one item: thus;

'I' in: 'I could have danced all night' is a simple noun phrase,

and: 'Some people' in: 'Some people are mad' is a complex noun phrase

The second linguistic use of the word is in sentence structure where a complex sentence refers to a sentence which contains one or more dependent clauses. Neither of these definitions correspond exactly to the more everyday use of the word complex, to describe language that uses involved argumentation, or even just language which is difficult to read.

A great many tests have been devised to test for the latter type of complexity. They are too numerous to mention here, but are given an overall review in Harrison (1980). Harrison assumes a neutral viewpoint in relation to all the tests he reviews but does note that in terms of predicting reading difficulty some tests work better with some materials and readers than others.

No Stylistician appears to have used reading tests to distinguish different styles, but their use among writers of Administrative language is beginning to be recommended (Cutts and Maher, 1980 and H.M. Customs and Excise) The likelihood that reading tests really can predict reading difficulty in Administrative texts is held by some linguists in the field to be doubtful (see Campbell and Holland, 1982). One of the main arguments against the tests is that whilst there is some psycholinguistic evidence to suggest that what reading tests measure, (generally sentence length and word length in syllables) (eg. Coleman (1962)), does cause difficulty, this difficulty is caused by factors which reading tests measure only



indirectly (eg. the frequency of passives, which tends to correlate with longer sentences (see Gough (1965))).

It seems that the only relevance therefore which reading tests and periodicity measures might have to this study is the fact that their use is now encouraged among administrative writers. It is possible that administrative writers may now be tacitly or overtly indulging in 'formula fitting' which according to Campbell and Holland (1982) is the writing of texts which will give low scores on reading tests. The use of reading tests and periodicity measures therefore, together with information gleaned from government style manuals seems a possible way of pinpointing differences between styles deliberately aimed to be either informal or formal in AdE.

#### 1.4.6 Style Variables and AdE.

To summarise this section and conclude the present chapter it is important to come to a decision about what variables to consider in analysing AdE and how this will determine the format of the work to be undertaken.

It is already clear from what has been written so far in this section that there is a huge number of potential style variables, only a small number of which can be discarded as being of no use to this type of study. Having made this discovery the question arises as to whether to use all of the appropriate variables in researching AdE, or merely to select intuitively those which seem most likely to

yield results. If the first option is taken then there is a good likelihood that time will be wasted investigating variables which are of no stylistic significance. If this second option is taken then the researcher incurs the far greater risk of overlooking features of stylistic interest.

Where a great deal of research has been made in a field already, the second method would be appropriate: the general stylistic parameters of the variety under study being already known. To the authors knowledge however, no stylistic study has yet been done of AdE specifically, other than Chiu (1973) who examines only a very small aspect of AdE. Some authors have tried to compare 'official' texts with a multitude of other varieties (see Part 2 Ch 1 of this thesis). But these are usually legal rather than administrative texts. This thesis is therefore exploratory, and because of this it is better <sup>to</sup> run the risk <sup>of</sup> wasting time looking at uninteresting variables rather than to miss interesting stylistic effects.

The wisest way to make sure that all the potential stylistic variables in AdE are taken into account seems to be to follow the example set by Crystal and Davy (1969) and make a qualitative study of the variety. In order to make this manageable such a study would have to concentrate on AdE itself and not make comparisons with other varieties. Formality distinctions within AdE should be examined at this stage however. Such a qualitative study forms the subject matter of Part 1 of this thesis.

Whilst a qualitative study is an obvious first step in analysing AdE, research of this type raises the question of whether these intuitive observations might not be turned into hypotheses about the nature of AdE which can then be measured quantitatively. A quantitative study would be feasible at that stage because many variables would presumably have been rejected and the first study would provide the means to interpret the results of the second. At such a stage AdE might be compared with other varieties. A quantitative study would provide the possibility but not the certainty of adding support to the qualitative one and furthering knowledge about the use of quantitative techniques in stylistics, itself a far too neglected area of study. Part 2 of this thesis therefore comprises a qualitative study which is then discussed in the light of the contents of Part 1.

The chapter which follows this is a preliminary to the quantitative study in that it sets out the method of research followed.

## PART 1 - CHAPTER 2 - METHOD

### 2.1. General

The method of analysis for a qualitative study of the type to be conducted in Part 1 of this thesis is not of the same type as that used for the quantitative study in Part 2 (see pp. 235 - 284). This is because whilst the quantitative study is aimed at investigating hypotheses, to be generated through prior qualitative study, no such hypotheses are available in the case of the qualitative study.

The structure of this chapter is consequently somewhat simpler than that of the Method in Part 2. The Chapter covers: Design and Procedure, Texts, Reference Works.

### 2.2. Design & Procedure

The design of the qualitative study is not a 'design' in the experimental sense, and therefore cannot be classified as 'independent subjects', 'paired subjects' and so on. In order to carry out an analysis of what linguistic features appear to be stylistically distinct in Government Information Leaflets (GIL), it will first be necessary to establish what are the range of different types which represent these features in English generally. This is a rather general statement and needs some exemplification.

An example of what is meant by 'feature' might be Noun Phrase

Postmodification. In order to carry out the study in Part 1 it is necessary to find out, for English generally, what the possible patterns of Noun Phrase Postmodification are. This means in English exploring the syntactic categories which may function as Noun Phrase Postmodifiers. If available in the literature, information is also required on what use styles other than GIL make of the possibilities in Noun Phrase Postmodification, and whether certain possibilities are chosen more frequently in English generally, than others. Information about what the scope of particular structural features is in English generally will need to be obtained by use of a single reference grammar.

Note that elements of syntactic structure which are obligatory and not optional, e.g. the agreement between verb inflections and subject, are not of importance in setting out the descriptions carried out above. Being obligatory they carry little opportunity of expressing style, except where syntactic rules are sometimes deliberately violated, for example in poetic language.

After setting out a description of the scope of any given item of linguistic structure in English generally then attention can be turned to a corpus of GIL texts. The aim in examining such texts will be to do two things, firstly to note any particular items of linguistic structure that are unusually absent from the texts or that occur with a particularly high frequency, the assumption being made that these features are possible style markers in GIL. The

second aim is to attempt to ascribe the features thought to be stylistic to a particular stylistic dimension, for example do they provide information about the status relationship between reader and writer, or about the activity to which the text relates. In doing this, Part 1 of this study will be following the method of stylistic analysis suggested by Crystal and Davy (1969:p.20).

Note that because Part 1 is a qualitative study, isolation of features of potential stylistic interest will be made by reference to the author's own intuition, and not by any quantitative process of scoring variables on a text by text basis. This type of procedure is well precedented in Stylistic study, for example in Leech (1966), and attested by Sandell (1977) to be as valid as a quantitative analysis.

The chapter format which seems most appropriate to the design set out here is to let each separate chapter of Part 1 deal with a particular element of linguistic Structure. The possibilities for conveying stylistic information of the various elements of linguistic structure have already been discussed in detail in the previous chapter. The elements of linguistic structure to which individual chapters will be devoted in Part 1 are as follows:

Chapter 3....Noun Phrase

Chapter 4....Verb Phrase

Chapter 5....The Clause

Chapter 6....The Sentence

Chapter 8....Vocabulary

It is envisaged that each chapter will split into 2 main sections; the first setting out the scope of the particular element of structure to be studied in English generally, the second examining the use that GIL texts make of those structural possibilities.

### 2.3 Texts

The text corpus to be examined in Part 1 will be made up solely of GIL texts. The base (full) corpus of such texts was made by initially locating, by discussion with the National Association of Citizens Advice Bureaux, those Government Departments issuing leaflets to the general public. The list of such departments is set out below. It excludes departments like the Ministry of Agriculture and Fisheries which publishes leaflets of a technical nature intended for specialist groups.

Department of the Environment

The Home Office

Department of Employment

HM Land Registry

Department of Education and Science

HM Customs and Excise

Department of Health and Social Security

Office of the Health Service Commissioner  
Court of Protection  
Manpower Services Commission  
Department of Transport

Each Department was mailed directly using identical letters. The letter simply set out the aim of the research being conducted, and requested the department concerned to send single copies of all leaflets current at the time and available to the general public. It was assumed that departments would not select from their stock of leaflets. All departments replied to the request, with large volumes of leaflets being obtained.

For the purposes of Part 1 this base corpus was not 'sampled' in the true sense of the word, a large selection of leaflets was merely taken at random from the pile, and used to make the analysis. A true sample containing specific numbers of texts was however made for the purposes of part 2. This was not done in Part 1, as it was believed that an overview of the style of GIL texts would be most efficiently obtained by examining as large a selection of leaflets as possible. Where in the course of particular chapters quotations are made from individual leaflets these are referred to their original source.

### 2.3 Reference Works

Two reference works are used for the purposes of Part 1: A



University Grammar of English by R. Quirk and Sidney Greenbaum (1973), and 'Chapter 3' of Investigating English Style by David Crystal and Derek Davy (1969).

The first of these was chosen to fill the need for a standard reference grammar to both enable the scope of each element of structure studied to be described, and to ensure a consistent grammatical description and terminology throughout Part 1. This particular work was chosen precisely because it did not assume a specific linguistic theory. This decision was made partly because many grammars based on particular linguistic theories are not complete and therefore difficult to use for stylistic work, but also in the interests of having as neutral a grammatical description as possible.

Where Quirk and Greenbaum (1973) was unclear or further detail was needed on any point, the larger work on which Quirk and Greenbaum (1973) is based was used. This is A Grammar of Contemporary English by R. Quirk, S. Greenbaum, G. Leech, and J. Svartvik.

Though not primarily intended as a reference text, Crystal and Davy (1969), contains an important chapter (chapter 3) in which Crystal and Davy set out their framework for stylistic analysis. Crystal and Davy also provide a chapter devoted to linguistic description, this is not wholly compatible with Quirk and Greenbaum (1973) as it uses slightly different terminology. The main reason for not using

Crystal and Davy's linguistic description in Part 1 though is that it is not sufficiently detailed to allow the depth of analysis required here.

Crystal and Davy's framework for stylistic analysis involves the assignment of features believed to be stylistic to one of several dimensions of situational constraint (ibid p. 64). These dimensions and their applicability to the language of Government leaflets has already been discussed in detail (see pp. 23-28). A summary of the dimensions is given below and is copied from Crystal and Davy (ibid p. 66). The dimensions are listed in three groups: A, B, and C. These are explained below the list.

A

INDIVIDUALITY

DIALECT

TIME

B

DISCOURSE

(a) [SIMPLE/COMPLEX] MEDIUM (Speech, Writing)

(b) [SIMPLE/COMPLEX] PARTICIPATION (Monologue, Dialogue)

C

PROVINCE

STATUS

## MODALITY

## SINGULARITY

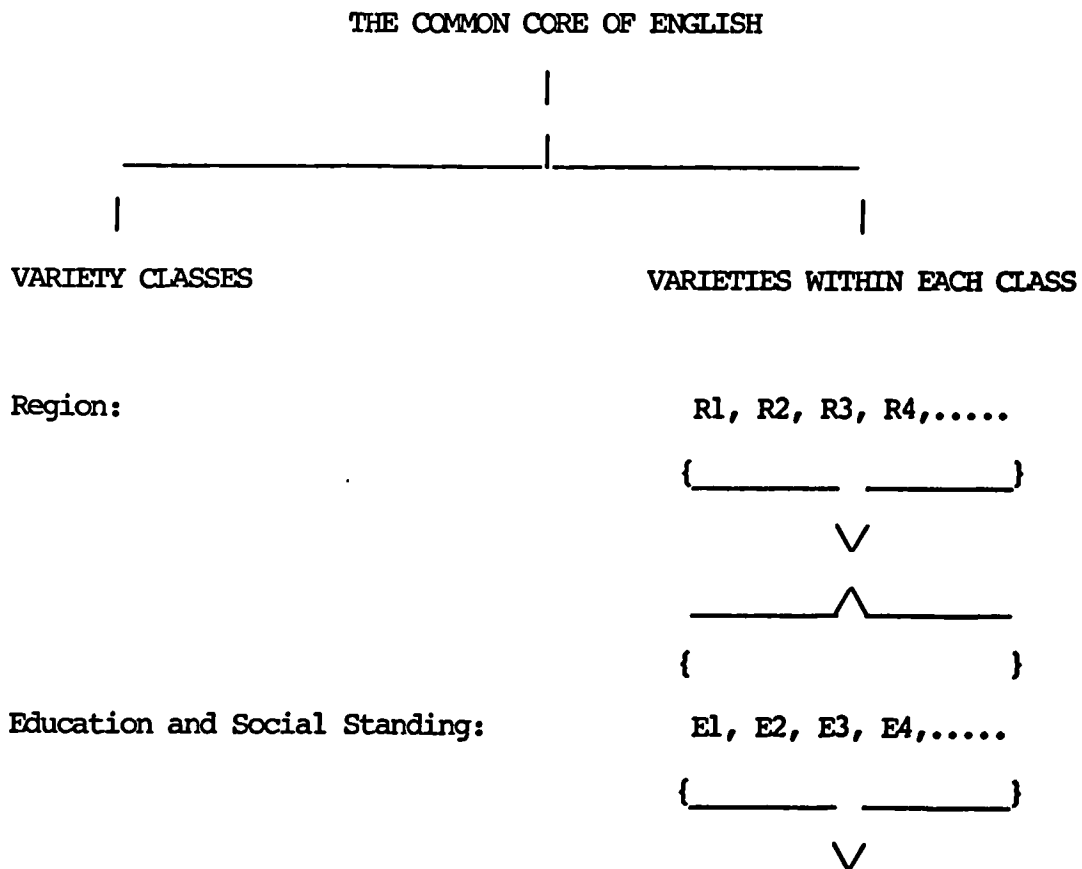
The first group of dimensions at A do not concern the present thesis, they are to do with idiolectal, dialectal and temporal language variation, and would not be considered as 'styles' by many writers. The complex dimension at B concerns the variation which arises in language as a result of whether speech or writing is chosen as the medium of expression, and of whether the language is in monologue or dialogue form. All these constraints are generally constant in leaflets which are usually in written monologue, but may be in dialogue (question and answer leaflets). These constraints are of only marginal interest in this study, though there will occasionally be need to refer to them.

The main dimensions of interest therefore for Part 1 are the first three at C: 'Province' which relates to 'the kind of occupational or professional activity being engaged in' (Crystal and Davy 1969:p. 71); 'Status' which relates to the social relations between the writer/reader; 'Modality' which relates to the form in which a text is expressed e.g. letter, essay etc. The final dimension at C; 'Singularity' concerns the use of linguistic features which are the preference of a particular language user. Crystal and Davy (ibid) intend this dimension to cover literary uses of language.

Whilst Crystal and Davy's framework is the one which will be used

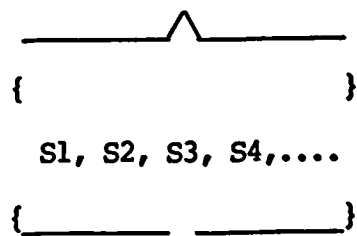
throughout Part 1, Quirk and Greenbaum (1973) also make stylistic distinctions in their Grammar which they refer to under the heading 'Varieties of English' (pp. 1-9). It is important before concluding the Method to see how Quirk and Greenbaum's 'varieties' correspond to Crystal and Davy's 'dimensions of situational constant', as Quirk and Greenbaum's categories are sometimes used when describing the usage of some linguistic features.

Quirk and Greenbaum set out their varieties schematically, and this schema is reproduced below.

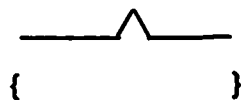


continued overleaf

Subject Matter:



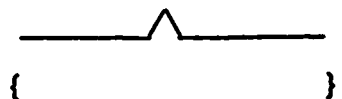
Medium:



M1, M2,.....



Attitude:

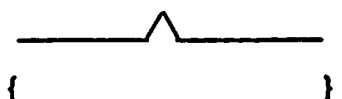


A1, A2, A3, A4,.....



---

Interference:



I1, I2, I3, I4,....

Quirk and Greenbaum's terms cannot be wholly equated with those of Crystal and Davy's (1969), because the former are the names of types of language whilst the latter are the constraints which produce the types. Insofar as comparison is possible, Quirk and Greenbaum's 'Common Core' features correspond to those in Crystal and Davy (1969:p65) which are not assignable to any particular dimension.

'Region' corresponds to Crystal and Davy's 'Dialect'. 'Education and Social Standing' together with 'Attitude' correspond roughly with Crystal and Davy's 'Status', 'Subject Matter' roughly to 'Province', and 'Medium' to 'Discourse Medium'.

The Crystal and Davy (1969) terms 'Discourse Participation' and 'Modality' have no parallels at all in Quirk and Greenbaum (1973). This is the main reason why Quirk and Greenbaum's framework is not used as a model in this study. Quirk and Greenbaum's term 'Interference' is not paralleled in Crystal and Davy. This refers to interference from a non native speaker's mother tongue when speaking English. This variety classification obviously has no part to play in the analysis of Part 1.

The brackets used by Quirk and Greenbaum in their schema are intended to convey the idea that varieties are not mutually exclusive but are inter-related (Quirk and Greenbaum 1973:p.8). In this respect their use of the term 'variety' seems misleading as it seems to imply clearly distinguishable language types that do not necessarily correspond. Crystal and Davy's (1969) terms seem more suitable on this count, it being clear that several different constraints may be operating simultaneously on the same utterance. However, Crystal and Davy (ibid), like Quirk and Greenbaum (1973), have to note that their situational constraints do not always operate independently (see Crystal and Davy 1969:p 89).

This section concludes this chapter on the Method of Analysis to be used in Part 1. Chapter 3 which follows sets out the analysis of the first syntactic category to be discussed in Part 1; the Noun Phrase.

PART 1 - CHAPTER 3 - THE NOUN PHRASE

3.1 General

English Noun Phrase structure falls logically under two distinct headings, which merit separate discussion. These are Premodification and Postmodification. Premodification will be dealt with first, initially as a phenomenon in English generally, then subsequently in Government Information Leaflets (GIL).

3.2 Noun Phrase Premodification in English

NP Premodification in English involves so many features that it is virtually impossible to illustrate them all in one sample NP. The following fictitious examples however, taken together, include all the possible elements of Premodification listed by Quirk and Greenbaum (1973). For reference these are:-

Predeterminer	(+D)	
Determiner	(D)	
Postdeterminer	(-D)	The abbreviations here
Adjective	(AJ)	will be used in subsequent
Participle	(P)	examples.
-s genitive	(GN)	
Noun	(N)	
Adverbial	(AV)	
Sentence	(S)	
Head	(H)	- neither pre- or post- modifier but the item modified



Examples of the elements of structure listed on the previous page:-

1. such a small man

+D D AJ H

2. the few good students

D -D AJ H

3. the enclosed lecture notes

D P N H

4. The Society's new headquarters

D N GN A H

5. The nearby house

D AV H

6. That I-can't-stand-it-any-more feeling

D S H

Each of the categories of Premodifier exemplified above will now be described separately in detail.

### 3.2.1 Predeterminer

Predeterminers consist of a small class of items which must precede

determiners, and are therefore defined by position. Semantically the function of predeterminers tends to be to express quantification.

The class includes:

Fractions e.g.

7. half his time

8. two thirds the amount

and quantifiers of the type: 'treble', 'double', e.g.

9. treble this sum

10. double the quantity

An additional type of predeterminer not specifically noted by Quirk and Greenbaum (1973), are those which function as emphasisers;

11. such an honest man

12. rather a large garden

### 3.2.2 Determiner

Determiners are again a small closed class which include:

articles, e.g. a, the, no;

demonstrative pronouns, e.g. this, that, these, those;

possessive pronouns, e.g. his, hers;

as well as items like; some, any, enough, every and each.

Quirk and Greenbaum (1973:p62) note that there are restrictions on determiner selection. For example, much may only precede non-count nouns.

### 3.2.3 Postdeterminer

Quirk and Greenbaum (1973) distinguish a further class of premodifiers which must follow the determiner, and are again like predeterminers a small closed class tending to function semantically as quantifiers, e.g.

13. the two main options

14. a few inconsiderable problems

### 3.2.4 Adjective

Unlike the classes of premodifiers described above adjectives are a

large open class. Examples include; blue, intimate, nice, reasonable etc. They tend to follow postdeterminers, but may be immediately preceded by an adverb which semantically serves to intensify the adjective's meaning, e.g.

#### 15. A really efficient service

Ordering of adjectives in the premodifying structure of an NP is more complex when there is more than one adjective present, and may involve their placing at intervals among other premodifying elements according to the adjective type; Quirk and Greenbaum (1973:p403).

Because adjectives are such a large class it seems possible that some styles may make use of particular types of adjectives which can be classified in some way. Certain sub-sets of adjectives are self evident and need no explanation, e.g. size, colour, provenance and so on, but one particular subclassification which has proved useful in stylistic (see below) work is Leech's (1966:p.129) attributive/designative dichotomy. This distinction is used by Leech in his description of Advertising English.

Leech classifies adjectives as 'designative' if they further define the NP head by placing it in a distinct category. An adjective so used describes a permanent and universally observable property, e.g. 'blue' when used to describe a pack of blue envelopes. Attributive

adjectives, on the other hand, attribute a quality to the head which is concerned with the response of the writer/speaker rather than any inherent property of the thing described e.g. 'delicious' when used to describe a peach.

### 3.2.5 Participle

Verbal participles used in NP premodification are used adjectivally, and treated as adjectives by some authors e.g. Leech (1966:p.129 ff). Participles may be of either the '-ed' or the '-ing' type for example:

16. the enclosed envelope

17. his engaging manner

Participial premodifiers should not be confused with denominal forms which resemble '-ed' verbal participles but are in fact derived from nouns and not verbs, as Quirk and Greenbaum (1973:p.398) show. Examples are; wooded, fluted, vaulted etc, and in this chapter are referred to as adjectives.

### 3.2.6 -s genitive

The -s genitive serves semantically to show a possessive relationship as in:

18. John's car

the possessor need not be a single proper noun but a string, e.g.

19. The Duke of York's hat

### 3.2.7 Noun

Noun premodifiers are like adjectives a large open class. They may often form very predictable relationships with the headwords that they modify, e.g.

20. The ticket office

21. The railway timetable

Noun premodification is also noted for its ambiguity. Leech (1966) notes that:

'As a class, the noun modifier has a semantic relation to its head which is multiply ambiguous to the point of vagueness.'

Whilst with most of the previously discussed postmodifiers it is quite clear what relationship they bear to the head noun, the situation is not so clear with the noun premodifier. For example;

## 21. a pony trap

could possibly refer to a lure to catch ponies or a pony drawn vehicle, and perhaps even a pony acting as a trap. Quirk and Greenbaum (1973) refer to this type of problem as a 'reduced explicitness relation'.

Campbell and Holland (1982:p.162) note that long strings of noun premodifiers are a feature of American Bureaucratic English. A good example of this phenomenon can be found in a companion paper by Charrow (1982:p.183):

'health education program evaluation measurement techniques'

Note that not every premodifier modifies the head word 'techniques'; 'health' modifies 'education', 'education' modifies 'program' which in turn modifies 'evaluation', and so on. The modification sequence is therefore recursive.

### 3.2.8 Adverbials

Adverbs have already been seen operating as adjectival intensifiers, but may also be found directly modifying the head word, e.g.

## 22. His far-away cottage

This example is taken from Quirk and Greenbaum (1973:p.395), but other examples are not easy to imagine. It seems likely that this type of modification is restricted.

### 3.2.9 Sentence

Quirk and Greenbaum (1973:p.395) demonstrate the possibility of a sentence occurring as an NP premodifier, for example;

23. 'His pop-down-for-the weekend cottage' (ibid)

Quirk and Greenbaum call this type of premodifier 'playful'. It seems likely to be a restricted class of premodifier, whose use however seems likely to denote a relatively informal style.

Having now described premodification in English generally, section 3.3 which follows will examine the types of premodifier to be found in GIL.

### 3.3 Premodification and GIL

This section sets out to examine the use made by GIL texts of the total possibilities of NP Premodification. The presentation of the elements of structure in NP structure follows exactly the same order as above in section 3.2



### 3.3.1 Predeterminer

In spite of the fact that Crystal and Davy (1969:p.41) remark that they: 'feel sure that there is much to be said about the way different varieties make use of predeterminer contrasts,' GIL texts make very little use of this category. Very seldom is the predeterminer slot filled in a GIL NP. In the NPs where a predeterminer is present this seems most often to be 'all'. The use of this particular lexical item may be due to the need for GIL writers to make the reader aware that the requirements of certain legislation are all embracing, e.g.

24. include tips and receipts from all sources

25. write down all the children who live with you

26. The new structure applies to all Goods vehicles

Very often this predeterminer is distinguished graphologically (not reproducible here) by bold type and/or block capitals, indicating that a considerable amount of importance is placed on the semantic inclusiveness of 'all' in GIL. GIL shares this need for inclusiveness with Legal English. This is evident when reading Crystal and Davy's comments on particular hiring agreement (1969:p.210):

(example on following page)

‘Note that the hirer is liable - a term already mentioned as being extremely exact in meaning - but liable in respect of damage - which is a term we used to exemplify flexibility. and not just damage but any damage, caused by any breach of any stipulation ! It would be difficult to be more inclusive than that,...’

‘Both’ is yet another frequently occurring predeterminer in GIL which serves a similar semantic purpose to ‘all’ in stressing inclusiveness in GIL, this time in relation to two people or things, e.g.

27. For a couple, both the man and the woman should sign

Note that those predeterminers which are evaluative, e.g.

28. quite a sight such a nice man rather a good buy

are notably absent from GIL, which as a medium for purely neutral and objective information about citizen’s rights and obligations, avoids these lexical items.

### 3.3.2 Determiner

Two determiners *appear* to characterise GIL texts; these

are 'any' and 'no', often with graphological emphasis, e.g.

29. ...provided that no relief from customs duty was obtained.

30. Any breach of these conditions renders the vehicle liable to forfeiture.

Both these determiners stress the importance to the reader of adhering to the regulations concerned, and like some of the predeterminers discussed above, function semantically to either completely include or exclude everything; in 29. it is clear that even small amounts of relief from duty are excluded, and in 30, that even minor breaches will result in forfeiture.

GIL again bears a close relationship to legal English on this point, the following are from holiday booking conditions and from an example in Crystal and Davy (1969:p.196):

'No agent or employee of the company has authority to vary these conditions'

'the hirer...shall be or remain liable in respect of any damage... by reason of any breach...of any stipulation...'

Determiners whose semantic function is imprecise quantification, like 'much', 'some', and 'enough' are notably absent from GIL. This

seems to be because, although GIL has been shown so far to favour the imprecision of blanket exclusion or inclusion, this is rarely true in quantification where figures are usually exact, e.g.

31. vehicles weighing over 1525 kgs.

### 3.2.3 Postdeterminer

The postdeterminer position has already been observed to be concerned almost uniquely with quantification (see sect 3.2.3). Not surprisingly then in a province whose concern is the explanation of citizen's rights and responsibilities; how old their cars must be before being tested, how much wine may be had in a duty free allowance, and so on, GIL frequently fills the postdeterminer slot in NPs.

The numerals involved in quantification are rarely written out in full, and would undoubtedly stand out less if this was done. Examples of typical GIL postdeterminers are:

32. If total proceeds exceeded `10,000 ....

33. persons born before 6th April 1919

34. within 3 years of the date of its exportation.

### 3.2.4 Adjective

Adjectival modification is restrained in GIL, in that one does not

see the multiple adjectival modification so often found in Advertising English. The adjectives used appear to form a fairly closed small set, and are largely designative rather than attributive. The adjective items which occur most frequently in GIL appear to be those concerned directly with the province of administration. These include; 'concesssionary', 'general', 'exempt', 'chargeable', 'additional'. Here are some examples of these items in use:

35. Concessionary rates for farmers and showmen

36. the general conditions required by this resolution

37. ...may be imported free of VAT by an exempt person

38. an additional personal allowance

39 chargeable assets

Semantically, the adjectives used in any GIL text making a direct explanation of legislation appear to be highly precise. Adjectives rarely seem to possess the wide range of meanings in GIL texts that they would in other varieties. What is perhaps of greatest interest is that the meanings of adjectives are likely to be the subject of prior agreement amongst the members of a given government department because of the necessity to make leaflets legally correct.

Adjectival reference may therefore vary from one province (in this case Departments or Ministries) to another. This point becomes clearer when examining text samples: The adjective 'heavy' has a very specific meaning when explaining vehicle taxation, as has the adjective 'light' in the NP 'Private Light Goods' found on car tax discs. 'Heavy' in this context refers to 'weighing over 1525 kgs', according to D.O.T. texts.

In a similar way, the Inland Revenue restrict the meaning of the adjective 'blind' to those persons who are registered as blind with their local authority, the adjective 'disabled' has a similarly restricted meaning. In the specialisation of its adjective meanings GIL can again be paralleled with legal English where there is mutual understanding between lawyers on the exact meaning of 'terms of art'. This point is made by Crystal and Davy (1969:p.210).

'The technical terminology or special vocabulary of the law is remarkable not for the fact that it contains a mixture of words, some with exact and some with less exact meanings, but because in many instances the degree of exactness is the subject of a kind of tacit agreement between lawyers.'

Note that in GIL texts the choice of certain head noun and premodifying adjectives are highly predictable, good examples of these from GIL texts are:

39. annual income; standard conditions, contractual obligations,

weekly amount, statutory fee, fair conditions and further information.

This phenomenon is not solely a property of GIL texts; Pytelka (1972) notes exactly the same type of feature in the English of Business and Commerce. Pytelka calls the phenomenon 'automation'. Crystal and Davy (1969:p208) note characteristic collocations in legal English as well, for example; term of years, and upon the death of. It seems likely from this evidence that many linguistic styles, particularly those where the province involved is one associated with a trade or profession, use automation, alongside their inventory of technical terms.

Some of the automatised phrases of GIL exemplified in example 39 on the previous page, are self explanatory, and no doubt occur in other styles. 'Weekly amount' and 'annual income' fall into this category. Collocations like; 'statutory conditions' and 'contractual obligations', however have specialised meanings in GIL and can only be interpreted in terms of the particular legislation relevant to the text in which it occurs. It seems strange that such technical terms should occur in GIL because of the pressure from many sources on Government Departments to make their leaflets easily understandable. The following extract from a recent departmental style manual illustrates this pressure;

Jargon Avoid Departmental jargon and code words. If their

use is essential make sure that the words are listed in an "Explanation of terms section."

None of the texts from which the collocations of example 39 were drawn contained such an explanation of terms section.

Intensification of adjectives is notably absent in GIL. This absence is closely paralleled by the findings of Gibson (1979:p.153) who finds only a negligible number of adjective intensifiers in his 'stuffy talker' texts, some of which appear to be texts with an administrative function, not dissimilar to GIL. GIL has no requirement for intensification of its stock of uniformly designative adjectives since their meaning cannot easily be intensified. There is for example no;

40. very concessionary rate, or

41. highly statutory conditions.

'Rates' are either concessionary or not so in the administrative province which constrains the GIL style, and the same applies to 'conditions' in example 41. The contrary to this tendency is observed by Leech (1966) in Advertising English where intensification is used with (usually) attributive adjectives. The intention behind their use presumably being to further enhance the desirability of the goods advertised, e.g.:



unbelievably low prices, incredibly good value for money

### 3.2.5 Participle

The use of participial premodification allows for what is sometimes referred to as a 'nominal style'. The use made of this capability in GIL is extensive and is explained below. In contrast to the restrained nature of adjectival modification in GIL participial modification is very frequent, and it appears to be the 'ed' participle that is most often used. Examples abound in the data. Typical examples are:

42. Approved profit sharing scheme

43. a fixed reduction

44. def. erred annuity

45. the completed application form

46. before the proposed date of the wedding

The very frequent use of the '-ed' participle in GIL shows the fondness of GIL writers for a nominal rather than a verbal style. This may be because a premodified noun phrase using the '-ed'

participle is generally more economical on words than a corresponding phrase using an active verb phrase, for example compare the underlined portions in:

47. The limited grant aid, which must be claimed before the 3rd April,... (text sample)

48. The Government has limited grant aid, which must be claimed before 3rd April

This search for conciseness of statement may be partly due to the space constraints which 'leaflet size' makes on the writer, but also because the use of such a nominalising feature allows more information to be packed into a single simple sentence. In example 47. above, the required information has been stated in the premodification structure of the NP, allowing what had to be stated in a subordinate clause in example 48 in the postmodifying element. The VP of the sentence has not been reached, leaving this free for the expression of further information content.

Note that as well as being more concise, phrases like that in 47. allow impersonality of statement, whilst the verbal counterpart is more likely to be the opposite. This tendency to impersonality in GIL is repeated in the verb phrase, where passive constructions are frequently used. For more information on this point, consult Chapter 4 on the Verb Phrase.

The '-ing' participles are a little less evident in GIL than their '-ed' counterparts; examples of those that do occur include:

49. Her future husbands existing passport

50. any one of the participating garages

51. Local Vehicle Licensing Offices

52. qualifying loans

Note that here too the use of the participial premodifier allows a concise nominal style; compare 52 for example with:

53. the loans are for qualifying purposes.

Note also that participial modification is far more likely to be designative than attributive. GIL has already been observed to favour designative rather than attributive modification.

### 3.2.6 -s genitive

The use of the -s genitive in GIL illustrates another area in which this style makes full use of opportunities for avoiding personal reference. Many GIL texts make frequent use of the -s genitive because constant reference to the reader as one of a number of

administrative classes, e.g.; the claimant, the applicant, involves the writer in statements like:

54. The vehicle must be for the importer's own use..... and,

55. The claimant's weekly income.

instead of:

'...for your own use', 'your weekly income'

In some GIL texts the reader is never referred to as 'you', only by the sort of oblique method exemplified above. It must be observed at this point that GIL texts do not all conform to the same stylistic norms in this respect; in fact large numbers of texts appear to belong to two distinct types, those which personally address the reader and those which do not. Subjectively speaking there appears to be a difference of formality between these two text types, with the former being less formal than the latter. There is further subjective evidence that such a strong division exists in GIL, Reference will be made to this in subsequent Chapters of Part 1, as the status difference appears to be of considerable importance in GIL.

The -s genitive can be added to a phrase of considerable length, and Walker Gibson (1970:p.153) notes that his stuffy talker texts (which appear to contain some U.S. administrative texts) make full use of

these possibilities. To support this Gibson (op cit) cites the following example:

"the College Entrance Examination Board's Scholastic Aptitude and Achievement test"

Such strings do not however appear to occur in GIL.

### 3.3.7 Noun

Noun premodifiers tend to occur particularly in titles to documents, leaflets, benefits, or names of offices. Here are some examples:

56. Attendance allowance, Disability Allowance, Age Allowance

57. Goods Vehicle rate

58. Enquiries section, Enquiries desk

59. Order Book, Pension Book

60. Housing Benefit Supplement

It is clear that the administrative province which constrains GIL will need to label things in such a way; Housing Benefit Supplement must be distinguished from Housing Benefit itself, but also from the

similarly named 'Supplementary Benefit' to which it is partly related. Rates for Goods vehicles are dealt with separately to those for private cars. Administration is very much a labelling and compartmentalising activity, and for this reason noun premodification is heavily used in GIL, far more so than adjectival modification.

Labelling in this way can lead to quite lengthy phrases composing of nothing else but nouns. This tendency is most marked in U.S. Administration as Charrow (1982:p.183) reports, quoting the following example;

'health service program evaluation measurement techniques'

Again, in relation to U.S. administrative texts, Gibson (1970) reports that over 5% of total words in the texts he studied were noun adjuncts (noun modifiers).

Charrow (op cit) expresses anxiety that long 'noun strings' like the example above may intimidate readers. And a companion paper by Campbell and Holland (1982:p.163) reports research claiming to show that rewriting such sequences using prepositional phrases can result in easier comprehension. Thus an example like;

61. The Heavy Goods vehicle register....., would become:

62. The register for Goods vehicles of a heavy type.

It seems almost certain however that GIL writers opt for the noun string as in example 61 precisely because it is much neater and shorter, and conforms to the leaflet modality requirements of economy of statement. Example 62. in comparison seems clumsy and long winded, it seems unlikely that it would find favour with GIL writers.

### 3.3.8 Remaining Categories

Clause and Sentence Premodifiers were described as rare in section 3.1, and it was also noted that the sentence premodifier in particular appears to be a marker of an informal style. It is not surprising that very few such premodifiers exist in GIL, though there is one notable exception. This is the Inland Revenue's 'the pay as you earn scheme', which makes one suspicious that this was an attempt to give a friendly aura to that most disliked of obligations, paying income tax.

### 3.3.9 Summary

To summarise the functions of NP premodification in GIL these appear to fall into three main categories:

1. Quantification: an important function enabling GIL to convey a

wide variety of information on dates, amounts of benefits, rates of duty and so on.

2. Designation: the labelling function, enabling the classification and sub-classification<sup>a</sup> of a large number of services and departmental sections.

3. Economy of Statement: The function of allowing a nominal style, thereby enabling more information to be packed into a single sentence

#### 3.4. Noun Phrase Postmodification in English

The description of Noun Phrase Postmodification given here is adapted from Quirk and Greenbaum (1973:ch.13).

Three possible items of structure may occur as postmodifier to an NP, these are; words, prepositional phrases, and clauses. These three possibilities will be described separately.

##### 3.4.1 Single word postmodifiers

This type of postmodification is classified under the heading; 'Minor types of postmodification' by Quirk and Greenbaum (1973:p.392). The 3 main word types which may occur as NP



postmodifiers appear to be:

adjective e.g.

63. the last train possible, the only essay readable

adverb e.g.

64. the additional notes below, the free gift inside

reflexive pronouns e.g.

65. the director himself, the secretary herself

### 3.4.2 Postmodification by Prepositional Phrase

In contrast to the above, Quirk and Greenbaum (1973:p.388) note postmodification by prepositional phrase (PP) to be the commonest type of postmodification in English, being reported as three to four times more frequent than clausal postmodification. Examples of PP postmodification might be:

66. the cupboard under the stairs, the clock on the mantelpiece

The commonest type of PP postmodification is held by Quirk and Greenbaum to be postmodification using the 'of - genitive', for

example:

67. the home of the Viscount.

This is an example of the of-genitive being used to denote possession, but Quirk and Greenbaum (1973:s.4.69ff) note that it may be used to express up to seven different relationships, of which possession is only one. The others are:

subjective

68. the refusal of the committee

objective

69. an appraisal of the figures

origin

70. The steppes of Central Asia

description

70. a herd of sheep

measurement/partitive

72. a cruise of three weeks, a slice of the cake

appositive

73. the state of Connecticut

3.4.3 Clausal Postmodifiers

Clauses functioning as noun phrase postmodifiers, are either of the finite relative type (introduced by a relative pronoun: e.g. 'which', the general pronoun 'that', or the adjunct form e.g. 'where') or a non-finite clause. Non-finite clauses may be of one of three types; those using the -ing participle, those using the -ed participle, and those using the infinitive. Examples of all these types are now given, possible variant forms are given after oblique slashes.

finite relative clause

74. the street where you live/that you live on

75. the one that got away/which got away

76. the girl I love/whom I love/that I love

non-finite/-ing participle

77. the girl dancing on the table

78. the man leaning on the lampost at the corner of the street

non-finite/-ed participle

79. the letter enclosed

80. the solution decided upon

non-finite/infinitive

81. the thing to do

82. the horse to back in the three-thirty at Kempton Park

Note that whilst the simplest postmodifying clauses may consist only of a Verb Phrase as in example 78, they may have optional adverbials, as in most of the examples above, and these may be co-ordinated as in examples 81 and 77. A complement is also an optional feature, e.g.

83. the officer conducting the investigation

Note also that except where the relative pronoun introducing a finite clause refers to the subject, it is optional and may be omitted, e.g.

with relative pronoun

84. the man that I hate most of all

without relative pronoun

85. the man I hate most of all

The large amount of optionality in clausal NP postmodification set out above, clearly provides much potential for stylistic variation, this potential is further widened by the possibility of substituting finite relative clauses for non-finite clauses and vice versa, when referring to exactly the same set of circumstances, e.g.:

86. the solution adopted (non-finite clause)

87. the solution which was adopted (finite relative clause)

Quirk and Greenbaum (1973:p.381) note that the choice of a

particular type of clause postmodifier is connected with relative formality, this suggests that the choice of particular types of clause postmodifier may be a Status feature. Quirk and Greenbaum (ibid) do not elaborate on this. They do however observe that NP postmodification generally can be ordered hierarchically according to explicitness (ibid:p.377-379), with finite relative clause postmodification as the most explicit, and postmodification with prepositional phrases as the least explicit. Non-finite clauses occupy a mid-point in the hierarchy. Postmodification in its turn is held to be more explicit than premodification, (ibid).

#### 3.4.4. Multiple Postmodification

Unlike previous section headings on words, phrases, and clauses, multiple postmodification does not refer to the role of any one particular structural item in NP postmodification, but to the fact that there may be departures from the situation so far described, of having a single NP headword postmodified by a single postmodifier.

Quirk and Greenbaum (1973:p.393) observe that there are three possible variations on this situation. The first is to have one headword modified by several postmodifiers, the second to have one postmodifier modify several headwords, and the third is to have a postmodifier which is itself postmodified. Examples of these three variations are now given:

one headword-several postmodifiers

88. the man in the corner, with a beard, reading a newspaper

H            PM1                            PM2                            PM3

one postmodifier/several headwords

89. the girl, the man, and the dog standing by the sleazy cafe

H1            H2                            H3

postmodified postmodifier (embedding)

91. the man leaning on the lamp post at the corner of the street

PM1    PM2

(‘at the corner of the street’ postmodifies lamp post)

In the above examples, ‘H’ identifies headwords and ‘PM’ identifies postmodifiers.

Multiple postmodification further increases the potential for stylistic variation which postmodification in English generally, possesses: Even styles which employ postmodification to a similar extent may vary widely in how ‘elaborate’ their postmodification is. In other words, how much do they use multiple postmodification? and how embedded are the postmodifiers, or how many postmodifiers or headwords are used ?

Finally, Quirk and Greenbaum (1973:p.394) note that multiple postmodification may result in ambiguity which cannot be resolved by more careful ordering, and is a feature of the structure itself, the example which they give is:

'The smiles of delight on all the faces that he liked'

where it is unclear whether it is 'the smiles' or 'the faces' that 'he' liked.

Having now described the scope of NP postmodification in English generally, the next section (3.5) will go on to describe the use made by GIL of this.

### 3.5 Postmodification in GIL

#### 3.5.1 General

One of the most striking features of postmodification in GIL is the way in which the text types identified in the section on premodification above, (section 3.3.6) differ in their patterns of postmodification. The basis of the division into two text types was on the basis of whether they addressed the reader personally or not. The personally addressed (P.A.) texts appear to make greater use of clausal postmodifiers, particularly finite ones, e.g.:



92. employment advisers who can help advise you

whilst the impersonally addressed texts, (I.A.) appear to favour prepositional phrase postmodifiers, particularly the of-genitive, e.g.

an expression of the will of Parliament

Examined subjectively, P.A. texts appear to have a greater number of clausal postmodifiers than might be expected from Quirk and Greenbaum's (1973:p.387) observation that phrasal postmodification is three to four times more frequent than clausal postmodification in English generally. A similar but opposite trend appears to be the case for I.A. texts where there seem to be far more phrasal postmodifiers than might be expected. These qualitative judgements must however be accepted cautiously, they do not have the same standing as statistically verified figures, and should be the subject of further research.

A final general observation on the nature of postmodification in GIL must be that the sheer volume of postmodification in words, particularly in I.A. texts means that a substantial portion of texts must often fall within postmodifying structures. A striking example of this can be found in a text quoted by Vernon (1980), where the postmodifier is extremely lengthy and embedded:

93. a supplementary allowance of an amount which, when added to any unemployment benefit (including any earnings related supplement) to which you may be entitled, will equal the amount shown for the first of the benefit pay weeks stated opppsite.

This heavy use of postmodification in some GIL texts contrasts markedly with Crystal and Davy's (1969:p.113) observation that Conversational English makes very little use of it, and Advertising English which Leech (1966) finds to favour almost exclusively premodification.

Attention will now be turned in more detail to the different types of clause postmodifier discussed in section 3.4, and the way in which they are used in GIL

### 3.5.2 Word Postmodifiers

This type of postmodifiers was described as 'minor' in section 3.4. It is certainly minor as well in the role it plays in GIL. For this reason it is assumed that it has no part to play in making stylistic contrasts in GIL.

### 3.5.3 Prepositional Phrase Postmodifiers

This type of postmodifier, occurs very frequently in GIL texts which can be described as Impersonally addressed, though the range

of prepositions involved is rather small, the of-genitive being by far the most common item introducing postmodifying sequences. Here are some examples:

94. copies for perusal

95. the importation of private motor vehicles

96. hours and conditions of labour

97. periods of unemployment

98. all keepers of heavy goods vehicles

99. gross weight method of assessment

The main observation to be made about the high frequency of the of-genitive postmodifier in I.A. texts concerns the notion of 'explicitness' referred to in section 3.4. The of-genitive postmodifier lacks the explicitness enjoyed by the finite relative clause for example, which may express person and tense. The of-genitive is however economical on words, being generally shorter than a finite relative clause for example.

The constraints which determine the choice of this inexplicit, but

economical form of postmodification in I.A. may be largely due to Province. I.A. texts deal with subject matter which is often of a more technical nature than that of P.A. texts, and assumes that the reader already has an acquaintance with the subject matter. For example whilst a typical P.A. text might be an introduction to how to claim Supplementary Benefit, an I.A. leaflet is more likely to deal with Goods vehicle taxation, or VAT returns. Such technical leaflets as the latter clearly have no need to explain explicitly to the reader what is shared knowledge, the emphasis instead is on making sure that the leaflet covers its subject thoroughly, it must therefore be economical on words.

P.A. texts on the contrary, dealing as they often do with introductions to subjects which readers know very little about, use more explicit postmodifiers, and this emerges when clausal postmodifiers are examined. This is done in the next section

#### 3.5.4 Clausal Postmodifiers

P.A. texts make very frequent use of clausal postmodifiers, particularly finite ones, with deletion of the relative pronoun. Here are some examples:

100. the money you have coming in

101. the leaflet you want

102. any skills you have

103. things you need to have with you

104. copies you can look at

Note that several of these examples are paraphrases of what would undoubtedly be much shorter sequences in an I.A. text. In one case, that of example 100, the paraphrase is of a single word 'income', presumably deemed too technical a term for the claimants of supplementary benefit. Example 104. is directly paralleled by example 94. above from an I.A. text; 'copies for perusal'. This is evidence not only for the translatability of clausal postmodifiers in P.A. to phrasal in I.A. but of a general tendency of P.A. texts to avoid departmental jargon. This seems to further support the informal hypothesis expressed earlier that I.A. texts assume shared knowledge on the part of the reader (including knowledge of the jargon used), whilst P.A. texts do not.

In addition to this, the inclusion of the second person pronoun, placed in a position of prominence by the frequent absence of any preceding relative pronoun, and present in all the P.A. examples above, seems to indicate a different status relation between reader and writer than that which holds for I.A. texts. In addressing the reader as 'you' the P.A. writer appears on an equal standing to his

reader. The tone of the text seems subjectively to be friendly. This type of relationship would not be appropriate to the transfer of technical information which occurs in I.A. texts.

Non-finite clauses have not yet been discussed. This type of postmodifier is to be found in both I.A. and P.A. texts though far more frequently in the former than in the latter. It appears to be essentially a feature of I.A. texts therefore. Here are some examples:

105. any customs duty previously relieved

106. every factory, workshop, or place occupied or used by him

107. those established for the trade or industry

108. foreign ports incorporated in the scheme

109. the gross train weight indicated in column 2 of the Ministry plate

The non-finite clauses found in GIL are almost exclusively of the -ed participle type. In terms of explicitness, non-finite clauses were held to be less explicit than finite clauses (see sect 3.4). This is because the non-finite clause the tense and person markers which are present in the finite clause. The switch from a finite

clause postmodifier to a non-finite one may therefore a switch to a more impersonal type of postmodifier, e.g.:

110. the stipulations which the management recommends for inclusion

111. the stipulations recommended for inclusion

Example 111. appears to assume that writer and reader have enough shared knowledge to make reference to the management (the agent) redundant. The resulting NP seems subjectively more formal but it is more economical on words than the finite clause of example 110. In all but one of the examples of non-finite clause postmodifiers (examples 105 to 109 above), no agent is mentioned.

Viewed alongside I.A. text's preference for the less explicit prepositional phrase postmodifiers, this use of the non-finite clause in postmodification in I.A. seems to further strengthen arguments that I.A. texts are more technical in subject matter and assume more shared reader/writer knowledge. The view also conflicts with the popular belief that bureaucrats use impersonal and non-explicit language merely to remain anonymous. The view taken here is that in a leaflet modality, reference to the agent if specified would only be in terms of a nouns like 'the department', 'the government', 'the administration' and so on. These are nouns which do little to dispel anonymity, and where they are understood between reader and hearer are redundant.

The situation differs in letter modalities where the official writing may indeed have some reason for being anonymous, the convention for example that all decisions made in a department are the ultimate responsibility of the departmental head.

#### 3.5.4 Multiple Postmodification

Examples of all three types of multiple postmodification set out in section 3.4 occur in GIL, e.g.:

##### single head/two postmodifiers

111. a woman who is to be married in the U.K. and who will be going abroad immediately after the wedding

##### two heads/single postmodifier

112. hours and conditions of labour

H1                      H2

##### postmodified postmodifier

113. details of pay relating to your own full time employment

PM1                                      PM2

Inspite of the popular belief that Administrative language is often



unclear, no instances of genuine ambiguity produced by multiple postmodification were found in the data. What does have to be noted however is that embedded multiple postmodification is quite heavily used in I.A. texts, frequently involving more than 2 postmodifiers. This in itself can result in rather long NPs. Added to this is the tendency to be discussed later (Chapter 5 on the clause) of GIL to make frequent use of long clause adverbials. When clauses are used as postmodifiers this can result in NPs of the rather fearsome type set out below. This example was quoted at the beginning of section 3.5. Here it is annotated.

a supplementary allowance of an amount which, when added to any un-

PREMOD                      HEAD                      PM1                      PM2                      ADVERBIAL OF PM2

employment benefit (including any earnings related supplement) to

PARANTHETIC CLAUSE INSERTED INTO ADV OF PM2

which you may be entitled, will equal the amount shown for the first

PM3 (POSTMODIFIES 'benefit'      REMAINDER OF PM2      PM4      PM5

IN ADV OF PM2)

of the benefit pay weeks stated opposite

PM6

PM7

In this example the main verb of the clause of which this large NP forms a part has not yet been reached, and the reader must of course

hold all this information in his mind before he proceeds to take in the remainder of the clause. There is nothing ambiguous about this passage though. GIL clearly has need of such sequences where complex pieces of legislation cover the information that has to be explained to the public, and the style is therefore not surprisingly reminiscent of legal language, where precision of statement rather than elegance is required. Whilst Gowers (1962:Ch.3) has insisted that this type of stylistic borrowing should not take place in Administrative English, it clearly does, and is an example of the type of constraints placed upon GIL by the subject matter with which it may be called to deal. It leads to the hypothesis that there may be no very clear cut line between the styles of legal and Administrative English, but a grey area where the two merge.

### 3.5.5 Summary

In summary NP postmodification in GIL illustrates well the differing constraints operating on I.A. and P.A. type texts. The personality or impersonality of address involved is not interesting for its own sake but because of the Province and concomitant Status constraints which bring it about. These constraints are the same ones that bring about the choice of predominantly (explicit) finite clause postmodifiers in P.A. texts, but (less explicit) phrasal and non-finite clause postmodifiers in I.A. texts.

The Provinces of I.A. and P.A. appear to differ, one concerned with

the imparting of technical information, the other with general, introductory information. Status relationships are related to these Province differences; in I.A. the relationship is that between a reader and writer who are strangers to one another, but share some knowledge about their subject matter, in P.A. the writer tries to assume a more personal relationship with the reader, though there is little shared knowledge: The aim here may be to establish confidence between reader and writer by assuming a friendly personally directed style. This confidence presumably already exists in the case of I.A. texts.

Finally, multiple postmodification, especially embedded postmodification, highlights the constraints which a province concerned with explaining legislation to the general public imposes. In doing this multiple postmodification in GIL suggests links with Legal English in addition to those which were observed in GIL premodification.

This section summarising NP postmodification in GIL completes Chapter 3 on the Noun Phrase. Chapter 4 which follows deals with the Verb Phrase.

Note: Text continues on page 117

## PART 1 - CHAPTER 4 - THE VERB PHRASE

### 4.1 General

The three aspects of the verb phrase which provide most potential for stylistic variation are; tense, aspect, and mood. All three of these areas will be covered in this chapter, a section being devoted to each. The format of examining the structural patterns in English generally, then looking at the use made of these by GIL will be kept to throughout. Voice will be dealt with as a part of Chapter 5, on the clause, which follows.

#### 4.2.1 Tense in English generally

Tense and time are not one and the same thing, as Quirk and Greenbaum (1973:p.40) open their discussion about tense by observing:

'Time is a universal, non-linguistic concept with three divisions: past, present, and future; by tense we understand the correspondence between the form of the verb and our concept our time'

Tense in English may only take one of two forms; past or present, in use these do not refer always to past and present time, as will be seen. It is hardly necessary to exemplify the forms of the past and

present tenses here, but to clarify matters, the present tense will be used to refer to the following paradigm of forms:

I wish  
you wish  
he wishes  
she wishes  
we wish  
they wish

The label 'past tense' will be used to refer to the following paradigm:

I wished  
you wished  
he wished  
she wished  
we wished  
they wished

Note that tense is a property of finite VPs and that non-finite VPs cannot show tense. In addition, Palmer (1974:p.33) notes that Imperative verb phrases, e.g. Go! Sit! do not show tense distinctions.

The present tense may refer to a variety of time situations.

Combined with perfect aspect, the present tense may refer to an event in the past which still has relevance in the present, e.g.

114. I have lived in London since last June.

Even the simple present, i.e. the present without any aspectual contrasts, may refer to past time, in a fictional text, e.g.

115. As I write, it is the year, 1914...

But it is to the simple present and to future time that the simple present is most used to refer. However even within these time relationships, the usage of the simple present may vary. It is to these various usages that attention is now turned.

The simple present tense is most frequently used to refer to habitual activity. This fact is attested by Palmer (1974:p.60). An example of this use is:

116. I attend the badminton class every Tuesday

The simple present may also however be used to describe timeless truths according to Quirk and Greenbaum (1973:p.41), e.g.

117. The earth is round.

and in situations where the act of speaking/writing is simultaneous with a performed action, e.g.

118. I, Edward, solemnly declare...

This type of use is labelled 'performative' by Quirk and Greenbaum (1973:p.42). It tends to act as a stylistic marker of formal ceremonial language, and legal documents, e.g.

119. I name this ship 'Queen Elizabeth 2'

120. I hereby declare that, I, Matthew Jackson....

A final use of the simple present tense to refer to simple present time is that of commentary; this usage, is predictably found frequently in sports commentaries, e.g.

123. and Borg makes a fine backhand return to Connors...

where the action on the tennis court and the commentator's description of it are virtually simultaneous.

The simple present may alternatively be used to refer to future time, frequently in temporal and conditional subordinate clauses, according to Quirk and Greenbaum (1973:p.49), e.g.

124. What will happen when Stanislavsky claims diplomatic immunity?

125. I will/shall not withdraw my claim unless you withdraw yours.

Note that the use of WILL and SHALL in this type of example to refer to future time. The interpretation of these auxiliaries poses a problem, and is discussed by Quirk and Greenbaum (1973:p.47). They note that the use of these two auxiliaries provides 'the closest approximation to a colourless, neutral future.' But also that it is very difficult to separate the modal and future functions of these auxiliaries. This point is supported by Palmer (1979:p.111). For instance example 125 above, could be interpreted as indicating either future reference or lack of willingness. The first interpretation being one of reference to time and the second to mood.

In main clauses, the simple present tense is observed by Quirk and Greenbaum (1973:p.41) to express certain events in the future, e.g.

126. Tomorrow is Wednesday

127. 1984 ends next Friday

and to refer to timetabled events, e.g.

128. The match starts at two, the bus leaves at twelve and arrives back at six.



Finally there is a problematic use of the simple present tense that does not fit into any of the categories so far described. This is exemplified in the following;

121. Valmai lives in Gaerwen.

122. John Idris works at the Power Station.

where the use of the simple present seems to indicate a durational state rather than a habitual activity, conveying a meaning very much like the simple present progressive, described under Aspect below. Palmer (1974:pp.70-74) claims that the reason for this is the contrast between Stative and Dynamic verbs. He points out that verbs like live and work refer to states, unlike dynamic verbs like kick, hit, and strike which refer to activities. Stative verbs appear to have duration as part of their meaning, and are therefore not used with the simple present progressive unless very limited duration is being emphasised, e.g.

123. Valmai is living in Gaerwen (until she can find somewhere else)

124. John Idris is working at the power station (this week)

The simple past tense has only two main functions according to Quirk and Greenbaum (1973:p.42), both of these being in relation to past time. The first use is to describe an action activity which took

place at a particular point in time, the second to describe an activity taking place over a period of time in the past but now finished.

An example of the first type of use mentioned would be:

125. I staked my claim back in '49

and the second;

126. I played for Rovers between 1983 and 1985

Palmer (1974:pp47-48) adds to this an 'unreality' use, an example of this being:

127. He acted as if he were already in charge.

Quirk and Greenbaum refer to this type of usage under the heading of 'mood', classing it as an instance of the subjunctive in English. Palmer (ibid) does not admit of English having a true subjunctive mood.

This unreality use of one of the two past tense forms of the copula can be contrasted with the use of the other form 'was' to refer to real conditions, for example:

128. He was already in charge

Finally, it must be noted that in fiction, the simple past may refer to future time, e.g.

129. In the year 2001, the starship Constellation left the planet earth on its way to another Galaxy.

#### 4.2.2 Aspect in English generally

So far in this chapter only simple VPs have been considered. These are those which are only marked for simple present or past tense. Complex VPs involve aspect and tense marking. Discussion of complex VPs therefore involves a description of aspect in English, and the way in which it may combine with tense.

Aspect in English may be perfective or progressive. Perfective aspect is marked in a VP by the presence of both the verb HAVE and the -ed participle. Progressive aspect is marked in a VP by the presence of both the verb BE and the -ing participle. Aspect; either perfective, progressive, or both of these, may be combined with tense to produce the following complex verb phrases:

#### present perfect

130. I have lost my pen

present progressive

131. I am working down in London (at present)

past perfect

132. I had been trying to unlock the door for 10 minutes, when I realised that it was the wrong key.

past progressive

133. I was driving down the high street, when a policeman stopped me.

present perfect progressive

134. I have been digging the garden, but I think I will stop for a break now

past perfect progressive

135. I had been watching the house for some hours, (when I suddenly saw something strange)

All the different combinations of tense and aspect set out above have different meanings, and these are not discussed.

The present perfect exemplified in example 130 above refers to a time period which began before the present time but which ends in the present. The time span can be very great, as in:

136. These rocks have existed since the world began

or very brief as in:

137. I have seen him just now

note that it is the adverbial which indicates the time scale and not the VP. If no adverbial is present as in example 130 above, then the time period referred may also relate to the future: The pen described as lost in example 130, if found will be found in the future. This point is made by Palmer (1974:p.36)

The present progressive is most often used for referring to present activity according to Palmer (ibid). It describes activity in fact which presumably started in the recent past and will continue for some indeterminate period into the future. The present progressive however may also be used, like the simple present, to refer to future time. However it is not always easy to see how the two forms differ if at all in meaning. For example in:

138. The ship {leaves from } Nice tonight on the evening tide.  
          {is leaving }

Quirk and Greenbaum (1973:p.49) claim that in such an examples, involving Dynamic Transitional verbs like; ARRIVE, LEAVE, and COME, there is in fact no difference in meaning between the present an the present progressive. Palmer (1974:p.66) however claims that there is a difference, and that this is a matter of intention versus schedule. For example:

139a. I am leaving Nice tonight

expresses personal intent on the part of the speaker, whilst:

139b I leave Nice tonight

assumes the existence of some pre-arranged schedule.

The past perfect tends to refer to a time period beginning in the past and ending at a later point in the past, labelled 'a point of relevance' by Quirk and Greenbaum (1973:p.45). In example 132,. above, this was the point when the writer discovered that he had been using the wrong key. The past perfect thus has the effect of setting an event in the past in a time context.

The past progressive may also has the effect of setting a past event in a context, but unlike the past progressive tends to refer to an activity that was taking place in the past but which was interrupted or curtailed when something else happened. This is the meaning

conveyed by example 133. above. When compared directly with the simple past tense, the past progressive is shown by Quirk and Greenbaum (1973:p.45) to have the ability to express incomplete action. To demonstrate this, compare:

140a. The ship was sinking

with;

140b. The ship sank.

Example 140a above allows that the ship was actually prevented from sinking, i.e. the act of sinking is not complete. Example 140a does not allow of this possibility.

The present and past perfect progressive forms predictably tend to combine the notion of limited duration or incompleteness associated with progressive aspect, with the notion of current relevance associated with perfective aspect. In example 134. above, on page 125, the activity involved in digging the garden is incomplete, but has been going on for some limited duration in the recent past. The point of current relevance is the present time when the speaker announces that he is interrupting his activity to have a break.

In example 135. above, again on page 125, using the past perfect progressive, all the action takes place in the past. The speaker

describes an activity of limited duration taking place in the past (watching the house) which was interrupted at a more recent point in the past. This more recent point is the point of relevance. In this particular utterance the point at which the speaker saw something strange going on.

#### 4.2.3 Mood in English Generally

This section is an examination of the role of the modal auxiliaries in the English verb phrase. The modal auxiliaries are small set of auxiliary verbs, used to express concepts like: possibility, ability, permission, obligation, probability, etc. In specialist texts within the literature, e.g. Palmer (1979:p.9) and Huddleston (1976:p.333), there is a good deal of argument about the precise linguistic definition of the term 'modal', and the exact set of verbs that make up this category. This thesis is concerned with primarily stylistic matters, and avoids discussion of these points by assuming the definition of 'modal' to include only those auxiliaries classified by Quirk and Greenbaum (1973:pp.52-28) as such.

The modals listed by Quirk and Greenbaum (1973:ibid) are:

CAN/COULD

MAY/MIGHT

SHALL/SHOULD



WILL/WOULD

MUST

OUGHT TO

NEED

DARE

CAN

CAN may be used to express; ability, permission, and theoretical possibility. Here are examples:

ability

141. I can type, but not very well.

permission

142. You can go now.

theoretical possibility

143. The loan can be extended if necessary

Quirk and Greenbaum (1973:p.52) note that CAN expresses a relatively informal sort of permission. Permission is expressed more formally by MAY, (below).

## COULD

COULD is formally the past tense of CAN, and can be used to express simply past ability. However it is not restricted to past time reference and can take on a wider range of meanings like; future permission, and both factual and theoretical possibility. COULD also allows the expression of 'contingent possibility' (Quirk and Greenbaum 1973:p.53). This is the expression of a hypothetical possibility, one whose truth has not yet been realised for some reason. Here are examples of all the meanings mentioned above;

### past ability

143. I simply could not make it work, no matter how hard I tried.

### future permission

144. Could I ask you to intervene in this matter?

### theoretical possibility

145. It's getting late now, the shop could be shut, or then again it might not be.

### factual possibility

146. We could go to the concert

contingent possibility

147. If we had a piece of bent wire we could pick the lock

MAY

MAY has a range of meanings which appear to closely parallel those of CAN. Like CAN, May is able to express permission and possibility. The possibility expressed by MAY however is usually factual rather than theoretical as in the case of CAN (Quirk and Greenbaum 1973:p53), for example:

factual possibility

148. The train may be a few minutes behind time (because of a delay)

The same utterance could bear the rather more unlikely interpretation that the Railway company may deliberately run the train later than usual in order to connect with another service.

The difference between the permission expressed by CAN and MAY however is held by Quirk and Greenbaum (1973:p.53) to be one of stylistic choice constrained by formality. For example:

permission

149. You may leave now

Compare example 149 with the less formal 'You can go'.

### MIGHT

MIGHT is reported by Quirk and Greenbaum (1973:p.54) as expressing both theoretical and factual possibility, and occasionally, permission. The theoretical possibility expressed by MIGHT seems intuitively much more tentative than that expressed by CAN. Compare for example:

150. The loan might be extended (in exceptional circumstances).

with example 143 above:

151. The loan can be extended if necessary.

The factual possibility expressed by MIGHT is likewise far more tentative than that expressed by MAY. Compare for example;

152. The train might be a few minutes behind time.

with example 148. above in which MAY replaces MIGHT in the same utterance. MAY in this utterance expresses a strong factual possibility. The use of MIGHT in the same situation seems to indicate a more hesitant expression of factual possibility.

The example given by Quirk and Greenbaum (1973:p.54) of MIGHT expressing permission is:

'Might I smoke in here?'

### SHALL

The use of SHALL to refer to future time has already been discussed (see sect 4.2.1 above). It also has a modal function, the main use of which is, according to Quirk and Greenbaum (1973:p54), to express intention, and this only in the first person. The same authors (op cit) also note a stylistically marked use of shall in legal and 'quasi-legal' texts. This is noteworthy as GIL texts have already been shown to behave similarly in some ways to legal texts (see previous chapter on the NP). Other restricted uses are; to express willingness (when used in the second and third person), and insistence. Examples of these uses are given below:

#### intention

153. I shall be there at six-o-clock sharp

#### legal use

154. All parties shall agree to abide by the conditions of the agreement.

willingness

155. They shall be free to choose their own destiny

insistence

156. You shall say you are sorry

SHOULD

Modal SHOULD displays a wide range of uses according to Quirk and Greenbaum (1973:p.55). These uses include; obligation, logical necessity, putative, contingent uses as well as the expression of 'formal real conditions'. Examples of these uses now follow:

obligation

156. You should listen to what your elders say

logical necessity

157. A first class stamp should make sure that they receive it on  
time

putative use

158. It amazes me that he should have been so bold

contingent use

159. I should be very pleased to meet him if he can manage to come

formal real conditions

160. Should you require any further assistance, do not hesitate to write.

WILL

Modal WILL is able to express very much the same range of meanings as modal SHALL, that is; willingness, intention and insistence. It also has, in addition, the power to express predictions of various types. The 'willingness' use of WILL is often used in polite requests according to Quirk and Greenbaum (ibid). For example:

161. Will you have another slice of cake?

The use of WILL to express intention parallels that of SHALL in reference, but seems much less formal. Compare for example:

162. I shall be there at seven.

with:

163. I will/'ll be there at six.

The insistence use of WILL parallels the insistence use of SHALL, and there appears to be very little difference between the two. for example:

164. He shall/will do as I say.

In speech however, will would need emphatic stress in order to convey insistence.

The predictive uses of WILL are also expressable by other means. In specific predictions, WILL refers to logical necessity in the same way as that expressed by SHOULD above. For example:

165. The parcel will have arrived by now

In 'timeless predictions' the use of WILL parallels that of the simple present tense:

166. Water {will evaporate} when subjected to heating.  
          {evaporates    }



WILL may also be used to express habitual predictions, for example:

167. They will drive you round the bend if you don't watch them

### WOULD

Formally the past tense of WILL, WOULD can be used to express insistence like WILL, but in the past, for example:

168. You would decide to do it your way, it's no wonder it ended in disaster!

WOULD may also however refer to present time and has a use similar to that of WILL to express willingness, e.g.

169. Would you open the window?

This is perhaps even more politer than the use of WILL in:

170. Will you open the window?

the choice between the two being perhaps determined by status constraints.

Like WILL in example 168 above, WOULD may also be used to refer to habitual activity, though the WOULD, referring as it does to past

and not future time like WILL, does not have the strength of a prediction. For example:

171. He would go for a walk round the town every evening just as it was getting dark

WOULD also has a contingent use like SHOULD above, and it is very difficult to discern any difference of meaning between the two. Compare WOULD and SHOULD in:

172. I should/would love to go to the party if I could.

SHOULD in this context may be being replaced by WOULD and therefore be stylistically related to time, intuitively SHOULD appears more formal than WOULD.

Finally WOULD can be used to express probability, usually when referring to a past event. For example, the utterance:

173. That would be the manager you are talking about.

might follow a conversation in which a speaker described a person he did not know, and the listener responded with a guess as to the his identity.

## MUST

Modal MUST is used to express both obligation, and logical necessity (like SHOULD), for example:

### obligation

174. All residents must vacate their rooms by eleven a.m. on the day of departure.

### logical necessity

175. There must be some rational explanation for his disappearance

Note that the obligation and logical necessity expressed by MUST can also be expressed by HAVE TO. The latter form seems intuitively less formal than MUST, though this is not mentioned by Quirk and Greenbaum (ibid) Quirk and Greenbaum (1973:p.56) observe that HAD TO replaces MUST when expressing obligation in the past, e.g.

176. We had to leave early because of the bad weather.

Palmer also notes (1979:p.65) that in the negation of obligation, NEED NOT has to be used, as with modal MUST, 'not' negates not the modality but the event, for example, in:

177. You must not tell Agatha.

the meaning conveyed is that there is an obligation on the listener not to tell Agatha, whilst in:

178. You need not tell Agatha.

it is clear that there is no obligation on the listener to tell Agatha.

In a similar way, **MUST** cannot be used to express the negation of logical necessity, **CANNOT** has to be used instead; **MUST NOT** expresses an obligation not to do something, for example:

179. This cannot be the case.

180. This must not be the case.

#### OUGHT TO

**OUGHT TO**, like **MUST**, also expresses obligation and logical necessity. Quirk and Greenbaum (1973:p.57) note however that these modals are not in free variation, **OUGHT TO**, both **OUGHT TO** and **SHOULD** expressing a 'less categorical' logical necessity and obligation than **MUST**. This becomes clear when all forms are compared:

### obligation

181. You ought to/should go to the dentist.

182. You must go to the dentist.

### logical necessity

183. There ought to/should be a way to solve this.

184. There must be a way to solve this.

DARE is not described as a modal here since it does not occur in GIL texts.

This section on Mood ends the discussion of the Verb Phrase in English generally. The section which follows will examine the use made by GIL texts of the patterns so far described.

## 4.3 The Verb Phrase in GIL texts

### 4.3.1. Tense

GIL texts are characterised by a relative scarcity of full finite VPs. This is no doubt partly due to the fact that many of the VPs in In GIL texts are operating in non-finite postmodifying clauses, (see

earlier chapter on the Noun Phrase), or as non finite dependent clauses in sentence structure (see Chapter 6 on the Sentence). A subjective analysis of the proportion of finite to non-finite clauses in GIL texts, suggests that both are equally well represented, giving about 50% of each type. Non-finite VPs do not show tense, and as a result the potential stylistic significance of tense in GIL texts is perhaps rather limited.

However it is quite clear that GIL texts are very conservative in their use of both tense and aspect, making extensive use of the simple present to the exclusion of most other forms. This tendency appears to be true for both I.A. and P.A. texts.

The main use of the simple Present in English generally is held by Palmer (1974:p.60) to be to refer to habitual activity, as reported on p.119. The simple present is used to fulfill this function in GIL texts though it is clearly not its main function. GIL texts have most need of reference to habitual activity when asking claimants questions about their income, hours of work and so on:

185. Are these earnings what you normally receive?

4

186. How many hours a week do you normally work.

The habitual use of the simple present often appears in this form, where the emphasis is on the claimant providing information on

his/her habitual and therefore normal circumstances, this is obviously because benefits are worked out on the base of normal average weekly income. Another use is to describe invariant rules, like:

187. Light goods vehicles pay a single standard rate of duty.

or alternatively to state a general rule with exceptions, for example:

188. Light goods Farmer's vehicles also pay a flat rate of duty, but at a concessionary rate

Palmer explains (1974:p.60) that in principle what the simple present does is to merely report. He appears to argue that this use however is not the most important in English generally, for the simple non-linguistic reason that if the speaker can observe an event or activity, so can the hearer. There is therefore no reason to report it. This argument does not hold however for GIL texts where the writer and reader are generally separated by space and time from one another, having in addition only a minimum of shared knowledge, especially in P.A. texts. The simple reporting function of the simple present in GIL texts is very well used in GIL texts and is, possibly, its main function. Leaflets abound with examples like:

189. This note explains the position regarding losses.

190. Part time means less than 30 hours a week.

191. This change puts lorry taxation on a much fairer basis.

192. These notes do not of course tell you everything about income tax.

The reporting function of the simple Present is important in GIL precisely because its information giving texts need frequently to point out things that the reader might not otherwise be able to make out for himself; 189 above, explaining the meaning of a note for example, or 190. making quite clear what the department means by 'part time work'. 'Part time' is a term with a rather vague definition in English generally, and would be open to a variety of interpretations.

Example 191. departs from the pattern formed by the remaining examples, in that it does not report fact, but the departments own (favourable) interpretation of the legislation which it is explaining. This was not observed to be a widespread function of the simple Present in AdE, but shows linguistic evidence of the governments need to make sure that new legislation is favourably accepted by the public.



Because many of the verbs commonly used in GIL, e.g. BE, HAVE, LIVE, APPLY, OWN, DEPEND, are concerned with constants regarding peoples lives, about which the administrator often wants to know, these tend to be stative. The simple present in GIL often therefore expresses duration rather than habitual activity. For example;

193. Is your permananent home in the U.K.?

194. Are you a lone parent?

195. Do you have a FIS book already?

196. Write down all the children who live with you

197. Does either of you live at a different address

198. Where no fixed reduction applies, state nature and amounts

199. You cannot claim these expenses if you own your own home.

200. The duty payable depends on the gross train weight.

It was noted when describing the uses of the present tense in English generally (p. 121), that the simple present in temporal and conditional subordinate clauses, frequently refers to future time. This type of future reference is well used in GIL, though the use of

the simple present to express an unconditional future is not. The reason for this appears to be the need for the administrator to foresee possible future problems and cater for these eventualities by what he writes in his text; for example:

201. If you need help or further information please ask me

202. Unless the vehicle qualifies to be admitted duty free, it will normally be liable to duty as shown in part III.

203. Release from these conditions will only be allowed if the customs charges are first payed

204. If the person is not present at the time of the importation, a written declaration is required of him

205. Have a look at all the self-service vacancy boards, if you're not sure which types of work might suit you, or if you decide to do something different.

In these conditional clause examples, the future time reference appears to vague and imprecise. It might be expressed by saying, 'if at any time in the foreseeable future...'. In the temporal clauses that follow reference is to some process or procedure which is contingent on some future event. Again the precise time in the future is not specified, because it depends on the future action of

the reader (referred to as 'the owner' and 'the importer' in these I.A. leaflets).

206. Grant of the relief will be considered when the owner arrives

207. When all customs duty and/or taxes due are paid, the importer will be given a customs form

208. The applicant should fill in and sign the two enclosed forms and should attach them to the completed application form when they send or take it to the passport office for their area.

Since so much of the future for the administrator will only be decided by what action individual members of the public take, it is not at all surprising that the simple present to refer to an unconditional future is little used in GIL texts: Earlier, on p. 121, the use of the simple present in main clauses was observed to be confined to events which the speaker could be certain would take place.

The performative use of the simple present has a small but important part to play in GIL texts. This is indirectly the result of the fact that many GIL leaflets, besides consisting of explanatory text, often include an application form as well (if the benefit or service described is one that can be applied for). This happens most often in leaflets of the Department of Health and Social Security, and the

Department of Employment, who are responsible for administering most state monetary benefits in the U.K.

Personal communication with the National Association of Citizen's Advice Bureaux revealed that such forms, when completed and sent to the appropriate department, have by law to be processed. This is a legal obligation which is matched by a claimant's legal right to seek redress if his claim does not receive attention. A consequence of this is that the properly completed form constitutes a claim for benefit in law, and is in consequence a legal document. It is usual therefore for the form to end with a space for the claimant's signature and a statement of which the following are examples:

209. I/We declare that I/We have read the instructions on the form..

210. I/We claim Family Income Supplement.

211. I certify that this is a true likeness of M....

The performative use of the simple present is therefore a simple pro-form, an unsigned declaration which becomes active the moment the claimant signs the form. These performative statements are often accompanied by warnings, for example:

212. WARNING: To give false information may result in prosecution.

The performative use of the present again demonstrates how close the administrator's task is to that of legal practitioner's. Much of the administrator's work is covered by a complex of legislation to which he must adhere. This in turn places constraints on administrative texts, so that inevitably at times, as with performative declarations, and warnings, they assume quite a formal style. This no doubt accounts for some of the criticism levelled at leaflet writers for the way in which they 'intimidate' the public, e.g. Vernon (1980:p.35):

'...people under what is likely to be the most telling emotional strain of their lives deserve more consideration than the WARNING in the declaration at the bottom of the form.'

The simple past tense is rather rare in GIL texts. The administrator is generally unconcerned about past events but rather with present circumstances of claimants etc. The main exceptions to this rule are tax and pension matters, where age and past incomes are obviously of some importance. The use of the past tends to occur in interrogatives, for example:

214. If you were born before 6th April 1924, please enter date(s) of birth

215. If you started this job less than 5 weeks ago, give the starting date

The simple past is also used in GIL to set out information which is conditional on something having happened in the past. This is another way of dealing with the type of situation already referred to in future conditional clauses, where the event is seen as potential but not yet having occurred, for example:

216. A vehicle is admissable free of customs duty provided that:

(a) it was previously exported from the EEC (and)

(b) it was not exported under.....

The unreality use (see p.123) of the past tense is occasionally used in GIL texts to cope with rather unreal or hypothetical situations which do not in fact hold at the present time. For instance the situation of two single people wanting a joint passport for a honeymoon abroad after they are married, with consequent name changes. The Passport Office copes with this in the following way:

217. The applicant should complete this in full as if they were already married.

Having discussed the simple present and past tenses, the next section will deal with progressive and perfective aspect in GIL.

The simple past is also used in GIL to set out information which is conditional on something having happened in the past. This is another way of dealing with the type of situation already referred to in future conditional clauses, where the event is seen as potential but not yet having occurred, for example:

216. A vehicle is admissable free of customs duty provided that:

(a) it was previously exported from the EEC (and)

(b) it was not exported under.....

The unreality use (see p.123) of the past tense is occasionally used in GIL texts to cope with rather unreal or hypothetical situations which don't in fact hold at the present time. For instance the situation of two single people wanting a joint passport for a honeymoon abroad after they are married, with consequent name changes. The Passport Office copes with this in the following way:

217. The applicant should complete this in full as if they were already married.

Having discussed the simple present and past tenses, the next section will deal with progressive and perfective aspect in GIL.

#### 4.3.2 Aspect in GIL

When discussing tense in GIL, it was observed that very little use was made of the simple past tense. This is partly because administrators deal mainly with current events and their possible future consequences, but also because when occasionally past time is referred to, the present perfect is used. The present perfect which expresses a past that is currently relevant is better suited to the constraints of the administrative province, than the past expressed by the simple past tense which merely reports a past event. Past events are hardly ever referred to by the administrator unless they are currently relevant, for example to a claim in progress. Examples of the present perfect in GIL are:

218. If NO, and you haven't claimed FIS before, please enclose...

219. The notes are there to help you, but they have been kept short.

220. How have taxation classes changed?

221. Qualifications: What exams have you passed?

Progressive aspect is very poorly represented in the data, it is only rarely that GIL needs to express duration through use of the progressive, either past or present. It has already been observed when discussing the simple present in GIL that the stative nature of



many verbs used ensures that duration can be expressed without using the progressive. The present progressive is however sometimes used to express limited duration, for example:

222. If you are experiencing employment problems....

223. Put YES if you are bringing children up on your own.

224. Put NO if you are living with someone as husband and wife.

225. If the person on whose behalf a vehicle is being temporarily imported.....

227. Could you take a stop-gap job while you are looking around

Note that it has to be admitted that 'limited duration' in some of these examples is a matter of opinion: For example whilst an administrator might class 'living together' as a temporary relationship, a couple to whom this applies might consider themselves to have a permanent relationship.

In conclusion GIL makes very little use of aspectual contrasts. Subjectively, the vast majority of finite VPs in GIL appear to involve the Simple Present, with its many uses. GIL does however make very interesting use of the system of English Modals. Mood in

GIL is discussed in the section that follows.

### 3.3.3 Mood in GIL

The main use of modal CAN in GIL is in P.A. texts. On page 130. three possible uses of CAN were noted: ability, permission, and possibility. Simple expressions of ability are very rare in GIL, but both possibility and permission uses are found very frequently. It is however rather difficult in GIL to decide to which of these two a particular use of CAN belongs. Consider the following examples:

228. You can ask to see one of our employment advisers.

229. You can get further information and leaflets from our offices.

230. Alternatively an application can be made with the assistance of the personal application branch.

Whilst example 229. seems to indicate simple possibility, both examples 228. and 230. seem to confer a right on the reader to use a particular departmental service, as well as expressing the possibility of obtaining departmental help. This seems to indicate that the permission and ability uses of CAN are not distinct as has to be tacitly inferred from Quirk and Greenbaum (op cit), but may both be present at the same time. Palmer (1979:p.61) supports this

idea, noting:

'...to say what is possible is often to say that the speaker will not object, i.e. that he gives permission'.

and even goes on to suggest (ibid:p.149) that there is a four level gradation in the Province of Rules and Regulations (to which GIL might be held to belong) between granting permission and saying what is possible. Palmer's gradation of permission and possibility in Rules and Regulations is set out below:

1/ Giving permission

2/ Reporting rules and agreeing with them

3/ Reporting rules

4/ Saying what is possible

It seems misleading to the present author to describe the the relationship between possibility and permission in Rules and Regulations in terms of discrete levels. There is insufficient evidence to allow this. The gradation is perhaps better seen as a cline. Subjectively, it seems that the meaning expressed by CAN in P.A. GIL texts seems nearer the bottom of this cline (i.e. the possibility end) than the top. There will be need to refer again to

this cline when discussing Modal MAY.

COULD

Of the 5 possible uses of Modal COULD listed earlier on page 131, it is the theoretical possibility, and contingent possibility uses which are most used in GIL. For example:

theoretical possibility

231. You could get another £3.65 a week on top of your Child Benefit.

232. You could still be entitled to benefits worth £££s.

233. 'Yes' answers to 3 questions mean you could get FIS.

contingent possibility

234. Ask yourself what you could do if you had the chance to try.

235. But local jobs are not always easy to find. Are there jobs further away you could take

236. Could you take a stop gap job?

Subjectively, COULD seems far more frequent a modal in GIL P.A. than in I.A. This may be a function of the type of use to which this modal is put in GIL. The theoretical possibility use of COULD tends to emphasise what might be in store for the potential claimant of a service or benefit. For example the possibility of being entitled to £££s in example 232 above. This usage of COULD parallels the frequent use COULD to express theoretical possibility in some forms of advertising, for example:

237. Shopping at XYZ superstores could save you £££s .

and the following example from Crystal and Davy (1969:p.223):

238. 'Could' it be your after-dinner coffee that keeps you awake at night

Whilst this usage of COULD is appropriate in an introductory text of the types usually found in P.A. GIL, it would be inappropriate in I.A. where it has already been observed (in discussing the NP) that the reader is already expected to have at least some familiarity with the subject matter, and possibly already has a good reason for seeking more information. The reader of an I.A. text therefore, presumably does not need any further incentive to take further action after reading the text, whilst the P.A. text reader might be expected to need the encouragement provided by the use of COULD.

The contingent possibility use of COULD appears to function in GIL

make friendly suggestions to the reader about possibilities which may be available to him if only he thought in a different way. Example 235 on page 156 above for example endeavours to encourage the reader to look further afield for jobs. The friendly tone of the suggestion made with COULD in P.A. GIL would appear to be a function of Status constraints. P.A. texts have already been observed to be less formal than their I.A. counterparts. COULD would not appear to be as stylistically appropriate in I.A. precisely because these texts are more formal in terms of status

#### MAY

It was noted above on page 132 that for English generally, MAY parallels CAN in allowing expression of permission and possibility, though the possibility expressed by MAY tends to be factual rather than theoretical as in the case of CAN. This was observed to be a trend rather than an absolute rule. In expressing permission, MAY was observed to be more formal than CAN.

As might be expected from the discussion of CAN above, MAY in GIL does not tend to distinguish clearly between the expression of possibility and permission. However unlike CAN which was observed to express a meaning closer to possibility, on the cline between what is possible and what is permitted, MAY in GIL seems to express a meaning closer to permission than possibility, for example:

239. Tax on the benefit obtained from the exercise of a share option, may in certain circumstances be paid in instalments.

240. You may claim for any necessary travelling expenses which you had to pay in carrying out your work

241. An employer may make a contract with agricultural workers engaged in husbandry to provide food, drink (not intoxicating), cottage or...

242. The court may remit or postpone the payment of the fee or part of the fee where in the opinion of the court hardship might otherwise be caused to the patient.

MAY appears to have a meaning closer to that of permission, because every one of the examples above seems to refer to either departmental rules and regulations, legislation, or at least departmental policy. However, note that all these examples are ambiguous. The reader cannot possibly work out for himself, without consulting the department concerned, whether the texts above indicate permission or simple possibility.

GIL does have instances where MAY quite clearly expresses possibility rather than permission, such examples are however fairly rare. here are some examples:

243. Some wife's earned income allowance may be due.

244. It may be to your advantage not to claim the married man's allowance.

There is likewise at least one instance in GIL where MAY clearly expresses permission. This is when MAY occurs in a VP together with the verb CLAIM. Evidence for this is that when GIL writers wish to express possibility with CLAIM the VP must also contain BE ABLE TO. The permission expressed by MAY with CLAIM refers in all probability to legislation covering the benefits or services to be claimed. Examples of MAY with CLAIM are given below:

Permission use:

245. You may claim for

EITHER a relative...

OR

246. You may only claim for the contribution which provides death benefit.

247. You may claim if your wife is a registered blind person.



Possibility use:

248. If you made a loss in lettings, you may be able to claim tax relief

249. You may be able to claim relief for tax paid overseas

250. If you or your wife were born before 6th April 1919, you may be able to claim

In view of the observation made earlier on page 156 that MAY is more formal than CAN when expressing permission, it is not surprising that the more formal (in Status) I.A. texts in GIL make far more use of MAY than they do of CAN. Subjectively, the reverse seems to be true for P.A. texts. The notion of MAY expressing permission being more formal than CAN used for the same purpose is well supported by Palmer (1979:p.60). Palmer observes:

'it seems fairly clear that MAY is far more formal than CAN; in the MAY example above (not quoted here), the situation is that of a trial. This is supported by Ehrmann (1966:p.12) who noted that most of the occurrences of CAN for permission were in dialogue.'

MAY and CAN are not however mutually exclusive in I.A. and P.A. GIL, It is however clear that a GIL text containing frequent instances of

MAY is likely to be impersonal in address, and one containing frequent instances of CAN, personal in address.

### MIGHT

Modal MIGHT in GIL is used mainly to make tentative statements about what is possible. for example:

251. If you had a well paid job before, you might well be expecting too much pay.

252. You might be suitable for a TOPS training course

Like Modal COULD, MIGHT in GIL involves the expression of a friendly suggestion to the reader again about tentative possibilities. Like COULD, MIGHT tends to be confined to P.A. texts, and like COULD its usage in GIL would appear to be constrained by the need for less formality in P.A. texts.

### SHALL

GIL texts, occasionally make formulaic use of modal SHALL to express willingness, for example in:

If you need more help or information I shall be glad to give it

However modal WILL is more frequently used for this task, perhaps because SHALL in this context has a formal and perhaps old-fashioned stylistic meaning in addition to its referential concept of willingness. Otherwise SHALL is completely absent from GIL texts, in spite of the legal and quasi-legal uses reported by Quirk and Greenbaum (1973:p.54), and Palmer (1979:p.63). This absence of a modal widely used in legal English seems odd in GIL when it is considered how close legal and administrative provinces have so far proved in this thesis. SHALL does have other uses besides that of saying what the citizen should and should not do, in legal documents. These uses are reported fully on page 134, but can be summarised here as: intention, willingness, and insistence. None of these uses however occur in GIL.

One possible explanation for the absence of SHALL in GIL is that it has a connotational meaning of authoritarian discipline through its well known legal use, in addition to any referential meaning. GIL has been, and still is the subject of much public criticism, not least from the government where administrative language was described in a recent White Paper (Cmd 8504) as: 'lengthy, legalistic, and intimidating'. It is perhaps not surprising then that GIL writers may wish to avoid some words with legal connotations.

A second, perhaps simpler, explanation is that GIL text is not composed of rules and regulations, but a report on rules and regulations. It is not the duty of Government Departments to

make clear what a citizen shall or shall not do, but to point out what the law says, this function of the administrative process is possibly better conveyed by the use of Modal MUST which expresses duty rather than a strict obligation, or modal SHOULD which makes a more tentative expression of obligation. This interpretation is to some extent supported by the more frequent use of MUST and SHOULD in GIL texts (see later).

### SHOULD

SHOULD is frequently used in both I.A. and P.A. texts in GIL. SHOULD, like SHALL, may express obligation, and it is this particular use which is prominent in GIL texts. Palmer (1979:p.69) claims that whilst the writer using SHOULD is imposing an obligation on the reader, its use allows for the fact that the obligation may not be fulfilled. This would be appropriate when the writer was not himself able to enforce an obligation. Government Departments are not law enforcement agencies, but one of their prime aims is to inform the public of their duties. This appears to be one of the main functions of GIL texts. A possible explanation of SHOULD's widespread use throughout GIL might be that SHOULD fulfills this function by allowing the reporting by a government department of what legislation involves, and enabling it to recommend to the public what they should do.

Here are some examples of SHOULD in use in GIL:

253. You should include all the interest

254. If your pension is received 4 weekly or quarterly, the amount entered should be based on the weekly rate of pension.

255. The following are NOT taxable and should not be included on your return:

Alternative explanations of the use of SHOULD in GIL might be that either, in common with COULD and MIGHT it is a tentative and therefore a more friendly form than SHALL, (though this would not explain SHOULD's common use in the more formal I.A. texts), or that SHOULD is actually being used where no legal obligation exists but the department concerned still discerns the need to place obligation on the public. It is possible that all these factors in some way constrain the use of SHOULD in GIL.

#### WILL

In spite of the range of meanings ascribed to WILL on page 136 above: willingness, intention, and insistence, this modal appears to be reserved largely for reference to future events in GIL, though the willingness use is sometimes encountered, for example in formulaic statements like:

256. I will provide details on request

found in many texts.

Future reference may involve a neutral future, as for example in:

257. Please enclose birth certificates or some other proof of the child's age. These will be sent back to you.

258. You should include all the interest. I will allow any exemption due.

WILL however is more frequently used in GIL to express a future that is conditional on some event, usually action taken by the reader, for example:

259. If you have just started work, you can still claim now and we will ask your employer for details.

260. If a relative feels that it would be in the patients's interests, no objections will be raised by the court.

261. If you get married after the first tax month you will only get part of the allowance.

#### WOULD

WOULD is a rarely occurring modal in GIL texts. Of the range of uses

outlined on page 138 (above), only that expressing tentative willingness is used in the data. The form is confined largely to interrogatives in P.A. texts, for example:

262. Would you be prepared to move somewhere else, even to a different part of the country?

The use of the tentative modal WOULD only in P.A. texts appears to match the similar use of COULD (see earlier in this section). The use of WOULD in the example above appears to express a friendly suggestion. The choice of WOULD in P.A. texts appears therefore to match the informal Status already observed in P.A. texts.

#### MUST

MUST is widely used in both types of GIL text. MUST in GIL expresses obligation on the reader, but it is not always clear who exactly is imposing this obligation, the law itself, or the department concerned. This problem can be clarified by moving beyond Quirk and Greenbaum's (1973:p.56) observation that MUST simply expresses obligation. Palmer (1979:p.61 & 91), observes that the obligation imposed by MUST may be either neutral, expressing dynamic necessity, or oriented towards the subject, who takes responsibility for imposing the obligation (deontic necessity). In GIL there are sometimes cases where it is clear that the obligation is being

reported and therefore expresses dynamic necessity, for example:

263. In the 1896 Act there are certain additional provisions about deductions or payments for the subjects mentioned in paragraph 16. These are:

(1) ...

(2) Bad work or damage to the employers property

Any deduction or payment must be reasonable and must not exceed the actual cost to the employer caused by the worker.

In this example it is quite clear that the department is merely reporting an obligation to the reader. This is a well established function of some types of GIL text where the original legal text on which the leaflet is based is obviously not intended to be read by the average member of the public, it frequently occurs in publications by the Department of Employment which try to set out in simple terms how the law affects people at work.

The general tendency in GIL however is for uses of modal MUST to be ambiguous on the point of who takes responsibility for imposing the obligation, for example:

264. If you are a married man and your wife is living with you, you must show all her income.



This example could easily be read in two ways as follows:

(1) The department requires that you show all your wife's income.

(2) The law obliges you to show all your wife's income (you may be prosecuted if you don't).

Similar dual interpretations can be made of:

265. The relative must be maintained by you.

266. To take advantage of this provision an election must be made.

There is no obvious way of solving these ambiguities for the reader, though for the stylistician, this use of MUST in GIL along with the permission uses of CAN and MAY, provide interesting information on the situational constraints that govern GIL style. These constraints arise because of the unclear status of government departments, which while instruments of government authority are not that authority themselves.

#### UGHT TO

Modal UGHT TO is notably absent from GIL. It was noted on page 141, earlier that both SHOULD and UGHT TO may express hesitant

obligation. However only SHOULD is used in GIL texts. Existing stylistic information about the distributions of these two modals is scant, but subjectively it seems possible that OUGHT TO might be more common in conversational than in formal written English. It is not possible for this thesis to investigate such a general hypothesis, but nevertheless this could provide a starting point for further research.

The discussion of OUGHT TO completes this section on MOOD in GIL, it simply remains to summarise the points made in this chapter before concluding and moving on to discuss GIL clause structure in Chapter 5.

#### 4.4 Summary

In summary of the VP in GIL, GIL makes a very narrow selection from the tense and aspect system of English, using mainly the simple present tense throughout. This seemingly simple situation is more complex on further examination as GIL makes very full use of most of the different uses of the simple present. Unlike English generally, the main function of the simple present in GIL is not to refer to habitual activity, but to report, the simple present is also widely used in temporal and conditional clauses to express a conditional future usually dependent on action to be taken by the reader. The performative function of the simple present has a small but important role to play in declaration sections in GIL texts, and

suggests a possible further parallel with legal English.

The simple past is not frequently used in GIL, its use being confined to pension and tax leaflets, where age and the previous year's financial status have importance. The use of the past tense to express conditionality, dependent on past events, parallels the conditional future use of the simple present.

The use of present perfect is observed to be the most frequent method of referring to past events in GIL, this is appears to be related to the present perfect's ability to express a past with current relevance. This tense/aspect combination is seen to suit the constraints of the administrative situation in which the only reason to refer to the past appears to be if it is relevant to a current problem or claim.

GIL poses distinct problems for the neat description of mood. It is seldom possible with modal CAN, MAY or MUST to say that the meaning expressed in GIL corresponds exactly to traditionally defined categories like 'permission' and 'obligation'. GIL texts are frequently ambiguous to the extent that it is often difficult to pinpoint who is taking responsibility for granting permission or imposing obligation, the department concerned or the law. This failure to be able to classify modal uses accurately in GIL provides an insight into the status constraints which operate in the administrative situation: whilst government departments serve

authority, they are not that authority itself.

Finally, the tentative modal forms COULD, WOULD, and MIGHT, are observed to be unevenly distributed in GIL, occurring more frequently in the (less formal) Personal Address texts. These modals are perceived subjectively to be more friendly and less formal than the corresponding forms; CAN, WILL, and MIGHT. Their more frequent use in P.A. GIL than in I.A. is held to be due to status constraints. P.A. texts are introductory texts written with the intention of encouraging readers to claim benefits or find out more about a particular service, this purpose requires a relatively informal style.

## PART 1 - CHAPTER 5 - THE CLAUSE

### 5.1 General

This chapter provides a description of the Clause in English generally and in GIL, however Clause structure will only be described here in the areas not covered by other chapters. For example, little attention will be given to what NPs typically function as subject or object in English generally and in GIL; this subject has already been well discussed in Chapter 3 on the Noun Phrase. Also, being restricted to clause structure, this chapter will not describe different clause types (main and subordinate) and the way in which they may join to form sentences. This will be dealt with in Chapter 6 which follows.

There is one element of sentence structure which does come within this chapter however, and that is the simple sentence, since a clause may function on its own as a simple sentence. In fact some authors, including the ones on whose work the description of the clause in this chapter is based (Quirk and Greenbaum:1973), title their chapters on clause structure; 'The Simple Sentence'. For the purposes of description the structure of the clause and of the simple sentence are identical and can be described under the headings of: Subject, Verb, Object, Complement, Adverbial, and vocative, In the sections which follows, these will be referred to in abbreviated form as follows: S, V, O, C, A, voc, for example:

268. In the meantime John had phoned Bethan

{ } { } { } { }  
A S V O

269. Cedric is an absolute idiot

{ } { } { }  
S V C

269A. You can't do that, Harold!

{ } { } { } { }  
S V O voc

Each of these categories of clause structure will be the subject of separate discussion in the description that follows. In addition to this, this chapter examines what Quirk and Greenbaum (1973:p.168) refer to as 'relations of grammatical paraphrase', in a discussion of Voice (passive and active) in English generally and in GIL.

## 5.2 The Clause in English

### 5.2.1 Subject

The subject of a clause is an obligatory part of any clause (see Quirk and Greenbaum 1973:p.167) and is usually an NP in English. In declarative clauses the subject occupies initial position in the clause. The verb of the clause must agree in person and number with the subject. The NP occupying subject position may range in length

from a single word, for example, a pronoun, to a heavily modified NP, for example an NP with several premodifiers and postmodifiers. Examples of possible subject NPs might be:

pronoun

270. He won't succeed

noun

271. Reginald can't possibly come.

premodified NP

272. The automatic on-line data processing unit has stopped working

postmodified NP

273. Facilities for disabled people are available.

Pre- and post-modified NP

274. The new unemployment and debt advice centre manned by volunteers and funded by the local council is now open.

occasionally, elements of structure other than the NP may occur  
Subject, for example:

### Verb

275. Skidding is to be avoided at all times.

### Prepositional Group

276. In the Pyrenees will suit me fine for this summer's holiday

### Clause

277. Smoking of cigarettes is prohibited within this area.

#### 5.2.2 Verb

The verb position in clause structure must be obligatorily filled, and may only be occupied by a verb phrase. The verb must agree in person and number with the subject of the clause. The verb phrase itself has already been discussed in the previous chapter (chapter 4), and will not be further discussed here.

#### 5.2.3 Object

The object is not an obligatory part of clause structure except where the verb position of the clause is occupied by a transitive verb, e.g. HIT, (where an object is required to complete the meaning of the verb). Object position in the clause, as with the subject, is generally occupied by an NP, but may be filled by the same



structural items as also occasionally operate at subject position.

See above for these structural elements.

A clause may have more than one object, generally referred to by the labels 'direct' (D.O.) and 'indirect' (I.O.). An example of a clause containing both a direct and an indirect object might be:

278. Hywel gave Harri the news last Saturday

S    V    I.O.    D.O.    A

Note that in the passive paraphrase of an active sentence the object becomes subject, for example:

279. Harri was given the news by Hywel last Saturday

S        V        O

#### 5.2.4 Complement

Complements may be either a noun phrase, adjective, or clause. They must be co-referential with either subject (subject complement: S.C.), or object (object complement). Here are examples of both types:

#### subject complement

280. Branwen is an idiot

S    V    S.C.

object complement

281. They considered Stanislavsky a fraud.

S            V                    O            O.C.

Note that a complement cannot become the subject of the passive paraphrase of an active clause.

5.2.5 Adverbial

The adverbial component of a clause is generally optional, though with some verbs there may be an obligatory adverbial, for example with SEEM:

282. Cuthbert seems in high spirits

S            V                    A

The adverbial may take the form of a prepositional group as in the example above. It may alternatively be realised by an adverb or an adverbial phrase, and occasionally a noun phrase For example:

adverb

283. He normally never forgets

S            A                    A            V

adverb phrase

284. I will have to consider your offer very carefully

A

noun phrase

285. I have been writing this chapter all week

A

Note that, as exemplified by example 283, a clause may contain more than one adverbial.

Perhaps one of the most stylistically interesting aspects of adverbials, as well as their inclusion or non-inclusion, is their mobility within the clause. Adverbials may occupy initial, medial or final positions in the clause, for example:

initial

286. Studiously, he turned from one page to another.

medial

287. He was impatiently stamping his foot.

(interrupting the V.G.)

final

288. I shall reply to your letter without delay.

#### 5.2.6. Vocative

The vocative is an optional clause element which addresses the intended recipient of the information in the clause, for examples:

289. Don't forget you gloves, Mary!

#### 5.2.7 Linking and Apposition

In the examples discussed so far, each clause element has generally been realised only by a single word or phrase. This need not necessarily be so, many elements of clause structure can be realised by more than one phrase. These phrases may be linked to each other or in apposition. More detail is now given on these relationships.

Here is an example of linking at subject position:

290. Our language, our culture, and our heritage must not be allowed  
to disappear.

The example above contains 3 NPs at subject position, linked together by commas and 'and'. Here is an example of apposition at subject position.

291. Jesse James, a latter day Robin Hood was famous for his wild exploits.

In this example the two elements of the subject are co-referential, the second noun phrase in the subject provides additional information about the first, either phrase could be deleted and still leave a whole clause. Apposed subjects may also appear in brackets:

292. Form 16B (application for a Zebra licence) is obtainable at most of our offices.

#### 5.2.8 Passive and Active clauses.

The grammatical paraphrase relationship between active and passive clauses has its basis in the following relationship: A clause containing a noun phrase as object; for example:

293. Esmerelda slowly ate the obscenely huge banana split

S	A	V	O
---	---	---	---

may become a passive clause in which the object NP becomes subject, for example: !

294. The obscenely huge banana split was slowly eaten by Esmerelda.



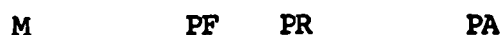
The subject NP of the active clause may either appear in a phrase with 'by' as above in 294, or alternatively be omitted completely, still leaving a complete clause, for example:

295. The obscenely huge banana split was eaten.

where the only marker of the passive is the passive verb phrase consisting of the past tense of the copula + the passive auxiliary (the -ed infinitive form).

In complex VPs the passive auxiliary (PA) always occupies final position after modal (M), perfective (PF) and progressive (PR) auxiliaries, for example:

296. Simon will have been being interviewed by now.



Note that the semantic category of the agent of the action denoted by the verb remains co-referential with the subject NP of the active clause. Those passive clauses which retain the subject of the corresponding active clause in a 'by' phrase therefore retaining the agent. Those passive clauses which omit the 'by' phrase also omit

the agent. The latter type of passive will be referred to from now on as 'agentless'.

Palmer (1974:pp.86-87) discusses the 'meaning' of the passive. By this Palmer seems to be referring to a set of possible motivations on the part of the writer to choose a passive rather than an active clause. These motivations are said to vary according to whether the passive is agentless or not.

The motivations put forward by Palmer for using the agentless passive centre around the assumption that the writer does not want to refer to the agent, either because it is:

- 1) unknown, or
- 2) irrelevant, or
- 3) undesirable

Use of the passive accompanied by an agent can be motivated, according to Palmer (ibid) by the desire to give the object NP of the active clause more prominence by making it the subject of a passive clause, and/or to give the subject NP less prominence. This might be due to one of three reasons:

- 1) Because the writer considers the object NP more important and wishes to emphasise it.

- 2) In order to keep the same subject in successive clauses, for example to change:

297. Penelope began to read but Gerald interrupted her.  
to;

298. Penelope began to read but she was interrupted by  
Gerald.

- 3) In order to place a particularly long subject NP, or perhaps several co-ordinated NPs in a less prominent position, for example:

299. The motion was passed by a vote of 32 to 12 by a  
joint meeting of the administrative staffs of the  
two major firms involved in the venture.

This discussion of the relationship between active and passive clauses ends this section on the clause in English. The next section goes on to examine the nature of the clause in GIL.

### 5.3 The Clause in GIL texts

#### 5.3.1 Subject

On pages 174-175 above it was noted that the subject of a clause may



be realised by any one of four different grammatical categories: Noun phrase; Prepositional phrase; Verb or Clause. However it is extremely rare in GIL that the Subject is realised by anything other than an NP. Stylistically there is little remarkable in this since in English generally the NP appears to be the most frequent grammatical category found in subject position (Quirk and Greenbaum 1973:p.170). What is interesting to the stylistician however is the way that different types of subject NP are distributed in the Personal and Impersonal address texts of GIL. The distribution of NPs consisting solely of a single personal pronoun is particularly uneven, as well as those NPs whose postmodification is complex (involves more than one postmodifier).

As regards distribution of different NP types, the tendency in GIL appears to be for P.A. texts to favour NPs consisting of a personal pronoun only, and for I.A. texts to use more complex NPs, often co-ordinated or postmodified, and sometimes with complex postmodification. Subjective examination of sample texts in the data suggests that the ratio of pronominal NPs to other NP types in P.A. may approach 1 to 1 in some texts, though by no means all, whereas in I.A. texts it appears to be generally low (about 1 to 8).

Examples of typical Subject NPs from the data are given on the next page in the underlined portions of each clause:

Personal Address

300. You can ask to see one of our employment advisers
301. You could get another £3.65 a week on top of your child benefit.
302. We will ask your employer for details.
303. You must sign and date the form
304. You can get further information and leaflets from our offices

Impersonal Address

305. Tax on the benefit obtained from the exercise of a share option may in certain circumstances be paid in instalments.
306. Any deduction or payment must be reasonable.
307. The court may remit or postpone payment
308. The applicant must surrender with the application form, any previous passport held.
309. Customs duty, car tax and VAT must be paid at the port of entry.

Pronouns are semantically far less explicit than a full noun phrase, as both writer and reader must be aware of what the pronoun used refers to. In P.A. texts although the referent of the pronoun may seem to be obvious, an assumption appears to be made on the part of the writer that the reader is also the applicant for the service described in the text. In I.A. texts on the other hand writers generally prefer to refer directly to 'the applicant' as in example 308 on the previous page.

When subjects refer to the writer or his department, the rather vague 'we' may be used in P.A. as in example 302. The precise reference of 'we' might seem irrelevant to the reader, though not entirely. Applicants frequently need to correspond with departments who are processing their claims. In I.A. the reference to the appropriate authority is generally more precise; note reference to 'the court' in example 307.

The distribution of the different types of subject NP in GIL text types appears to be constrained by the purpose of the text. In P.A. texts which generally aim to provide only introductory information it seems acceptable to refer to the department as 'we'. This also appears to equalise the status relations between reader and department, i.e. to put it on a basis of 'we' and 'you', as well as making these relations more informal. In I.A. where texts carry far more detailed information, and are perhaps more reference than introductory texts, the precise identity of the people and organisations discussed needs

to be made more clear. Status is again equal, on the basis of 'the court' and 'the applicant' but more formal.

### 5.3.2 Verb

The verbal component of clauses in GIL is not discussed in detail in this chapter, but see the section below on the Passive, and Chapter 4 on the Verb phrase. It is however noteworthy that the distribution of contractions in the verbal components of GIL clauses is uneven. Verbal contractions are seldom found in I.A. texts, but are relatively frequent in P.A. texts. For example:

#### Personal Address

310. ..even if you don't have your pay slips.

311. ... even though you haven't applied them [skills] in a job  
before

#### Impersonal Address

312. Goods do not qualify for relief unless...

313. This relief applies only to those who have not  
previously claimed exemption

It seems likely that the option to contract some verbal items is not available in I.A. texts due to the more formal Status already commented on in I.A. texts. Verbal contractions are noted by Crystal and Davy (1969:p.113) to be a marker of Conversational English, and are therefore perhaps more common in informal speech than in writing. The appearance of contractions in official texts, which might normally be assumed to be formal does seem to provide some evidence that the writers of P.A. texts are making a conscious effort to reduce the apparent Status differences between Government Departments and the general public.

### 5.3.3 Object

Object NPs in GIL clauses, like Subject NPs, show a similar tendency in the distribution of pronominal and non-pronominal NPs. That is the tendency for P.A. texts to use pronominal NPs and for I.A. texts to use non-pronominal NPs. When observed at object position this trend however is much weaker than at subject position. Whilst object NPs in I.A. texts appear subjectively to be almost entirely non-pronominal, there seems to be a ratio of about 2 pronominal NPs to eight non-pronominal NPs in P.A. texts.

This tendency might be explainable by reference to the same constraints as those detailed for subject position on page 185.

Examples of object NPs in I.A. and P.A. clauses are given below.

Personal Address

314. We won't tell him (that you are claiming FIS).

315. The office will tell you when this will be..

Impersonal Address

316. Please read the introduction to the enclosed notes.

317. An employer may make a contract with agricultural workers engaged in husbandry to...

Not all GIL clause objects are NPs. Object position may be filled by a subordinate clause, typically introduced by 'that' in GIL. The distribution of object clauses appears however to be somewhat uneven, I.A. texts having object position filled by a clause a little more frequently than P.A. texts. In addition P.A. object clauses tend to omit the subordinator, for example:

Personal Address

318. Let your friends and relations know you are looking for a job.

V

I.O.

D.O. clause

319. We won't tell him [your employer] you're claiming FIS.

#### Impersonal Address

320. If a relative feels that it would be in the patient's interests...

321. I certify that the information given on this form is correct

The perceived difference in the ratios of clausal to phrasal objects between I.A. and P.A. seems rather too small to indicate any strong stylistic diversity, but the tendency of P.A. texts to omit the subordinator in subordinate clauses, may perhaps be a result of the Status constraints already observed. Quirk and Greenbaum (1973:p.317) note that the subordinator 'that' (very frequent in I.A. subordinate clauses) is frequently omitted in informal English.

#### 5.3.4 Complement

Complements, when they occur in GIL are very often adjectives or adjective phrases. The adjectives used are a rather limited set, being constrained by the administrative nature of GIL. The adjectives used are generally participial, or de-nominal. Here are some examples:

322. A vehicle whether used or not is admissable free of customs

S                      ADV                      V                      COMP

duty, provided....

323. If you are unemployed...

324. If your job prospects have been affected by a disability...

325. Death Grant is not payable ...

326. What income is taxable ?

Complements do not appear to show unevenness of distribution, when P.A. or I.A. texts are considered, the most noteworthy feature in stylistic terms appears to be the set of adjectives used: unemployed, payable, taxable, admissable etc. These particular lexical items are clearly constrained by the administrative province underlying GIL. Note too that the adjectives occurring in complements tend to be designative rather than attributive, that is they describe factual rather than putative qualities of the clause subject. This might be expected in official texts which must keep to the facts, and can be compared to advertising language where it is common to attribute qualities like 'delicious', 'refreshing' to products in order to enhance their image.

In consequence of the designative nature of adjectives used in GIL clause' complements, intensification is notably absent. Designative adjectives cannot be readily intensified, for example a person cannot be 'very unemployed' or an income 'extremely taxable'.



#### 5.3.4. Adverbial

Adverbial position in GIL clauses is perhaps the most interesting aspect of the GIL clause stylistically. Ignoring for the present adverbial clauses which are perhaps best discussed at sentence level, GIL clause adverbials may take the form of either an Adverb phrase, Noun phrase, or Prepositional phrase. GIL therefore makes full use of all 3 types of phrasal clause adverbials available in English.

The relative proportions of the differing types of phrasal clause adverbials vary between I.A. and P.A. texts. I.A. writers appear to favour Prepositional phrases over other types, and P.A. writers Adverb phrases, though all types are used in both P.A. and I.A. texts. Overall however, P.A. clauses appear less likely to have their adverbial position filled than I.A. clauses. In addition there appears to be a tendency for P.A. clause adverbials to be shorter than I.A. adverbials.

Whilst Adverbial position is most likely to be final in all GIL texts there appear to be far more initial adverbials in I.A. compared with P.A. texts. Finally, I.A. texts appear to have a tendency to co-ordinate several adverbial phrases at adverbial position, P.A. texts tend to have only one phrase per adverbial.

The general picture then of adverbial usage in I.A. and P.A. texts

is for P.A. to use fewer adverbials, mostly of a simple type, than I.A. texts.

Examples of clause adverbials in both I.A. and P.A. texts now follow.

#### Personal Address

327. If you are bringing up children alone ...

328. Do you have a FIS order book already ?

329. Are you getting Child Benefit now ?

330. How much help can you get this week ?

331. Ask at your local council office

332. Claim on this form

#### Impersonal Address

333. Customs duty, car tax and VAT must be paid at the port of entry.

334. An estimate of the customs charges may be obtained if desired,

from HM Customs and Excise at the address given in paragraph

276

335. Please write clearly in capitals.

336. If any of these conditions are broken before payment of the charges....

335. Death grant is not payable in respect of X.....

336. In approved cases Supplementary Benefit can be obtained.

337. Alternatively application may be made to the Court by any person....

338. In certain small cases the appointment of a receiver may not be necessary.

339. In the 1896 Act there are certain additional provisions...

In I.A. some adverbials may be formulaic, set phrases which have a specialised function in the administrative province. Some examples from the data include: if applicable; under the foregoing provisions; in the circumstances specified; in exceptional cases; in  
BLACK INK.

Because there is such a great deal of difference in adverbial usage between I.A. and P.A. texts, some attention ought to be paid here to the possible factors constraining it. P.A. has already been observed to be a modality for conveying introductory information to the reader in a relatively informal style, whilst I.A. conveys detailed information rather more formally. The I.A. modality approaches that of a reference text rather than that of an introductory one.

Adverbial usage in GIL seems likely to be constrained by the factors in the last paragraph. The adverbial's semantic function in conveying information about where, when, and how administrative processes take place is likely to be needed more in I.A. where comprehensive information is expected. Sometimes one adverbial phrase is not enough for the I.A. writer and needs to be supplemented. In P.A. texts less detailed information is needed, adverbials are therefore often not present or kept short.

The fact that adverbials more often occupy initial position in I.A. than in P.A. lends some support to the view expressed in the last paragraph. The placing of an adverbial in initial position lends greater importance to the information expressed, and in I.A. this seems to indicate the general level of importance attached to adverbials. Adverbials frequently found in initial position are those concerned with exceptions; e.g. 'in exceptional circumstances'.

The observations made here about clause adverbials parallel those for sentence adverbials. This adverbial type is examined in Chapter 6.

### 5.3.6 Vocative

Vocatives are a non-occurring category in GIL. This is not surprising as few situations can be imagined in which a GIL writer would want call attention to a particular type of reader by naming them. If used, the vocative would tend perhaps to make GIL read more like advertising copy. For example;

340. Single Parents! claim FIS now if you are on a low income.

### 5.3.7 Linking and Apposition

Apposition is not often found in GIL, its main use is in titles of benefits, where the apposed item is usually paranthetic, and enclosed in brackets. The bracketed item usually refers to the departmental code for a leaflet or form. For example:

340. In subsequent years the normal renewal reminder (form V11) will be sent to keepers.

Linking of subject NPs, though not very frequent in GIL, is very largely confined to I.A. texts, for example:

341. Customs duty, car tax, and VAT must be paid at the port of entry.

Such an observation might be expected for I.A. texts where there has already been observed to be a need for more explicit reference than in P.A. texts. Linking at subject position in I.A. texts makes it possible to include more precise information than if a single subject NP, for example 'The appropriate taxes', had been used.

#### 5.3.8 Passive and Active clauses

In spite of the widely held assumption that official texts use passive clauses a great deal, there are clearly more active than passive clauses in both I.A. and P.A. texts. This feature quickly becomes clear if only a rough count is made of passive and active clauses on any page of GIL texts. However, this is not to say that the distribution of passive and active clauses in GIL is not stylistic. GIL makes frequent use of the passive, but mainly in those texts already described as Impersonal address texts, where a subjective analysis suggests that as many as one in every four clauses may be passive. Many of these passives are also agentless.

If comparing GIL with the language of scientific journal articles, the proportion of passive to active clauses may not seem high, nor as consistent between texts. However, it must be recalled that whilst the use of the passive in scientific articles may be

determined by editorial policy, this is not so in GIL, where at least one departmental style manual openly discourages their use (Forms and Notices Unit 1983). The fact that the passive is nonetheless frequently used in GIL, especially in the more formal (I.A.) texts, suggests that there are certain strong constraints governing its use in GIL. This is tantamount to saying that the choice of a passive or active clause is not a matter of individual choice in GIL, but the result of other factors. The rest of this section is devoted to a discussion of these factors.

One way in which passive use might be explained in GIL is in relation to the motivations for passive use suggested by Palmer (1974:pp.86-87). These motivations were cited on pages 181-182 (see earlier), and prove helpful in explaining some passive use in GIL.

As the main type of passive in GIL appears to be the agentless passive of Impersonal address texts, these provide the starting point for discussion. Palmer (ibid) suggests that there are three possible reasons for using the agentless passive. These are essentially concerned with the agent itself, and whether it is unknown, irrelevant, or undesirable.

True 'unknown' agents are quite rare in GIL, at least in the sense noted by Palmer, in his example:

'He was killed' (ibid p.86)

where the killer is a mystery figure (or figures) whose identity is obviously unknown. In GIL agents are frequently unknown, because their identity is equal to that of a 'hypothetical agent', a person who may or may not be co-referential with the reader of the text, and whose duties or responsibilities need to be explained. For example:

342. The object of this leaflet is to explain what should be done when a mental patient has assets which need to be protected or which cannot be dealt with or used for his benefit.

The three underlined clauses in this example all refer to a 'hypothetical agent' whose duties are those of managing a mentally ill relative's personal and financial affairs. The agent is technically 'unknown', though note that this agent could have been made 'you' on the assumption (likely to be true in most cases) that the reader himself has to perform the functions described. This would have resulted in little distortion of meaning, and would be the likely method of presentation in a Personal Address text. The 'unknown agent' explanation of passive use in GIL seems rather weak when viewed like this.

Authors might consider an agent to be irrelevant if it is not necessary for the complete understanding of a clause. The decision about what is irrelevant may however be a subjective one. 'Irrelevant' agents might be those whose identity can be assured,



for example: 'the government', 'the department', 'the statute' and so on. In addition, it might be considered immaterial to the reader, who exactly is providing a benefit, laying obligation, making definitions, etc, because most sources are 'official', i.e. some form of governmental authority. Here is an example showing both immaterial and assumed agents:

343. Wages are defined as money or other thing had or contracted to be paid.

In the first clause in this example, some authority, possibly the statute, case decision, or departmental directive, is the agent of the first clause. The agent of the second clause can be understood to be a hypothetical employer. In the second clause the agent is redundant and therefore irrelevant, because use of the lexical items 'wages' and 'paid' makes the identity of the missing agent clear. It is also in a sense 'unknown', because no particular employer is being referred to.

Whilst the irrelevant agents omitted from the example above might be re-inserted were the text to be paraphrased in P.A., with a likely change to active instead of passive clauses, the insertion of the agents would add little to the meaning of the text, serving perhaps most of all to reduce formality, for example:

344. The law defines wages as money or any other thing you have from

an employer or which he contracts to pay.

In the introductory type of text typical of P.A., example 344 might be stylistically more appropriate though no more informative and less concise than example 345. In a I.A. text where previous knowledge of the subject is generally assumed and a high density of information content is necessary, the more concise form of example 343 would be stylistically more appropriate.

'Undesirability' of agents, like 'irrelevancy' is a rather subjective classification. It is easy to imagine a situation where the writer of a text might desire to avoid mentioning an agent, he might for example be writing something libellous and wish to avoid identifying himself as the agent in a clause, therefore remaining anonymous. The reader of the same text is likely to feel that the identity of the agent would have been most desirable.

There is very little evidence to suggest that GIL writers omit the agent in passive clauses in order to preserve anonymity, though note that it is often a strong convention in administrative circles for all correspondence to be signed by the director or manager (who bears ultimate responsibility), rather than by its author. This might be considered a type of anonymity, one in which individual officials are not mentioned by name, and might justify the writer excluding the agent in a passive clause. A possible example of this is:

345. If the modification applies in your case it will have been considered before the decision on your case was made.

The writer of this text, may indeed be responsible for considering cases and making decisions, and may need to remain anonymous because of the convention mentioned above. However, by far the best explanation appears to be the one mentioned earlier of irrelevancy. In a large department where decisions are standardised, there seems little reason to name the individuals responsible for making them. Examples like 345. appear to be witness to the simple fact that the only identity the official has is that of representing the department, or the government for which he works.

To summarise, the subjective hypothesis put forward in this section so far is that the agentless passive seems most likely to be the result of the general irrelevancy of expressing agents in GIL. They are rarely essential to the meaning of what a GIL text has to convey and are omitted in texts where space is at a premium, i.e. in I.A. In P.A. texts, whilst still technically irrelevant agents are included, in order to reduce formality. This is allowable in P.A. where conciseness is of less importance.

There are occasions in GIL where the passive agent is relevant, and is included, possibly because it is of some importance, for example the name of a service or benefit. Such agents are often lengthy, and it seems not unlikely that this in itself may provide another reason

for using the passive in some GIL texts. Possible examples of this are (agents underlined):

346. An allowance is available for animal foodstuffs consumed by a beast of burden used by the worker in his occupation.

(11 word agent)

347. The appointment of a receiver may not be necessary, the matter being dealt with by a simple order authorising the application of a patients property for his behalf.

(13 word agent)

This reason for using the passive has been termed 'weighty NP postponement' by Palmer (1974:p.87) and 'weighty agents' by Svartvik (1966:p.137). Svartvik observes that weighty agents subject to postponement tend to be of 8-8.5 words long. Both of the above examples fall into this category.

However, not only the long agent in example 346, but also the importance of the subject of the passive clause may have influenced the use of the passive. Palmer suggests 'thematization' as one reason for using the passive (1974:p.87), where the most important information in the clause is made subject and given prominence by being placed initially. In the case of example 347. it is 'animal

foodstuffs' which are clearly the most important information item in the clause. Thematisation therefore seems to be at work here as well as weighty NP postponement.

Other possible instances of thematisation at work in the data are:

348. A fixed reduction is one that has been agreed by the Inland Revenue.

349. Where the patient's Social Security Benefit or Pension is being received and used for his benefit by the hospital or other agent appointed on his behalf....

All the possible reasons for passive use in GIL discussed so far have been related to objective factors. One should not however rule out subjective factors, particularly the effect which any connotational meaning associated with passive use might have. This is in essence a semantic factor. It might be argued that through its use in scientific texts, the passive has come to take on a non criterial meaning of authoritative objectivity. GIL writers may make use of this in using the passive, particularly in the high information content modality and formal status I.A. texts, which may have technical subject matter. Quirk (1968:p.169) seems to suggest this when he says:

'The composing official is often afraid - and often with good reason - of being accused by his superiors or the public of lacking a proper command of dignified, remote, impersonal English.'

This section on the Passive completes the main body of this chapter on the clause. The chapter itself will now be summarised before proceeding on to the penultimate chapter in Part 1 on the Sentence.

#### 5.4 Summary

To summarise, the GIL clause is distinguished stylistically in the following aspects of clause structure. Subjects are nearly always NPs in GIL, but the subject NPs of P.A. texts are very frequently single pronouns. I.A. texts on the other hand tend to be complex NPs, or linked NPs. This irregular distribution of subject types appears to be due to be constrained partly by Status (formality) and partly by the difference between the introductory text modality of P.A. and the reference text modality of I.A.

The verbal component in P.A. texts is notable stylistically for the frequency with which verbal contractions occur. This seems to suggest the influence of status (formality) constraints. Contractions are very frequent in speech, their use in official written texts seems to indicate lower formality in P.A. texts

Objects show a similar but less pronounced tendency to be simple pronominal NPs in P.A. and more complex NPs in I.A. This is likely to be due to the same constraints determining stylistic choices at subject position.

Complements in GIL tend to be adjectives or adjective phrases. Lexically, these adjectives are drawn from a small set of administrative adjectives of a predictable nature, e.g. taxable, payable, etc. The adjective stock used in GIL is distinguished stylistically by being either de-nominal or participial. Semantically, the adjectives used are almost exclusively designative rather attributive, intensification in adjective phrases is notably absent. Complement type, and adjective use appears to be constrained by province factors in GIL. The subject matter of GIL texts is factual and sometimes technical, dealing always with administrative matters.

Adverbial position in GIL is very frequently filled, though I.A. and P.A. texts differ in the adverbial type favoured and the rate of use. I.A. texts tend to use more adverbials than P.A., and favour prepositional phrases. Several such phrases may be co-ordinated in one adverbial. P.A. uses less adverbials, and these tend to be simple adverbs or adverb phrases. P.A. adverbials are on the whole shorter than I.A. adverbials. This is partly due to the rarity of co-ordinated phrases in P.A. adverbials. These differences in adverbial use seem to be constrained by the introductory text

modality of P.A. texts and the reference text modality of I.A. texts. Prepositional phrases, especially co-ordinated ones carry a great deal of information content in I.A., whereas the simple adverbials in P.A. like 'now', 'already' and 'alone' convey much less information.

GIL is characterised by frequent use of the passive, but much more so in I.A. than in P.A. texts. However, in no text do there appear to be more passive than active phrases. The situational constraints determining passive use are complex, and vary according to whether a passive clause has an agent specified or not. Agentless passives may be chosen in GIL because the agents are frequently irrelevant. This may be because they can be assumed from context, or because of the administrative convention of departmental rather than individual responsibility. Passive clauses with a specified agent may be used specifically to enable thematisation of an important piece of information to take place. Alternatively it may be the result of the need to remove a lengthy subject NP from initial position in a clause.

All these factors are ultimately province constraints, being concerned mainly with the business of administration itself. However status constraints may be at work too, the passive being used to produce a formal impersonal style of English thought to be more suitable to administrative texts.



## PART 1 - CHAPTER 6 - THE SENTENCE

### 6.1 General

Beyond the clause in GIL there is relatively little of interest to the stylistician. Chapter 3 of this thesis demonstrates how much of the special flavour of GIL writing comes from the extensive use of possibilities at NP level, particularly in I.A. texts where postmodification is heavily used to increase the density of information carried in the text. The scope of the simple sentence has already been investigated in discussing the clause. Complex sentence structures result from the inclusion at Subject, Complement or Adverbial position of dependent clauses. Clauses occurring at complement position do occur in GIL and were discussed in the previous chapter. Clauses are notably absent at subject and at object position. It is the adverbial position in the GIL sentence which is of most interest to the stylistician, particularly the semantic function of the adverbial types used.

This chapter will follow the now familiar procedure of examining sentence structure in English and then discussing the use made by GIL of this.

### 6.2 The Sentence in English

The terminology for sentence description in this chapter is taken

mainly from Crystal and Davy (1969:pp. 45-49). Sentences will be classified to type according to whether they are MAJOR or MINOR, and whether they are SIMPLE, COMPLEX, COMPOUND or MIXED in structure. These terms will be fully explained.

The fundamental on which this description of the sentence is based is the distinction between MAIN and DEPENDENT clauses. A dependent clause is one which must be linked to another element of structure to have any meaning. A type of dependent clause already encountered in the chapter on the Noun Phrase is the relative clause. The dependent clause (underlined) in the sample sentence:

350. The form which you must use is obtainable from your local tax office

has no meaning if removed from the NP in which it is a dependent structure. Dependent clauses cannot function independently as a major sentence, only as a minor sentence (see below).

A main clause may occur on its own as a (major) sentence. It is therefore an independent item of linguistic structure which can be at least partly understood without reference to any other structure. Note however that a sentence may make anaphoric reference to a linguistic item in a previous sentence. In the example below, the pronoun 'them' can only be understood in relation to 'the eggs' in the previous sentence:

351. Break the eggs into a clean bowl. Now beat them thoroughly with a whisk.

Main and dependent clauses may be combined in English to form complex sentences, for example:

351. Whizzo washes your clothes whiter, even with heavy soiling!

MAIN CLAUSE

DEPENDENT (ADVERBIAL) CLAUSE

#### 6.2.1. Major and Minor sentences

As with the clause, discussed in the previous chapter a sentence must contain at least a Subject and a Verb, for example:

352. Mary cried.

S V

Some verbs may require other obligatory elements of structure too.

For example the obligatory complement with BE:

353. James is an idiot

S V C

Sentences which satisfy these attributes are major sentences. Those which do not are classed as MINOR in Crystal and Davy (1969:p49).

These are sentences used independently, which however may only consist of a single element of clause structure or combinations of such elements, e.g. Verb Adverbial:

354. Cleans whiter

This sentence might be a minor sentence used in a soap powder commercial for the 'Whizzo' soap powder mentioned above. Minor sentences are noted by Leech (1966) to be characteristic of advertising English.

#### 6.2.2 Simple and Complex Sentences

Single clause sentences are the simplest form of major sentences, these will be given the name 'SIMPLE SENTENCE' in all future discussions. Whilst all sentences which contain more than a single clause are in some way more complex than Simple sentences, the practice of Crystal and Davy (1969:p.48) is followed. These authors regard only those sentences whose adverbial is realised by a dependent clause as Complex. Sentences where Subject, or Complement are realised by a clause are not considered to be complex. The rationale behind this choice is well covered in Crystal and Davy (ibid).

An example of a complex sentence according to the definition given here might be:

355. Bill will explain the matter to you if you wish.

MAIN CLAUSE

ADV CLAUSE

An example of a simple sentence containing a clause at subject and at complement position might be:

356. Whoever said that needs his head examining.

SUBJECT CLAUSE

V

COMPL CLAUSE

### 6.2.3. Compound and Mixed Sentences

Where a sentence consists of two or more clauses, none of which is dependent, the sentence can be considered to be a compound sentence. The clauses in such sentences are generally joined by a co-ordinating item like 'and', 'but', 'or' etc. For example:

357. He chose not to go but Mary thought otherwise.

The subject of a second clause may be omitted if it is co-referential with the subject of the first:

358. Fred thought for a moment and decided on his plan.

A mixed sentence is simply a compound sentence where one of its component clauses contains a clause adverbial. The sentence therefore has both the attributes of complex and compound sentences.

A great deal of information can be carried in such a sentence. For example:

358. Blend the sugar and the butter, add the eggs, beat the mixture

CLAUSE 1

CLAUSE 2

CLAUSE 3

well, and add the vanilla essence before sieving and adding the

CLAUSE 4

CLAUSE ADVERBIAL (co-ordinate)

flour.

#### 6.2.4. Adverbial

The adverbial component of a sentence, if realised can provide a great deal of additional information in an otherwise simple sentence, for example, the sentence:

359. Brian promised to meet me.

can have an added adverbial:

360. Brian promised to meet me, just as I requested, before leaving for Rome, provided that I told nobody else about it.

To the original single clause has been added an adverbial composed of 3 adverbial clauses, all of which have clearly separate semantic

functions. The first adverbial relates to Manner, the second to Time and the third to a Condition. These semantic functions are all among those frequently fulfilled by adverbial clauses. Other important semantic functions of adverbials in English are:

Reason/cause

361. Vicky did that because she hates me.

Concession

362. You can't do that even if you do have permission.

Location

363. First find the stop tap, located underneath the sink.

Purpose/result

364. The scheme was changed so as to make it more attractive to investors

Note that the choice of semantic function in an adverbial may be closely related to province. For example, recipes frequently need to make reference to time and place, and so are likely to make frequent use of time and location adverbials:

365. When you have beaten the eggs, fold them carefully into the

TIME ADVERBIAL

LOCATION

mixture

ADVERBIAL

### 6.3. The Sentence in GIL

Much of the potential material for this chapter has already been discussed elsewhere. It has already been observed in chapter 5 that subject position is not usually filled by a clause in GIL (see page 183). The verbal elements of the sentence have already been adequately dealt with in chapter 4 on the verb phrase. Sentences containing object clauses were discussed on pages 188-189 of chapter 5 above. Like the Subject in GIL, Sentence complements are hardly ever clauses. What remains for discussion in this chapter is the nature of adverbial clauses in GIL and the proportions of the different sentence types: Major/minor, Simple/Complex/Compound/Mixed etc.

#### 6.3.1 Major/Minor Sentences in GIL

Almost all sentences, without exception, are of the major type in I.A. texts. P.A. texts do however make use of minor sentences, particularly when the modality of the text is promotional rather than merely introductory. The connotational associations of minor sentences with advertising English produce a 'persuasive' text, not



unlike an advertisement:

366. And these things free!

367. Saves up to `35.

368. More money each week for families on a low income.

369. Need more money?

370. More Help!

All these sentences come from leaflets explaining cash benefits to poor or unemployed readers. The function of such sentences is very likely to be to persuade the reader that claiming a particular benefit is a good idea. This is in spite of the social stigma that is often associated with benefits for the underprivileged, and possibly to counteract it. Whilst 'style borrowing' is obviously at work here, the texts remain essentially GIL texts in other respects, for example the heavy use of adjectival premodification found in Advertising English is not present.

#### 6.3.2 Simple/Complex Sentences

Whilst it is universally assumed that official language is 'complex', and that this complexity goes hand in hand with

formality, one should be extremely wary about equating this 'complexity' with the term 'complex' as used to describe sentences containing adverbial clauses, here. A subjective assessment of GIL texts seems to indicate that the proportion of Complex sentences is fairly constant for any given GIL text, I.A. texts do not appear to use more, complex sentences than P.A. texts.

Note that because a major sentence is not simply either Simple or Complex, but may fall into one any of 4 categories; Simple, Complex, Compound, or Mixed, it follows that similar proportions of complex sentences in I.A. and P.A. texts do not mean that they have the same proportion of simple sentences. On an initial subjective appraisal this appears to be true for I.A. and P.A.. P.A. texts appear to contain rather more Simple sentences than I.A. texts, which in turn have slightly greater portion of Compound and Mixed sentences.

This observation is not at all surprising considering the modality differences between I.A. and P.A. P.A. has the modality of an introductory text, and has already been observed to have a relatively low density of simple non-technical information. The simple sentence therefore suits the needs of P.A. texts well. The reference text modality of the I.A. text with its greater density of information, often technical, favours sentences which can carry a much greater amount of content than P.A. texts. This need is partly answered in the use of complex sentences, where the adverbial clause is used to add extra information, and partly in the use of

compounding two (main) clauses in a sentence, where the additional content comes in the extra clause. Compounding of clauses and the addition of adverbial clauses serve slightly different purposes in GIL. It is the function of adverbial clauses that will be discussed here.

### 6.3.3. Adverbial Clauses in GIL

A subjective analysis would appear to indicate that over half of the adverbial clauses in many GIL texts are concerned with either detailing conditions or concessions. For example:

371. A vehicle whether used or not is admissible free of customs duty, provided that it was previously exported from the EEC

372. If you do not do so you may pay an incorrect amount of tax.

373. If you are unemployed don't use the form...

374. Release from conditions will be allowed, only if the Customs charges are first paid.

375. If your job prospects have been affected by a disability, illness or accident, help is available from the Disablement Resettlement Officer

376. Ownership of a vehicle may be counted from the date on which the importer acquired the right to take possession of it, even though the actual taking possession may have been delayed.

377. To avoid losing money, post your claim now - even if you don't have your pay slips, birth certificates etc...

The frequent use of adverbial clauses in GIL may well be determined by Province factors. The administrative province which constrains GIL is concerned not only with making clear to the reader what the basic outlines of duties, responsibilities, services, benefits, laws etc are, but what conditions have to be satisfied, and what concessions may be allowed. Without this information the reader would not be fully informed. Both I.A. and P.A. texts have a need to present this type of information, though the conditions in I.A. texts (exemplified by examples 371, 374 & 376 above) tend to be more precise in reference than those of P.A. texts.

The use of conditional clauses is also frequently found in Legal English according to Crystal and Davy (1969:p.203). They claim that 'if clauses' are the main source of sentence complexity in Legal English. All administration in government is ultimately referable to the legislation which controls it. It does seem therefore that Legal texts with their many conditional adverbial clauses may be the indirect cause of their frequent use in GIL.

Whilst conditional/concessive adverbials as a single group may account for up to half the clause adverbials in a text, No other single semantically defined group of adverbials can be pinpointed as accounting for most of the remainder. Time, Manner, Location, and Cause are all clause adverbial types used to some extent in GIL. For example:

Time

379. The office will tell you when this will be.

Manner

380. ....the applicant should complete this in full, as if they were already married.

381. This change puts lorry taxation on a fairer basis, by enabling the Government to relate tax levels much more closely to the road cost imposed by different groups of vehicles.

382. ...call in when you can and keep in touch by telephoning in.

Location

383. Wherever possible, provide full names and addresses for all applicants

Reason

384. In order to ascertain the correct amount of tax, it is important that you fill in each section correctly.

All these adverbial types might be predicted when the administrative province constraining GIL is considered. Administration as a process has a clear need for setting exact times by which things must be completed, the location of things either in texts (or physically in the case of buildings and offices), the reasoning behind particular requests, and so on. This goes some way towards explaining why the information in many GIL sentence is incomplete without the addition of an adverbial clause.

It has already been pointed out that the presence of adverbial clauses does not necessarily result in what the layman refers to as complexity in official language. There are occasions however, when GIL texts (particularly I.A.) have so much need to use the extra informational content possible in adverbial clauses, that the sentence can be overloaded with adverbials. This seems to happen when the conditions concerning a point of administration are very involved. The resulting adverbials are often ponderous and lengthy, for example:

385. ...provided that no relief from customs duties was claimed at export and any customs duty on foreign parts incorporated

in it previously relieved, has later been paid or repaid.

Were such an adverbial sequence is combined with the tendency for I.A. texts to use long NP postmodifiers, the resulting sentences can reach great lengths:

386. Where a customary holiday on which the worker is not required to work for the employer immediately preceded a period of annual holiday or annual holiday and additional annual holiday required to be allowed under the foregoing provisions of this paragraph, together with any customary holiday, exceeds the number of days constituting the worker normal working week then, notwithstanding the foregoing provisions of this paragraph the duration of that period of annual holiday or additional annual holiday may be reduced to one day, and in such a case one day of annual holiday may be allowed on a day on which a worker normally works for an employer (not being the workers short day) in the holiday season or after the holiday season in the circumstances specified in sub-paragraph (2) (b) of paragraph 12.

This sentence comes from a code of practice leaflet issued by the Department of Employment. The province involved is not law, as the text concerned is not a legal document. However the similarities to legal English are obvious. This example shows how the constraints operating on GIL may lead writers to produce a style very close to a

legal one, demonstrating that there may be no clear divisions between one style and another.

#### 6.3.4. Compound and Mixed Sentences in GIL

Compounding, like the addition of adverbial clauses, is a process of sentence construction which results in greater information carrying capacity (when compared with the simple sentence). As a process it is seen on its own in Compound sentences, and combined with adverbial clauses in Mixed sentences. Unlike the addition of an adverbial clause compound does not result in more information on conditions, concessions, etc, but in the inclusion of a further main clause. The information in a compound or mixed sentence could therefore often be alternatively expressed in two simple sentences, or a simple and a complex sentence, for example:

387. Jack went up the hill to fetch a pail of water. Jill did the

COMPLEX SENTENCE

ADV CLS

SIMPLE

same.

SENTENCE

388. Jack went up the hill to fetch a pail of water and Jill did too.

One of the advantages of compounding over two separate sentences, is



to omit mentioning a subject twice if it is identical in two separate clauses. This gives greater economy of statement, and in GIL texts, particularly I.A. where space is at a premium, this is a possible reason for the use of compounding. Two typical examples of compounding in I.A. are:

389. The applicant should fill in and sign the two enclosed forms and should attach them to the completed application form.

390. This Notice gives information about importation of private motor vehicles into the U.K. and supersedes the November 1980 edition.,

This section on compounding completes the discussion of GIL sentence structure. This chapter will now be summarised, before make a final summary in chapter 7 of all the observations made in chapters 3 to 6 of Part 1.

#### 6.4 Summary

In summary the structure of the GIL sentence is notable stylistically mainly for the frequency with which its adverbial component is realised and also the frequent use of a particular semantically defined group of adverbials; those which are either conditional are concessive. There appears to be little difference in the frequency with which I.A. and P.A. texts use adverbial clauses,

though the reference of adverbials in the former is frequently more precise. High adverbial usage in GIL may be the result of Province constraints. The administrative province encompasses areas of work that are frequently concerned with the semantic functions of the adverbials observed, particularly those which have to do with conditions and concessions.

Sentence type in GIL varies between I.A. and P.A. texts. The frequency of Complex, Compound and Mixed Sentences varies little between the two text types, though I.A. appears to make slightly greater use of all three than P.A. The major difference in sentence type between the two text types is in the frequency of Simple sentences. P.A. texts have a larger proportion of these than I.A. This stylistic phenomenon may be the result of modality constraints; I.A. texts have the modality of a reference text type, whilst P.A. texts tend to be introductory texts. It may be that Simple sentences are more suited to the expression of simple introductory material in GIL whilst more detailed information may require the greater information carrying potential, and conciseness of non-Simple sentences.

The information carrying potential of sentences is held in this chapter to be enhanced by the addition of adverbial clauses allowing additional information about conditions, concessions, time, manner, location and cause. The conciseness of sentences is held to be enhanced by compounding which allows the co-referential subjects in

compounded clauses to be mentioned only once. This allows the greatest saving of text space when subjects are long in terms of words.

This section completes this chapter on the sentence in GIL. It now remains to conclude Part 1 of this thesis by making an overall summary of what has been discovered so far.

## PART 1 - CHAPTER 7 - SUMMARY

### 7.1 General

The aim of Part 1 of this thesis was stated in Chapter 2 (Method) (p62) to be that of:

'..firstly to note any particular items of linguistic structure that are unusually absent from the texts or that occur with a particularly high frequency, the assumption being that these features are possible style markers in GIL. The second aim is to attempt to ascribe the features thought to be stylistic to a particular stylistic dimension.'

The theoretical basis of Part 1 was observed to be qualitative (p61), further quantitative study was planned to be undertaken in Part 2 using Part 1 as a basis for selecting variables for quantitative study. The qualitative observations made in Part 1 therefore need to be summarised before proceeding to Part 2. This has in part already been done at the ends of chapters 3 to 6. It would be convenient however to reproduce all those summaries together. This is now done below, under appropriate headings.

### 7.2 Summary of observations on the NP in GIL

NP postmodification in GIL illustrates well the differing

constraints operating on I.A. and P.A. type texts. The personality or impersonality of address involved is not interesting for its own sake but because of the Province and concomitant Status constraints which bring it about. These constraints are the same ones that bring about the choice of predominantly (explicit) finite clause postmodifiers in P.A. texts, but (less explicit) phrasal and non-finite clause postmodifiers in I.A. texts.

The Provinces of I.A. and P.A. appear to differ, one concerned with the imparting of technical information, the other with general, introductory information. Status relationships are related to these Province differences; in I.A. the relationship is that between a reader and writer who are strangers to one another, but share some knowledge about their subject matter, in P.A. the writer tries to assume a more personal relationship with the reader, though there is little shared knowledge: The aim here may be to establish confidence between reader and writer by assuming a friendly personally directed style. This confidence presumably already exists in the case of I.A. texts.

Finally, multiple postmodification, especially embedded postmodification, highlights the constraints which a province concerned with explaining legislation to the general public imposes. In doing this multiple postmodification in GIL suggests links with Legal English in addition to those which were observed in GIL premodification.

### 7.3 Summary of the observations on the VP in GIL

GIL makes a very narrow selection from the tense and aspect system of English, using mainly the simple present tense throughout. This seemingly simple situation is more complex on further examination as GIL makes very full use of most of the different uses of the simple present. Unlike English generally, the main function of the simple present in GIL is not to refer to habitual activity, but in simple reporting, the simple present is also widely used in temporal and conditional clauses to express a conditional future usually dependent on action to be taken by the reader. The performative function of the simple present has a small but important role to play in declaration sections in GIL texts, and suggests a possible further parallel with legal English.

The simple past is not frequently used in GIL, its use being confined to pension and tax leaflets, where age and the previous years financial status have importance. The use of the past tense to express conditionality, dependent on past events, parallels the conditional future use of the simple present.

The use of present perfect is observed to be the most frequent method of referring to past events in GIL, this is seen to be related to the present perfect's ability to express a past with current relevance. This tense/aspect combination is seen to suit the constraints of the administrative situation in which the only

reason to refer to the past appears to be if it is relevant to a current problem or claim.

GIL poses distinct problems for the neat description of mood. It is seldom possible with modal CAN, MAY or MUST to say that the meaning expressed in GIL corresponds exactly to traditionally defined categories like 'permission' and 'obligation'. GIL texts are frequently ambiguous to the extent that it is often difficult to pinpoint who is taking responsibility for granting permission or imposing obligation, the department concerned or the law. This failure to be able to classify modal uses accurately in GIL provides an insight into the status constraints which operate in the administrative situation: Whilst government departments serve authority, they are not that authority itself.

Finally, the tentative modal forms COULD, WOULD, and MIGHT, are observed to be unevenly distributed in GIL, occurring more frequently in the (less formal) Personal Address texts. These modals are perceived subjectively to be more friendly and less formal than the corresponding forms; CAN, WILL, and MIGHT. Their more frequent use in P.A. GIL than in I.A. is seen to be due to status constraints. P.A. texts are introductory texts written with the intention of encouraging readers to claim benefits or find out more about a particular service, this purpose requires a relatively informal style.

#### 7.4 Summary of Observations on the Clause in GIL

The GIL clause is distinguished stylistically in the following aspects of clause structure. Subjects are nearly always NPs in GIL, but the subject NPs of P.A. texts are very frequently single pronouns. I.A. texts on the other hand tend to be complex NPs, or linked NPs. This irregular distribution of subject types appears to be constrained partly by Status (formality) and partly by the difference between the introductory text modality of P.A. and the reference text modality of I.A.

The verbal component in P.A. texts is notable stylistically for the frequency with which verbal contractions occur. This seems to suggest the influence of status (formality) constraints. Contractions are very frequent in speech, their use in official written texts seems to indicate lower formality in P.A. texts

Objects show a similar but less pronounced tendency to be simple pronominal NPs in P.A. and more complex NPs in I.A. This is likely to be due to the same constraints determining stylistic choices at subject position.

Complements in GIL tend to be adjectives or adjective phrases. Lexically, these adjectives are drawn from a small set of administrative adjectives of a predictable nature, e.g. taxable, payable, etc. The adjective stock used in GIL is distinguished



stylistically by being either de-nominal or participial. Semantically, the adjectives used are almost exclusively designative rather than attributive, intensification in adjective phrases is notably absent. Complement type, and adjective use appears to be constrained by province factors in GIL. The subject matter of GIL texts is factual and sometimes technical, dealing always with administrative matters.

Adverbial position in GIL is very frequently filled, though I.A. and P.A. texts differ in the adverbial type favoured and the rate of use. I.A. texts tend to use more adverbials than P.A., and favour prepositional phrases. Several such phrases may be co-ordinated in one adverbial. P.A. uses less adverbials, and these tend to be simple adverbs or adverb phrases. P.A. adverbials are on the whole shorter than I.A. adverbials. This is partly due to the rarity of co-ordinated phrases in P.A. adverbials. These differences in adverbial use seem to be constrained by the introductory text modality of P.A. texts and the reference text modality of I.A. texts. Prepositional phrases, especially co-ordinated ones carry a great deal of information content in I.A., whereas the simple adverbials in P.A. like 'now', 'already' and 'alone' convey much less information.

GIL is characterised by frequent use of the passive, but much more so in I.A. than in P.A. texts. However, in no text do there appear to be more passive than active phrases. The situational constraints

determining passive use are complex, and vary according to whether a passive clause has an agent specified or not. Agentless passives may be chosen in GIL because the agents are frequently irrelevant. This may be because they can be assumed from context, or because of the administrative convention of departmental rather than individual responsibility. Passive clauses with a specified agent may be used specifically to enable thematisation of an important piece of information to take place. Alternatively it may be the result of the need to remove a lengthy subject NP from initial position in a clause.

All these factors are ultimately province constraints, being concerned mainly with the business of administration itself. However status constraints may be at work too, the passive being used to produce a formal impersonal style of English thought to be more suitable to administrative texts.

#### 7.5 Summary of observations on the Sentence in GIL

The structure of the GIL sentence is notable stylistically mainly for the frequency with which its adverbial component is realised and also the frequent use of a particular semantically defined group of adverbials; those which are either conditional are concessive. There appears to be little difference in the frequency with which I.A. and P.A. texts use adverbial clauses, though the reference of adverbials in the former is frequently more precise. High adverbial usage in

GIL may be the result of Province constraints. The administrative province encompasses areas of work that are frequently concerned with the semantic functions of the adverbials observed, particularly those which have to do with conditions and concessions.

Sentence type in GIL varies between I.A. and P.A. texts. The frequency of Complex, Compound and Mixed Sentences varies little between the two text types, though I.A. appears to make slightly greater use of all three than P.A. The major difference in sentence type between the two text types is in the frequency of Simple sentences. P.A. texts have a larger proportion of these than I.A. This stylistic phenomenon may be the result of modality constraints; I.A. texts have the modality of a reference text type, whilst P.A. texts tend to be introductory texts. It may be that Simple sentences are more suited to the expression of simple introductory material in GIL whilst more detailed information may require the greater information carrying potential, and conciseness of non-Simple sentences.

The information carrying potential of sentences is seen to be enhanced by the addition of adverbial clauses allowing additional information about conditions, concessions, time, manner, location and cause. The conciseness of sentences is seen to be enhanced by compounding which allows the co-referential subjects in compounded clauses to be mentioned only once. This allows the greatest saving of text space when subjects are long in terms of words.

This brief chapter concludes Part 1 of the thesis; the qualitative analysis of GIL. Part 2 which follows aims to further research some of the observations in this qualitative analysis by subjecting them to quantitative study.

## PART 2 - CHAPTER 1 - INTRODUCTION

### 1.0 Rationale for Conducting Quantitative Stylistic Research

Very few linguistic Stylistic studies endeavour to validate or lend support to their subjective observations from quantitative evidence. The reverse is quite often true in Literary Stylistics where many articles have been written on the subject (cf. Bailey and Dolezel : 1968). This lack of quantitative support does not necessarily invalidate the results of linguistic stylistic studies, as reliance on native speaker intuitions is somewhat inevitable at many stages in stylistic research anyway as pointed out by Crystal (1971).

Stylistic research unvalidated by the use of quantitative analysis however has the disadvantage of not being replicable or open to re-testing. It also has the disadvantage of being open to disagreement between different individuals' intuitions, disagreements which cannot be easily resolved. Qualitative research therefore is somewhat limited in the conclusions which can be drawn from it. It is mainly for this reason that a quantitative study is being carried out here. It will hopefully provide the basis for further confirmatory work in this field, and provide a yardstick against which other styles can be analysed using exactly the same procedure.

The first step in any quantitative study is to define the problem to be solved.

## 1.1 General Statement of the Problem

Given that Part 1 of this thesis observed Government leaflets to be qualitatively distinct in style and to have two major sub-styles: Impersonal Address (I.A.) and Personal Address (P.A.), and that the theory chosen (see Introduction to Part 1) requires this distinctiveness to be relatable to situational dimensions (in the present case Province, Status, and Modality), how can the way in which these varieties are distinct be shown to have quantitative support? In addition; how can the extent to which these varieties are distinct from one another be shown to have quantitative support.

This general statement of the problem, essentially one of proving dissimilarity, also by implication involves the problem of how and to what extent are the texts referred to collectively as 'Government Leaflets, P.A and I.A. similar to one another? Do these texts form natural homogeneous clusters?

Note that demonstrating similarity here is equally as important as demonstrating difference. Although the latter alone has always been seen traditionally as the subject matter of linguistic stylistics. Testing similarity is an important prerequisite for stylistic analysis since any statement of stylistic difference made ultimately rests on an (often subjective) grouping of texts into groups of supposed linguistic similarity. These subjective groupings can clearly not be logically validated by the discovery of stylistic

differences based on the same groupings, support must be sought from other external sources.

### 1.2 Further notes on 'Difference' and 'Similarity'

Neither similarity or difference are easy terms to define clearly. The question immediately arises of how similar or different are the styles under consideration, and to what other styles are they different or similar? Any stylistic study is valueless to other stylisticians unless it answers these questions. It is not surprising then that a vast array of statistical techniques are used in the literature to discover different aspects of stylistic difference and similarity. This will emerge in the review of the more important statistical studies of style which now follows.

### 1.3 Format for Reviewing Previous studies

The review of each study below will take the following form. First a description of the theory behind the statistical technique used, followed by a description of the actual study itself; its aims and results. Finally an appraisal of the usefulness of the technique for stylistic research and its limitations will be given. This appraisal must be ultimately based on the two main criterion cited in 1.1 above; those of demonstrating stylistic difference and proving group similarity. In general the shortcomings of any one study in demonstrating difference are generally answered in another, but none

of the studies in the literature (to the author's knowledge) tackle the problem of establishing similarity.

#### 1.4 Some basic concepts associated with statistical research

There is a need to establish at this point definitions of some basic concepts that will be referred to frequently in discussing previous studies. These are; Independent and Dependent Variable, Significance testing and Levels of significance, and Alternative and Null Hypotheses.

##### 1.4.1 Independent and Dependent Variables

The term 'Independent Variable' (I.V.) generally refers to sets or groups of individuals (or the criteria for those groupings) between which the researcher wants to demonstrate a difference. In experiments these are often referred to as 'treatments', an example of this might be two sets of patients one administered drug A and the other drug B. Elsewhere the groups tend to be referred to as 'attributes' e.g. if men are compared with women, according to Leach (1979), (see sect 1.11.1 for a discussion of experimental and non-experimental studies).

The Dependent variable (D.V.) is the thing measured to assess the difference between the two groups. In the medical experiment mentioned above this might be the number of hours before the patient's temperature returned to normal. Each 'individual' or



'subject' (these can refer to things as well as people) on the I.V. has a score on the dependent variable.

Studies may have one or more I.V.s or D.V.s. No studies exist to the author's knowledge in Linguistic Stylistics which have more than one I.V. But studies do vary between those that compare variables only in pairs (bi-variate studies) and those which compare more than two variables at once (multivariate studies). These two approaches can be used to investigate different aspects of the same data.

Note that the researcher's choice of D.V. is subjective and that the results of any study will depend on how skillful he is in locating variables that will show up significant group differences. The choice of what groups to choose for the I.V. is similarly subjective and carries the same implications as well as the danger already mentioned in non-experimental studies that the attributes chosen may not reflect natural groupings when scores on the D.V. are considered

#### 1.4.2 Significance Testing and Levels of Significance

If the scores for two different groups of the I.V. on a D.V. are examined visually it is often possible to establish that they are indeed different. Such is the case in the example given on the next page:

	I.V.	D.V.
g	1	4
r	1	3
o	1	2
u	<hr/>	
p	2	9
	2	7
	2	7

Group 1 has obviously got much lower scores than Group 2, but the question that is most often proposed in statistical work is whether these groups are random samples from larger populations that one is really wanting to compare. We need to ask therefore, "Can the differences between the groups be attributed to chance? (in which case they must both be from the same population) or are they really from different populations?" The question of how large a difference must be before it can be considered not to be the result of chance alone is one of significance, and the cut off point between chance and significant differences one of significance level. Researchers vary in what level of significance to choose, though according to Plutchik (1983:p70) traditionally this is the 5% significance level. Such a level indicates that the observed differences are likely to occur by chance only 5% of the time or less in random samples from the same population.

Kenny (1982:pp106-7) refers to the same procedure as 'testing at the 0.05 level', but points out that if a researcher was being more cautious, for example when trying to confirm previous research a

more cautious level might be chosen, very often the 0.01 level. This would indicate that the researcher would only accept his results as being significant if the probability of their being the result of chance was less than one in a hundred.

In practice therefore the setting of a significance level is an area of quantitative research in which there is a certain degree of subjectivity.

#### 1.4.3 Hypotheses, Null Hypotheses and Alternative Hypotheses.

According to Kenny (1982:pl07) any significance testing requires a hypothesis to be tested. This hypothesis in Linguistic stylistics (if formalised, which is fairly rare) is usually of the nature that two sets of texts sampled randomly belong to different populations and therefore to different styles (see sect 1.11.2 for further discussion of these terms). However significance testing involves the testing of the opposing hypothesis which would hold if there was insufficient evidence to support the original hypothesis, in other words the hypothesis that both groups of texts belong to the same population.

The original hypothesis is usually referred to as the 'alternative hypothesis' and the opposing hypothesis as the 'null hypothesis'. So one speaks of 'Rejecting (or failing to reject) the null hypothesis at the .05 level of significance'; i.e. finding (or not finding) a big enough difference on the D.V. between the I.V. groups to pay any

attention to. The final section of this introduction is devoted to a statement of the hypotheses to be tested in Part 2.

#### 1.5 MARKWORTH AND BAKER (1980) - One way ANOVA - Bi-variate Study relating I.V. and D.V.

A recent example of bi-variate statistics being used on stylistic data is to be found in Markworth and Baker (1980:pp237-239). These authors use a one way analysis of variance (ANOVA).

Kenny (1982:pl27) explains that the purpose of the one way ANOVA is to test the significance between the D.V. means of several different groups (i.e. an I.V. with 3+ values). It tests the null hypothesis that the samples are independent samples taken from populations having the same means. As an ANOVA may deal with more than two samples any pair of which are potentially significantly different from each other an additional test may be required to establish which these groups are. This is usually either the Tukey or Scheffe test.

Markworth and Baker's study had an I.V. with 5 groups. These groups were 500 samples (in each group) of ten sentences in length, from 5 different sources of presumed stylistic difference, and were drawn from the Brown Corpus (see Kučera and Francis:1967), a one million word corpus of American English consisting of 500 texts from different sources. The 5 groups were; Fiction, Government Documents,

Learned Journals, Newspaper Reports, and Popular Journals. Markworth and Baker report on 4 comparisons; Fiction with all other styles, Government Documents with Learned Journals, Government Documents and Learned Journals with Newspaper Reports and Popular Journals, and Newspaper Reports with Popular journals. The groups were compared on a wide range of syntactic variables (36 in all). Counting of variable scores was done automatically by computer, hence somewhat approximately.

Markworth and Baker found that Fiction was differentiated from other styles by 14 variables (mainly units of Verb Phrase structure), but that the remaining comparisons yielded very little in the way of significant differences. Government Documents and Learned journals differed from Newspaper Reports and Popular Journals on only 3 variables: transitive main verb frequency, adverbial clause frequency, and the number of noun adjuncts. Newspaper reports differed significantly from Popular journals only in their proportions of declarative sentences. Learned Journals did not differ significantly from Government Documents on any variable.

Markworth and Baker's use of a one way analysis of variance at least shows that this statistical technique is sensitive enough to detect stylistic differences between styles. One major criticism of their method of scoring texts however is that they compared the average relative proportions of features per number of sentences for each sample and not the frequency with which particular authors made use of the structural possibilities available to them. For example, the

frequency of passive verb phrases was noted per 10 sentences, but as some sentences may be more elaborate in structure and contain more verb phrases this measure does not give direct information on how often an author chooses a passive rather than an active verb phrase, i.e. the proportion of VP's used that are passive.

The limitations of the one way ANOVA itself are such that variables are only examined singly for whether they distinguish different styles, whilst one can cite the total number of variables differentiating any two styles or two groups of styles there is no way of finding out whether all groups are significantly different from each other when all the variables are considered, or even whether some variables are in fact measuring the same stylistic effect. Markworth and Baker go on to remedy this deficiency somewhat by carrying out a discriminant analysis. This procedure is discussed in 1.7 below.

A discriminant analysis could have been used by Markworth and Baker as one method of checking up on the reliability of their style categories (the problem of establishing similarity discussed in 1.1 above). This opportunity is not fully realised, though the information presented (see 1.7 below) suggests that some of Markworth and Baker's styles do not form truly separate groupings.

1.6 SANDELL (1977) - Pearson Correlation Coefficients - Bi-variate study - comparing 2 D.V.s

This type of study is one in which the researcher tries to quantify the relationship between 2 sets of variable scores by statistical means. An example of such a technique is the Pearson product moment correlation coefficient.

Correlations can be produced for the scores of one particular group of an I.V. on two of the D.V.s but very often the I.V. groupings are ignored in this type of study. The emphasis anyway being on the relationship between one D.V. and another and not on that between I.V. and D.V.

The Pearson coefficient always varies on a continuous scale between +1 and -1. The sign in front of the coefficient indicates the direction of the relationship between the two variables and the magnitude of the coefficient indicates the closeness of the relationship. This means that a correlation of +1 represents a perfect positive correlation where scores on one variable are matched by scores on the other. -1 indicates a perfect negative correlation where for example high scores on one variable matched to low scores on the other. 0 indicates a total non-correlation between the two variables.

Pearson's coefficient takes account of both the ranking of scores and the absolute differences between them, (Kenny 1982:p79). Other

coefficients are available which work only on rank ordering, e.g. Spearman's rho.

Significance testing can be applied to correlations. Whether a correlation is significant or not depends on how large the correlation is (either positive or negative) and also the number of pairs of items being compared, (Kenny 1982:p88). This can be seen by examining the table of significance levels for Pearson's coefficient on the next page:

no. of pairs	lowest significant value (5% level)
3	.99
4	.95
5	.88
6	.81
7	.76
8	.71
9	.67
10	.63
15	.51
20	.44 from Kenny (ibid)

Because the lowest acceptable value gets smaller as sample size gets larger, even trivial correlations could be significant if the sample size was large enough.

Correlation coefficients may suggest (but cannot prove) evidence of



one of two kinds of relationships according to Kenny (1982:p86-7) either one variable causes the other (cause and effect), or both are being affected by a third unmeasured variable, or factor. Kenny contends' that the latter relationship is more likely in Linguistic studies. This is not necessarily true because when measuring proportions for example, longer sentences may have more noun phrases per sentence, and thus show a correlation. Such correlations however reveal only language universals<sup>and</sup> are of no interest to the stylistician (though he might be interested in the tendency of sentences being longer in some texts than others). Not all cases of cause and effect need be as transparent as this though, and great care therefore needs to be exercised in the interpretation of correlation coefficients.

No measure of correlation comes ready supplied with an interpretation, and it is at this point that subjectivity on the part of the stylistician is unavoidable. Even if he is sure that the results are not due to cause and effect he can only guess at the nature of the third unmeasured variable affecting the scores of both variables.

Sandell (1977:p124) provides an example of the use of Pearson's correlation coefficients with stylistic data. Sandell measured 5 texts in 5 samples all from the editions of the same newspaper. These were Advertisements from 1958 editions, Advertisements from 1968 editions, Foreign News telegrams from 1958 editions, Foreign News telegrams from 1968 editions, and Home Care editorials from 1968

editions. The samples had to have at least 100 words. Each group was made up of 14 samples. The 5 groups were measured on 12 (mainly syntactic) syntactic variables which Sandell believed might be associated with an Advertising style and intent to persuade. The variables were:

1. Average word length in letters
2. Average clause length in words
3. Average sentence length in clauses
4. Noun tokens as a proportion of total words
5. Verb tokens as a proportion of total words
6. Adjectives " " " " " "
7. Other parts of speech as a proportion of total words.
8. No of word types per 100 words not found in a comprehensive dictionary
9. Reinforcers as a proportion of total words
10. Ellipses per 100 words
11. Initial assonances per 100 words
12. End " " " "

Sandell's analysis was done manually by a team of counters (ibid:pl17).

Sandell used Pearson's correlation coefficients in an unusual way: He first produced correlation matrices for individual groups and then compared these with a correlation matrix for all groups taken together. The correlations were worked out between every possible

pair of variables (ibid:p123). Sandell directly compares only two correlation matrices one for the pooled groups and one for the separate groups. However he fails to explain how he produced a single correlation matrix for all the separate group correlations.

Sandell works on the principle that correlations for separate groups will not reveal evidence for persuasive intent but that pooled group correlations will. He therefore decides to assume that any increase in correlation magnitude of 10% or more on the pooled groups matrix when compared with separate groups is evidence that a variable is a marker of persuasive intent. Sandell then examines visually the intercorrelations between mutually correlating groups of variables and establishes a cluster of variables whose mutual intercorrelations have all increased by 10%. In this way Sandell is able to pinpoint a group of 7 variables all affected by a common factor, in a similar way to the sort of solution produced by mathematical means in a Factor Analysis (see section 1.8 below).

Sandell's technique is somewhat unusual. It would be more normal for him to have looked at differences in frequencies, (i.e. to take the variables separately rather than together) between one group and another or the pooled groups and one group.

In summary of Sandell's findings, he claims to discover increases in the tendency to use adjectives, ellipses and initial assonances due to persuasive intent. Note that Sandell's interpretation of the correlations in this way is due to an informed but subjective

assumption about the absence of persuasive intent within groups and its presence in pooled groups. Sandell (ibid:p123) admits this weakness.

The major limitations of correlation coefficients appear to be that they can only compare two variables at once. This only goes part of the way to solving the researcher's problem that rather than all measuring different aspects of style some of the dependent variables at least are measuring more or less the same thing. This problem was first raised in the section on ANOVA above.

However Sandell (ibid) does show that careful interpretation of a manually produced cluster of intercorrelating variables can help the researcher to proceed beyond the process of simply establishing relationships between 2 variables to positing the existence of unmeasured variables.

#### 1.7 MARKWORTH AND BAKER (1980) - Discriminant Function Analysis - Multivariate Comparisons Taking account of the I.V.

Discriminant Function analysis, sometimes referred to simply as Discriminant analysis, is a technique which begins by assuming that the groups of the I.V. are distinct and looks at which groups of variables most successfully show up the grouping (i.e. those that have markedly different scores for different groups).

Markworth and Baker (1980:p237) explain that this technique

endeavours to maximise the differences between the groups of the I.V. by creating functions of variables (discriminant functions) on which individual variables are given different weightings (some variables being more important in discriminating groups than others). These weightings are chosen as those which give a mean on the function of variables which differentiates 'best' between 2 groups according to Ehrenberg (1975:pp272-273). If there are more than two groups of the I.V. to be discriminated, more than one function may be required to discriminate between them. The number of functions produced is always one less than the number of groups being discriminated.

The correlations between each variable and each discriminant function can be worked out to give information on which are the most important variables participating in distinctions between groups.

Discriminant analysis in summary is a means for discovering which variables show the clearest distinctions between established groups. It also however can be used to obtain a good deal of additional information of interest to the stylistician. The nature of this additional information is described below.

The scores, on all variables measured, of individual cases, can be compared by Discriminant analysis with the functions produced, to find out into which of the established groupings of the Independent variable they best fit: Discriminant analysis might assign individual cases to the same groupings as those determined a

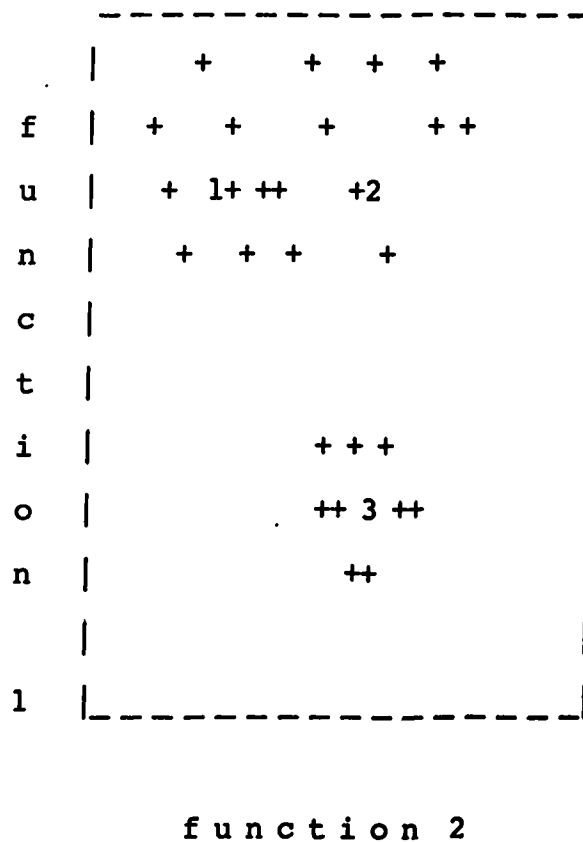
priori by the researcher, or alternatively to different groupings, indicating by the latter that they are more like the majority of individuals in that different grouping. The output of this information is option 14 on the version of Discriminant Analysis in the SPSS-X statistical computer package (SPSS inc:1983). It is clearly a useful method of checking on the researcher's own subjective stylistic text groupings.

All cases can be allotted scores on the functions produced by discriminant analysis, and a centre point (known as the group centroid) for the scores of each group (representing a different style) can be calculated. A plot of the scores on any two functions (plus the group centroids), with one function plotted on the x axis and the other on the y axis, can provide a visual display of how much stylistic distance exists between the co-ordinates of one groups members (Style A) and those of another (Style B). This gives an indication again of how well the styles are differentiated, as well as information on how similar individual texts are in style through examination of the distance of each group's co-ordinates from its centroid. Fig 1.7 (over page) gives a plot of fictitious data to clarify these points.

In Fig 1.7 both group 1 and 2 are not very clearly separated even though there is some distance between their group centroids. Both groups 1 and 2 are less compact than 3. Function 1 fails to discriminate adequately between groups 1 and 2 (whose group centroids have the same score on this function) though function 2

discriminates them somewhat. Group 3 is not discriminated from group 2 by function 1 but by function 2. Both functions 1 and 2 discriminate groups 1 and 2 but function 1 makes the clearest distinction. In order to get the full picture further analysis of the variables is needed in order to see which ones correlate highly with function 1

Fig 1.7 Plot of fictitious scores on 2 discriminant functions for 3 style variables.



There are 9 cases in each group  
 + = co-ordinates of individual cases  
 1,2 and 3 = Group Centroids

Markworth and Baker (1980:239-245) carried out a discriminant function analysis on data obtained on 36 syntactic variables and 5 style categories with 100 10 sentence long texts each. The 5 style categories were; Fiction, Government documents, Learned Journals, Newspaper reports, and Popular Journals. They report on only the first two discriminant functions which they consider gave a clear differentiation between fiction and non fictional styles. Government Documents were not well discriminated from Learned Journals by either of the two functions mentioned, the same was true for Popular Journals and Newspaper articles. However Government Documents and learned Journals together are differentiated from Popular Journals and Newspaper Reports (taken together) by functions 1 and 2.

Because of the failure of the first 5 group discriminant analysis to separate the non-fiction styles satisfactorily, Markworth and Baker went on to do a 4 group discriminant analysis on the 4 non-fiction styles. Markworth and Baker report on all the three functions for this analysis. Function 2 served to discriminate Newspaper Reports from Popular Journals, and function 1 served to discriminate both of these from Government Documents and Learned Journals, but only function 3 discriminated these latter 2 styles from each other.

Markworth and Baker report that 8 variables correlate with the first function in the 5 group discriminant analysis, these are



predominantly variables measuring verb phrase structure. For the 4 group discriminant analysis the number of variables correlating with each discriminant function are reported to be less, for example, nominalisations, passives and conjoined words all correlate positively with function 3 and noun adjuncts, and negatively with the same function.

Whilst Markworth and Baker make use of the possibilities of discriminant analysis to show how well different styles are differentiated, they do not use its capabilities to check up on their established style groupings. In view of their difficulties in managing to get functions to clearly discriminate all their styles it is possible that some of their texts are 'misplaced' and that Markworth and Baker's style groupings are not the optimum way of categorising their data stylistically.

The main limitation of Discriminant Analysis appears to be that it tries to fit texts into established groupings rather than looking at completely ungrouped data with a view to constructing the most natural groupings. This task is perhaps best undertaken by the large group of techniques collectively as Cluster Analysis, though there are distinct problems with this method as regards the definition of what is a 'cluster'. Cluster Analysis is further discussed in 1.9 below. As a means for providing groupings of style variables all functioning in a similar.

Discriminant Analysis differs from Factor Analysis (discussed below

in 1.8) in that the groupings of variables which result in Discriminant Analysis are based on the grouped cases whereas in Factor Analysis these are ungrouped. The result is that the emphasis in Discriminant Analysis is on finding groups of variables which best differentiate styles whereas in Factor Analysis the emphasis is on finding groups of variables which best differentiate cases (in this case individual samples of text).

#### 1.8 BIBER (1985) - Factor Analysis - Comparing more than 2 dependent variables

The basic principle of a Factor analysis is to reduce a large number of initial variables to a smaller number of derived variables, thereby summarising a section of the original data. Each Factor is a representation of the shared variance of a number of variables. In the case of a stylistician measuring texts on a large number of linguistic variables, each factor represents a group of linguistic features that have a tendency to co-occur in all texts, regardless of their initial grouping by the researcher into particular styles.

Ehrenberg (1975:p272) points out that Factor Analysis in common with many other statistical techniques should not be assumed to be more objective than it really is. Factor Analysis as Kim and Mueller (1978a:p9) point out, is a convenient label for a group of analytical procedures where there is in fact a good deal of choice (and therefore opportunity for making subjective decisions), rather than a technique with one fixed pattern for all eventualities. To

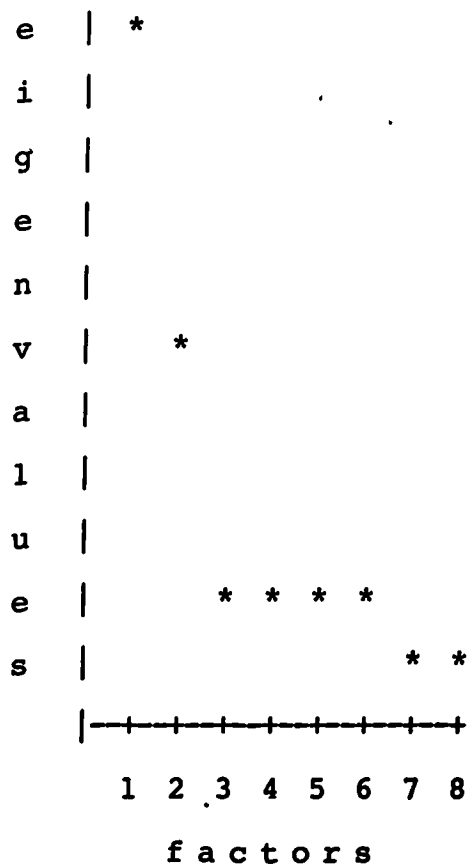
understand this latter point fully an outline of the factor analytical process is needed, and now follows.

The first stage in any type of Factor Analysis is the calculation of a matrix of Pearson Product Moment Correlation coefficients (described in 1.6 above). The number of variables going into the correlation matrix will of course depend on how many variables the researcher has chosen to measure, but a heuristic often suggested is to limit the number of variables to no more than half the number of cases or subjects in the study. Biber (1985:p342), however, (whose Factor Analysis on stylistic data is studied below) quotes Gorsuch (1983) as suggesting only one fifth the number of variables compared to observations. This would involve making a large number of observations per variable. Such a task caused no problems for Biber, who used a computer to count the scores for each observation, but is likely to be beyond the scope of a manual count if a useful number of variables are to be counted. The nature of the actual variables to include in the analysis is of course a subjective decision and this must be taken into account when examining and interpreting Factor Analyses.

In the second stage, weighted linear combinations of the original variables are calculated using the correlation matrix, these are the factors. Initially the factors 'extracted' from the data will contain a relatively large proportion of the shared variance among the variables (i.e. will discriminate cases well) but as more are extracted the percentage of variance accounted for becomes less and

less. Factor Analysis will go on extracting factors until it has accounted for all the shared variance. This often results in a very large number of Factors. This situation calls for an assessment of which factors are really worth considering and which are trivial. This is essentially a subjective decision, but one helpful suggestion made by Kim and Mueller (1978b:p44) is to draw up a graph of the factors and their eigenvalues (a measure reflecting the amount of shared variance accounted for by each factor. It can give a plot similar to that in Fig 1.8 below.

Fig 1.8 Plot of eigenvalues and factors for fictitious data



Such a plot is often known as a 'scree plot' because whilst the first two co-ordinates give a steep slope the remainder of the factors form as sort of scree as at the foot of a mountainside. The suggestion is made by Kim and Mueller (op cit) that the experimenter should ignore the factors forming the 'scree' and pay attention only to the first two. The plot on the previous page shows a very clear cut case but had factor 3 for example had a higher eigenvalue it would have been difficult to decide whether to include it or not. Even with the help of a scree plot then the decision about how many factors to extract remains subjective.

The final task of interpretation of the factors is done by the researcher and is therefore subjective, though guided by previous knowledge. Interpretation is facilitated by a list of the correlations of the original variables with the new derived variables or factors. These correlations are often termed 'loadings'. If several variables 'load' highly on one particular factor, and the researcher had subjectively identified these variables as those most indicative of informal texts, he might interpret the factor in question as a formality factor.

This final task and the one mentioned earlier of deciding how many factors to extract also arise for Discriminant Function Analysis, although in the case of the latter the problem is that of deciding how many functions to bother with and what they mean.

In order to make the job of interpretation as easy as possible, a

necessary prerequisite according to Biber (1985: p349), is to subject the factors to a 'rotation' which will result in variables correlating strongly with only one of the factors extracted, and not several. Many different rotations exist, some more widely used than others. Detailed information about rotations is not provided here but a comprehensive treatment of the subject is to be found in Kim and Mueller (1978b). The choice of one particular rotation rather than another is of course yet another opportunity for the researcher to exercise his own judgement. The rotation chosen changes the composition of the resulting factors. However, no one rotated solution according to Ehrenberg (1975:p267) can be considered 'correct' in Factor Analysis.

A final possibility in Factor Analysis is to calculate the scores (factor scores) for each case on the new variables or factors. In Factor Analysis proper these scores are estimates only (Ehrenberg 1975:p269). This is because Factor analysis always assumes that a certain amount of variance in the data is due to error. In the related technique of Principle Component Analysis no such assumption is made and therefore true component scores can be computed if the assumption of no error is correct, though this is rather unlikely.

The factor scores obtained can be used to give an idea of how well the different groupings of the independent variable are separated from one another. This can be done by plotting the scores on two different factors, on separate axes, for all cases in the data. The plot produced is reminiscent of the plot of discriminant functions

in fig 1.7 above. If all the scores for each group form dense clusters which do not overlap this is an indication that at least as regards the factors extracted the set of texts in each I.V. group are a homogeneous set. This is information of a very useful type for a stylistician.

Biber (1985) conducted a Factor Analysis on stylistic data made up of 545 observations (texts) measured on 42 variables covering aspects of Linguistic structure. Biber (1986) shows that the choice of variables was determined by examining relevant previous studies. The 545 texts were classified into 13 different groups according to provenance. These groups covered both spoken and written texts of varying degrees of formality. Biber (1986) shows that his study is mainly concerned with identifying the factors distinguishing spoken from written styles. Scores on all Biber's variables were calculated by computer techniques.

Biber carried out a Factor Analysis on his data and decided to extract 5 factors on the basis of a scree plot, (ibid 1985:p349). He rotated the resulting factor matrix using a Promax rotation (not available through the computing facilities at Bangor), and says that he interpreted 4 of the resulting factors. Biber calls his first factor "Interactive v Edited Text" (ibid 1985:p356) and reports 15 variables with positive loadings and 3 with negative loadings. Biber ignores all loadings smaller than .30.

Biber's second factor loads positively on 7 variables which Biber

considers indicative of an abstract style (p354). These are variables like the frequency of passives and nominalisations. The second factor also displays 5 negative loadings all considered by Biber to be indicative of frequent reference to the external situation. These negatively loaded variables include variables like time and place adverb frequency. The variables on which Factor 2 loads lead Biber to interpret this as an "Abstract v Situated Content" Factor

Biber (ibid:pp354ff) goes on to compute factor scores for all his factors and then plots the distribution of the mean factor score (on the first factor only) for each text/style grouping. Note that this is not the method of plotting factor scores mentioned above on page 260, where a check on group membership is facilitated. All in fact Biber is doing at this point is showing visually how far apart the means of each text group are considering their mean scores on Factor 1. Some measure therefore, of how well each group is shown to be stylistically different from its neighbours, on one factor, is all that is possible with this type of plot. Visually the plot shows 3 well separated groups. The first group is made up of clearly interactive spoken texts. The other two groups are non-interactive the first of these containing more spoken and less edited texts than the second.

Biber shows through his study that stylistic data can be analysed through the use of factor analysis and that meaningfully interpretable results can be obtained. Factor analysis emerges in



Biber's study as a useful tool for the stylistician in trying to pinpoint groups of variables all contributing to produce the same stylistic effect. Biber however uses a very large dataset well beyond what will be feasible in this study where only manual counting techniques will be used. The reliability of a study much smaller than Biber's would perhaps be more questionable.

Biber's study is disappointing in that he makes no attempt to check on the homogeneity of the original text groupings used in his sample. It seems likely for example that some rather vaguely defined categories in Biber's study such as his "Face to Face Conversation" are likely to contain a good deal of internal stylistic variation, for example from status differences among speakers. In addition, Biber's method of plotting the mean factor scores for each text grouping is very misleading as the factor scores for a few cases in each group may vary a great deal from the scores of the majority of cases. In this eventuality a mean factor score would not be a good representation of the scores of a total grouping and would of course mask untypical scores.

The following section investigates a possible method of solving the problem of testing for group homogeneity when all observations/texts are considered separately. This is the group of techniques known as 'Cluster Analysis'.

1.9 McENTEGART and LE PAGE (1982) - Cluster analysis - Grouping observations into natural clusters

The term 'cluster analysis' is misleadingly not a single method of analysis but a large number of techniques working on many different principles as Everitt (ibid:pp23-58) shows. Furthermore according to Baker (1985:p75) Cluster Analysis is a technique for sorting data. It is therefore not a statistical test like Discriminant Analysis (i.e. it does not involve significance tests usually).

Cluster Analysis is based on the assumption that in any group or cluster of entities each member will have more in common with its fellow group members than with the members of other groups. It is a multivariate technique in that the D.V. scores for all observations/texts on a number of variables are used as input. Prior groupings, for example into style categories are ignored.

In terms of their actual calculation nearly all Clustering techniques have in common the fact that they are based on a measure of distance related in some way to the performance of each subject/text on all the variables used. Cases are 'fused' into groups on the basis of the distance measure chose. The groups formed are subsequently fused together on the same basis until all the data has been accounted for, often by the fusion of all groups into one large group.

Clustering techniques vary according to Everitt (ibid:p52, pp60-64 & p75), in relation to; the assumptions they make about the data; the shape of clusters they tend to find; whether the clusters overlap; and the exact composition of the clusters. Added to this is the fact that many clustering methods require the analyst to decide subjectively, or in advance, how many clusters he wishes to find in the data, (Everitt op cit:pp64-67).

The consequence of this multiplicity of approaches is that different results may be obtained from the same data. This situation would not be so bad if a reliable system of validating clustering solutions were to be available. No such procedure exists according to Everitt (1980:p74) only heuristics like those suggested by McEntegart and Le Page (1982:pl16). The simpler of these involve trying to replicate the same clustering solution using several different methods, and supporting this with other non-cluster analytical representations of the data (for example plotting factor scores as in 1.7 above); and to check for concurrence with the researcher's own qualitative assesment of what solutions seem likely.

Another obvious area for subjective judgements with clustering techniques is that of where to stop clustering and consequently how many groups to accept. This may be fairly clearcut if a stage is reached where several very close knit groups have been obtained, all of which are very distant from one another. However if no such clear boundary exists a large set of groupings may be just valid as small set.

In spite of the inherent problems in Cluster Analysis the technique has been used in linguistic research, though not to the present author's knowledge in stylistics. McEntegart and Le Page (1982) use two different clustering techniques; Single Linkage, and Ward's Method, on *Diala* data. McEntegart and Le Page, working in Belize and St Lucia, obtained samples of speech recorded under varying conditions assumed to influence formality. The authors also scored informants on a variety of non-linguistic (i.e. cultural and economic) variables. Linguistic variables chosen for analysis were phonological variables thought to be socially marked. McEntegart and Le Page performed Cluster Analyses on the scores of all their informants on all the phonological variables. The resulting clusters were then tested for their degree of association with the socio-economic variables measured.

The technique used by McEntegart and Le Page could be transferred to stylistic work with individual informants being replaced by texts and the phonological variables of presumed social significance replaced by linguistic variables of stylistic significance. The clusters resulting from the analysis could be compared against original groupings (i.e. the same conditions on the I.V. as used for significance testing e.g. with ANOVA) determined by the researcher. This comparison would be qualitative and not involve the use of a test of association as did McEntegart and Le Page. The importance of making this sort of check on presumed stylistic groupings of texts has already been stressed on p.236 earlier.

McEntegart and Le Page did not obtain satisfactory results with Cluster Analysis on their data. They were unable to find any discrete non-overlapping clusters (McEntegart and Le Page's term for these are 'true typologies'), and only 4 overlapping (administrative) clusters for the Belize informants. No stable clusters could be found at all for the St Lucia data set (op cit:pl17).

McEntegart and Le Page's lack of success is not discouraging as regards the present study. In the qualitative study of government leaflets already completed (see part 1) very clear cut differences between at least two easily identifiable text populations have been observed (i.e. those between P.A. and I.A.). McEntegart and Le Page admit that the populations in their study were 'in a state of flux' and 'not ghetto populations' (op cit:pl16). Their lack of success with cluster analysis does not therefore seem surprising, and indeed appears to confirm intuitive judgements about what might be the outcome with this type of dataset.

Cluster analysis it seems can be cautiously accepted as an exploratory tool in stylistic work, there are too many problems with validation for it to be more than this. It does seem worth including in the present study however as the only statistical technique available expressly for sorting and classifying datasets without using any a priori grouping.

#### 1.10 General Summary of Information provided by the Statistical techniques reviewed

In summary the statistical techniques discussed in this Introduction as being useful for stylistic work fall into three main categories. Firstly techniques like Cluster Analysis, the plotting of Factor Scores and of discriminant functions which all provide checks on the status of the initial stylistic groupings of texts set by the researcher. Secondly the ANOVA to assess whether individual linguistic variables produce significant stylistic differences between pre-established text groupings. Thirdly techniques which facilitate the identification of variables functioning together to produce a particular stylistic effect, either disregarding the researcher's original stylistic groupings (Factor Analysis and Correlations) or taking these into account by trying to see which groups of variables maximally differentiate between the text groups (Discriminant Analysis).

#### 1.11 Basic Requirements of all the Research Techniques described above

All statistical techniques make some assumptions about the data to be input to them. These assumptions take the form of requirements about dataset parameters and must be adhered to if dependable results are to be obtained (though researchers commonly gloss over them). In addition to this, the nature of the study conducted, whether Experimental or based on Natural Observation will determine

exactly what type of conclusions can be drawn from the results. These points, of general relevance to all the techniques discussed above, are now discussed.

#### 1.11.1 Experimental v Observational Studies

The distinction between research founded on Experimentation and that founded on Natural Observation is an important one because it determines the exact nature of the conclusions to be drawn from the research. Robson (1973:p15-16) sums up the distinction as being one of control over the variables involved. In an experiment the researcher investigates the relationship between two phenomena by making changes in one and noting the effect of these on the other. The phenomenon which is subject to deliberate change is the Independent Variable (I.V.) (in a much stricter sense than that referred to before on page 239) and that in which change is observed the Dependent Variable (D.V.). In a carefully controlled experiment where nothing else is allowed to vary, or all other variations are totally random, the I.V. can truly be said to cause changes in the D.V., if any changes are found.

Because of the difference between experimental and non-experimental studies, the term 'Independent Variable' is sometimes replaced by the term 'Explanatory Variable' for a variable that plays the role of the I.V. but is not in the strict control of the investigator.

In a study based on natural observation, for example investigating

the relationship between smoking (I.V.) and cancer (D.V.), the I.V. is an attribute of subjects (unless you took a group of people and made some smoke and others not for x years) rather than something decided on by the experimenter. Other uncontrolled variables may show non random correlations with the tendency to smoke e.g. suffering from stress, and could themselves be the cause of cancer. These unwanted variables do not vary randomly and are beyond the researchers control. It is for this reason that a study based on natural observation can not be used as evidence to show cause and effect as in an experiment, but only evidence of an association between 2 variables. The association discovered must then be interpreted in the light of what other possible uncontrolled factors could have affected the result.

No true experiments have been carried out to date on stylistic data to the present author's knowledge. This would involve, if carried out, having texts written under instructions which could be assumed to affect their style, then measuring each text on a number of linguistic variables.

The only author who appears to consider an experimental design is Sandell (1977:p11) in his study of advertising language. Sandell concludes in favour of an observational study noting that whilst an experimental design is superior in the conclusions that can be drawn from it, any texts produced under experimental conditions would not be as authentic as a random sample of available texts. Sandell goes on to make the point that to some extent unwanted variables can be



controlled by careful restriction of the text populations for study, e.g. they can be standardised for time of writing/speaking. The same view as that of Sandell will be adopted here.

#### 1.11.2 Random Sampling

It is generally neither practicable or necessary for a researcher to submit all the available texts in a given style under study to analysis. The total set of such texts will probably be beyond estimate. The name given to such sets is the population. Leach (1979:ppl8-19) points out that a researcher usually investigates the parameters of any given population by analysing only a small portion of it, a sample, and then generalising the information obtained to the parent population. Such generalisations are only valid of course if the sample of subjects from a population is completely random. This means that although every individual is not investigated all individuals have an equal chance of being investigated.

Random samples can only be drawn if there is a clearly recognisable population to sample from. A frequent fault of statistical stylistic studies is to fail to make explicit what the populations being sampled are. This is true of of Carroll (1969). More recent studies using computer rather than manual counting of variables have used large corpora (in effect samples) (or sections of them) like the Brown Corpus as populations. In this case the reliability of the sampling depends on how well the parent corpus was sampled.

### 1.11.3 Underlying Population Distributions.

All the statistical techniques described in this chapter (except most Cluster Analysis ones) so far make stringent assumptions about the nature of the distributions underlying the populations from which samples to be tested are drawn. This is a natural requirement of the particular statistical concepts on which the techniques are based. All the techniques discussed so far are generally referred to as parametric statistics and require that the distribution of scores in the parent population, if sampled, would be 'normal'. There is nothing abnormal about other population distributions, but it can be shown that many statistical phenomenon, like human height, if plotted, display one particular symmetrical shape whose equation is well known. This is the normal distribution.

There are also other assumptions about similarity of 'variance', for example in ANOVA. Here the assumption is that the spread of scores in each group around its mean is not markedly different from that in another group. Data for analysis by ANOVA should be checked to make sure that this assumption holds true.

Parametric statistical techniques are often seen to be superior to those techniques which do not make the same assumptions about underlying population parameters (non-parametric statistics). This superiority is based on two things: the greater sensitivity and power of a given parametric test when compared directly with a non-parametric test designed to test the same thing, and the ability

of parametric statistical tests to tackle complex problems for which non-parametric statistics are not available, e.g. factor analysis.

Supporters of Non-Parametric Statistical Methods however point out the difficulty of establishing with certainty whether underlying population distributions are normal and the consequent gamble that is being taken with the statistical significance of the work done. Non-Parametric statistics tend to be mathematically simpler than their counterparts and therefore easier for non-specialists to understand.

For more complex studies where non-parametric statistics are not available, e.g. factor analysis, a possible method is to do only the first stage of the analysis (factor analysis is generally based on an initial correlation matrix, using a non-parametric technique e.g. Spearman Rho instead of Pearson 'r'). The resulting matrices could be processed by hand (see section above on correlations) to reveal groupings of mutually correlating or dissimilar variables. Kendall (1975) seems to think that acceptable results can be produced in this way.

Woods et al (1986) maintain that normally distributed scores on a variable in the parent population cannot be assumed in Language data except where samples are very large (scores approach normal as sample size increases due to the effect of the Central Limit Theorem - see Woods et al:op cit). If this is taken into account in stylistic work it seems that the rational choice for statistical

testing (i.e. confirmation of hypotheses) are non-parametric tests, the onus being on anyone who uses parametric statistics to show that the distributions concerned are normal if they opt to choose these more powerful tests.

Where instead of confirming hypotheses the researcher only wishes to explore the structure of his data to see what patterns emerge, and does not wish to confirm previously formulated hypotheses, it seems more admissible to use parametric statistics. This is true especially where an equivalent non-parametric test is not available. These results can be compared with manual analyses of non-parametric test outputs (see above). In this case however any generalisation of the researcher's findings to the populations is not really legitimate.

Before concluding this section, it must be noted that non-parametric tests are not entirely free from assumptions, the main ones being that the scores for analysis must be (when significance tests are being done) from a randomly sampled population (see 'Random Sampling above), that scores should be independent (i.e. not from the same individual/subject), and that scores should be continuous (i.e. all subjects should have different scores). Violations of the latter assumption can be corrected if not too severe, according to Leach (1979:p50). The assumption of independence for stylistical research means that in effect each text must be from a different author.

### 1.12 Implications of the foregoing for the Present Study

This study will clearly have to give consideration to two separate but related problems:

1. Exploration of the validity of style labels like AdE, Personal Address, and Impersonal Address using sorting and classification techniques like Cluster analysis.
2. Investigation of hypotheses about the role of linguistic variables believed to characterise the style of Government Leaflets, as well as further exploration of the possibility that groups of variables may share the same stylistic function.

### 1.13 Choice of Statistical Techniques in the Present Study.

Ideally all the different types of statistical techniques discussed in this introduction need to be used in the present study. This is because of the exploratory nature of the work to be undertaken and the consequent need to shed as much light as possible on the data. The desirable techniques to use in the present study are Parametric ones, and (in the case of Factor Discriminant Function Analysis) are the only available techniques. It has already been established that parametric techniques rely on assumptions about normality which cannot be guaranteed in Linguistic data.

Some method of testing the assumptions about normality for each and

every variable is needed in order that an informed decision can be made about what techniques to use. A statistical test of the null hypothesis that a sample of scores does come from a normally distributed population is possible with the Kolmogorov-Smirnov test, Leach (1979). This test would have to be carried out on each variable several times; once on the scores of each group separately as a prior to an ANOVA or similar test, and once over all the groups as a prior to all other tests.

The necessary prerequisite to the use of the ANOVA for individual variables are variables where it is possible to accept the null hypothesis of an underlying normal distribution for every group's scores and 'homogeneity of variance' (page 271). Even if the scores of only one group did not satisfy the requirements of normality it would not be possible to use a parametric test. The likelihood of every variable in this study satisfying the assumptions about normality in this way seem very unlikely: In the interests of uniformity, if anything more than one or two variables were to be non-normally distributed, a non-parametric test seems to be the best solution rather than using an ANOVA on some variables and the Kruskal-Wallis test (non-parametric equivalent to ANOVA) on others.

The situation is slightly more complicated with regard to other techniques. As a prerequisite to the use of a parametric correlation co-efficient both variables in the comparison (a correlation compares 2 variables) would have to satisfy assumptions about normality across all groups. Whilst satisfying this requirement seem

more likely than satisfying the requirements for the ANOVA, a uniform test is needed especially if a manual factoring technique is to be attempted (i.e. a technique involving multiple comparisons of many correlation co-efficients).

As a pre-requirement to Factor Analysis all the variables included in the technique would have to satisfy assumptions about normality across all groups. If only a small amount of variables failed to satisfy these assumptions then a Factor analysis may be possible on the majority of variables. If on the other hand many variables failed the test of normality, then a decision is not so easy to make. The choice is either to proceed with Factor Analysis on all variables anyway and treat the results with great caution, using a manual method to provide further support. The results of the Factor Analysis would therefore be only descriptive (no inference to populations, significance tests etc). The alternative is not to do factor analysis at all, instead satisfying oneself with only a manual analysis of multiple correlation co-efficients. The former option, provided that the result of the Factor Analysis is treated with caution, seems likely to provide the greatest amount of information.

The Cluster Analysis to be used in this study is non-parametric and therefore the problems discussed above do not arise. The problem does arise however of whether to standardise scores or not.

Discriminant Analysis requires all variables included in the

analysis to satisfy the assumptions about normality across all groups. If most variables fall into this category then Discriminant analysis should be performed. Even if this is not the case the Discriminant analysis could be used as long as the results were interpreted with great caution. In this respect the use of Discriminant Function Analysis is very like Factor Analysis.

#### 1.14 Plan of the Present Study

The design of the present study is set out in detail in the Method (Chapter 2). However, it is important to set out here the main factors to be considered in the design. These are based on the information obtained in the discussion of previous studies which forms the major part of this chapter (sections 1.5 to 1.9).

The main factors to be considered in the plan are as follows:

1. The formulation of questions to be answered. These are the hypotheses to be tested in Part 2.
2. Decide on an Experimental or Observational plan
3. Identify the Populations to be sampled and decide on a sampling technique. This by implication involves deciding on how many groups there should be on any independent variable.



4. Define the size of the data matrix to be used. The total number of scores which can be recorded will be determined by the time limitations involved alone, but for any given matrix of scores there are different permutations of numbers of variables and numbers of observations. which it seems fruitful to analyse. For example, a data matrix of 100 scores could be formed by 10 observations on 10 variables, or by 5 observations on 20 variables.

The choice of the exact combination of observations and variables is a delicate balance between several factors, firstly the need to include enough texts/observations on each I.V. group to make the study reliable, secondly the need to investigate a usefully large number of variables, taking into account the wish to discover factors if existent and to test the qualitative observations made in part 1.

5. Decide on what variables to study. This will inevitably be determined by the need to test the qualitative observations made in part 1. How many variables to study will be partly determined by the time available and the the number of observations over which they need to be measured (see 3 above).

6. Decide on how many groups to compare and therefore how many categories the independent variable will have. This will be determined by what styles seem most stylistically interesting to compare, but obviously the comparison of any one style will be partial, as there is insufficient time to make comparisons with all styles.

7. Decide how to score texts on individual variables. The standard method used in all studies reviewed in this introduction is to simply count occurrences of a particular structural item per sentence or per 100, 200 words etc. This method was criticised in 1.5 earlier as undesirable and therefore an alternative method is needed, at least for certain variables.

8. Decide on the exact forms of the statistical techniques to be used in analysing the data. To a great extent this will be determined by the statistical packages available at UCNW Bangor.

The first point (point 1.) forms the last section of this introduction and is set out below.

#### 1.15 Hypotheses

Not all the statistical tests to be conducted in the study necessarily need to have a formal hypothesis formulated along the lines suggested in the section on hypotheses (1.15) earlier in this chapter. This is due to the distinction between confirmatory and exploratory analysis. Confirmatory analysis as its name suggests tries to confirm formal hypotheses about the sample concerned with a view to making valid statistical inferences about the population from which the sample is drawn.

Exploratory analysis can be used as a different way of approaching the same sample of data if knowledge is insufficient to allow the

formation of formal hypotheses. No inferences of course can be drawn from such a study, but observations can be made about the structure of the data in the sample which can be confirmed in later confirmatory studies.

The work so far carried out in this thesis has been qualitative (see Part 1) and concentrates almost exclusively on an itemisation of the features of linguistic structure which appear to characterise AdE and its component styles I.A. and P.A. Part 1 therefore provides sufficient evidence to make hypotheses about the distributions of individual items of linguistic structure in at least 3 styles; I.A. and P.A. AdE texts and at least one other style chosen to make a contrast between AdE styles and other styles of English. In the actual study 4 styles are used; I.A. and P.A. AdE, Information leaflets from Non-Governmental institutions, and Newspaper articles. More Information about the reasons for this choice of styles to study can be found in Chapter 2 which follows.

The work of Part 1 does not provide enough evidence to allow the formulation of hypotheses about the structure of variables when all the styles to be studied are taken into account, though it may be useful in helping to interpret any structure found. No formal hypotheses will therefore be made regarding the correlation and Factor Analysis parts of this study.

A tacit assumption does exist however to the effect that the different style categories to be studied in Part 2 are discrete and

non-overlapping categories. It is not possible to test such an assumption however. Of the four methods of testing for group homogeneity which have been suggested in this introduction, only one; Discriminant Analysis is a technique which can be validated and therefore used in confirmatory analysis. However, the requirement of Discriminant Analysis that all variables to be tested should be drawn from normally distributed populations seems unlikely to be realised for all variables. This technique may therefore be used only in an exploratory way, and as a consequence no formal hypothesis will be put forward concerning the homogeneity of the style categories chosen.

Finally it has already been suggested in the previous section that the Kolmogorov Smirnov test be used to test hypotheses about the normal distribution for each variable. These hypotheses are set out below, and are followed by those for testing for significant differences between the I.V. groupings on each variable.

#### 1.15.1 Formal Hypotheses about Normality

Ideally a hypothesis needs to be put forward for each variable to be tested. As there are in fact to be 38 of these, general hypotheses are put forward here in which K stands for Variable 1, variable 2 etc up to variable 38.

There are 2 alternative and 2 null hypotheses to be tested:

### Alternative Hypothesis 1

That the scores for group \* on variable k are not drawn from a normally distributed population.

### Null Hypothesis 1

That the scores for group \* on variable k are drawn from a normally distributed population, ('null' i.e. no difference between the observed distribution and the normal distribution).

\* refers to group: 1, 2, 3 or 4 of the I.V.

### Alternative Hypothesis 2

That the scores for variable K are not drawn from a normally distributed population.

### Null Hypothesis 2

That the scores for variable K are drawn from a normally distributed population.

1.15.2 Formal Hypotheses for testing significance of each of the 4 I.V. groups on all D.V.s

### Alternative Hypothesis

That the samples of scores for variable K come from different populations.

### Null Hypothesis

That the samples of scores for variable K all come from the same population. ('null' - no difference between the populations)

### Note

In the event of the null hypothesis being rejected for any variable it will also be necessary to test, for each pair of groups on the I.V., the alternative hypothesis, that:

The samples of scores for group (a) and group (b) come from separate populations,

and the null hypothesis, that:

The samples of scores for group (a) and group (b) come from the same population,

where (a) may be any of groups 1 to 4 and (b) may be any of groups 1 to 4 except that group chosen to be (a).

This section on the hypotheses completes the introduction to Part 2 of this thesis. The next Chapter sets out the Method followed in conducting the study.

## 2.1 Design

### 2.1.1 General Theoretical Aspects

'Style Variables' are those linguistic features whose distributions show an association with particular situational contexts, e.g. advertising, business English etc. Conversely Linguistic features which are not style variables should have distributions which are not associated with any particular situational contexts. Any apparent variations in the distributions of the latter type of variable should be the result of chance variations or sampling error. In practical terms it would be a huge task to sample all situational contexts to discover whether any non-stylistic variables exist. It would be much easier to show whether a variable is stylistic in one particular context

### 2.1.2 Theoretical Aspects Specific to the Present Study

Government information leaflets seem to show stylistic distinctiveness in that their writers appear to favour certain Linguistic structural features and neglect others. This apparent distinction is claimed in Part 1 of this thesis to be a likely function of the situational contexts in which Government Information leaflets are written. The greater situational context pertinent to Government Information leaflets appears to be that of the Province



of (Central) Government Administration, and the Modality of Information Leaflets. The Medium is that of printed i.e. edited, written English. The lesser Situational Contexts within the Language of Government Information Leaflets (GIL) are determined by relation to the situational dimension of Status, in particular the degree to which the readers of information leaflets are Personally or Impersonally addressed, e.g. as 'you' or as 'the claimant'.

### 2.1.3 Practical Aspects

Two possible designs for the investigation of GIL suggest themselves, the first experimental and the second based on natural observation. These two possibilities are set out as 1. and 2. below:

1. Experimental: Obtain a random sample of subjects (possibly Government leaflet writers) and provide them with instructions to write text on the same topic which will be Government Information leaflets of either a personally, or impersonally addressed type and possibly a non-government leaflet text for comparison. In this way the experimenter could assign texts at random to being P.A. or I.A. GIL or non-GIL.

2. Observational: Obtain 2 random samples of authentic Government Information leaflets, one each of I.A. and P.A. texts. These would be selected on formal criterion which might be assumed to be related to how personally the texts address their readership, e.g. use of

the second person pronoun. For comparison one or more random samples could be taken from specifically non-GIL populations. In this study texts would be accepted as belonging to one of several different groups rather than being allocated to such groups by the experimenter.

Whilst design 1 has the notional advantage of being an experimental design, any benefits resulting from such, i.e. being able to establish cause and effect relationships would be lost unless the instructions to write text of a particular type were very reliable. Such reliability is likely to be very difficult to measure, because until this study is completed no very strong definition of what a GIL text is in Linguistic terms exists. In other words, just as people are not in control of gender (one cannot tell one person to be a man, another a woman) one cannot in practice get leaflet writers to switch between their styles. So just as comparisons of men and women have to be non-experimental, so do the comparisons of styles to be made here.

Moreover, the idea of an experimental study is not totally suited to the needs of the present study: This study sets out to explore the nature of Linguistic structure in GIL and to indicate links with situational variables, it is perhaps therefore too early to consider making inferences about cause and effect.

Design 2 above is also not without its problems either, the main one

being that of identifying clearly the populations being sampled, and choosing appropriate populations to compare. It is of little practical significance to be able to say that text group A differs significantly in style from text group B if this information can not be extrapolated to cover a clearly defined population of texts of which the text groups are only a sample. It is also of little practical significance if the contrasts between populations chosen for comparison could be due to any one of several situational variables, e.g. time of production, speech or writing, discourse or monologue etc.

Providing that the samples for design 2 can be taken from carefully selected and well described populations, this design appears to be the natural choice for the present study. The exploratory nature of this work requires that analysis be carried out on authentic texts, even if the results will be more limited in the interpretation which can be made of them. Future studies could be experimental with texts produced under experimental conditions compared with authentic texts in order to clarify their status as suitable material for research.

#### 2.1.4 The Independent Variable (I.V.)

After considering the time available to undertake the study, the number of variables desired, and the minimum number of texts desirable for each group of the I.V. it was decided that an I.V. with 4 groups was the largest possible study that could be attempted.

Group 1 on the I.V. is a sample of GIL Personal Address texts, and Group 2 a sample of GIL Impersonal Address texts. It was decided to compare these two groups with texts of the same leaflet modality but produced by Private Sector rather than Government Institutions. The information leaflets published by National Banks and Building Societies were selected to fill this need.

Group 3 will be referred to as BBSIL. This group was chosen because of the minimal differences that subjectively appear to exist between it and groups 1 and 2. It was expected that useful information about variables which relate to a general information leaflet Modality rather than the Public/Private sector Province distinction could be discovered in this way.

A second point of interest in relation to BBSIL emerged when some initial analysis was carried out: It became clear that all texts in the BBSIL sample were Personal Address. A comparison of BBSIL and GIL Personal Address with GIL Impersonal Address was therefore expected to yield interesting information on status differences.

For the fourth group of the Independent variable a situational context was needed which was totally unrelated to that of Information leaflets, but was nonetheless an example of printed edited text in wide Public circulation. The choice was an attempt to identify misleading variables in the other 3 groups which are relateable not to status or province but merely to the fact that texts are printed and edited with a wide public audience in view.

These variables would be those which failed to distinguish any of the 4 groups.

For group 4 the text population chosen was initially 'tabloid newspaper articles'. This was narrowed down to one newspaper only, after considering the evidence put forward by some authors (e.g. Crystal and Davy 1969, and Wallace 1981) that newspaper house styles may differ quite markedly. Random selections from among several newspapers may therefore have produced a non-homogeneous style which would not have been suitable for comparisons with other varieties. The Daily Mirror was randomly selected from a list of National Tabloid Newspapers to provide the sample for Group 4.

For all the groups concerned the assumption is made that each text sampled is from a different author and that scores for each text are therefore independent of each other. The design is therefore of the Independent Subjects type.

The variables of Modality, Status, Medium, and Audience have already been mentioned as variables taken account of in this study, the only other variable directly controlled for was Time. Time of publication was excluded as a factor affecting the results by standardising texts for for approximate date of publication or currency. This period is from October 1985 to the end of 1985.

All other variables not controlled for e.g. Dialect, Discourse Participation, Individuality etc are either obviously standard for

all texts as in the case of Dialect (Standard written English), or free to vary randomly as in the case of Individuality. Such variables are not expected therefore to affect the results of the study.

#### Summary of I.V. groups

1. GIL Personal Address
2. GIL Impersonal Address
3. BBSIL (Banking and Building Society Information Leaflets)
4. DMA (Daily Mirror Articles)

#### 2.1.5 The Dependent Variables

In the absence of previous research specifically dedicated to the analysis of GIL, the choice of dependent variables to study was made on the basis of the qualitative study conducted in part 1. Since Part 1 highlights more variables of stylistic interest than could possibly have been analysed in the time available it was decided to include only those linguistic features which appeared subjectively to make the strongest stylistic distinctions between P.A. and I.A. or those which seemed the most prominent markers of GIL style.

This selection of variables was made on an informed but subjective basis, and it was accepted at the time that the results of later analyses might show that some variables had been wrongly excluded. In order to lessen this problem as much as possible and to make sure

that no information gaps existed, some variables were added for the sake of closure. For example whilst complex and simple sentences were observed to make the main stylistic contrasts at sentence level, other sentence categories were investigated as well, i.e. mixed and compound sentences.

A further group of variables were added on which no qualitative study had been carried out. These variables were intended to extract information from the data about sentence cohesion.

The resulting list of variables was composed of 38. These variables are listed below as variables 3 to 40 (variable 1 is the I.V., and variable 2 is the number allocated to each text). Each variable is given its full definition, and listed under one of several categories.

#### VARIABLE LIST

##### SENTENCE STRUCTURE

3. Frequency of Complex Sentences - defined as the number of sentences containing an adverbial clause per 20 consecutive sentences. (relative or noun clauses are not considered as making a sentence complex in this thesis, see Part 1 Ch.6)

4. Frequency of Simple Sentences - defined as the number of

sentences consisting of only a single clause per 20 consecutive sentences

5. Frequency of Compound Sentences - defined as the number of sentences containing 2 or more clauses, all co-ordinated, per 20 consecutive sentences.

6. Frequency of Mixed Major Sentences - defined as the number of sentences containing at least 3 clauses, one of which must be adverbial and 2 of which must be co-ordinated, per 20 consecutive sentences

7. Frequency of Minor Sentences - defined as the number of any other independent sentence -like structure, not classifiable as any of 3-6 above because it is an incomplete set of sentence structures, per 20 consecutive sentences.

#### CLAUSE STRUCTURE

8. Frequency of conditional/concessive adverbial clauses - defined as the number of adverbial clauses which are either conditional or concessive, under the definition of Quirk and Greenbaum (1973: s11.23 - 11.28), per 20 consecutive adverbial clauses.

9. Frequency of Complement Clauses - defined as the number of clauses occupying complement position in a sentence per 20 consecutive sentence complements. Sentences containing this type of



clause are Simple sentences (i.e. counted at 4. above). Ch.6 Part 1 deals in detail with this.

10. Frequency of two-adverbial clauses - defined as the number of clauses, containing two or more adverbials (not adverbial clauses), per 20 consecutive clauses.

11. Frequency of a certain marked clause order - defined as the number of clauses where an initial adverbial was observed, per 20 consecutive clauses

12. Frequency of time (clause) adverbials - defined as the number of clause adverbials relating to temporal contexts, and defined as such by Quirk and Greenbaum (1973: s2.9), per 20 consecutive clause adverbials.

#### NOUN PHRASE (NP) STRUCTURE

13. Frequency of Pronominal Noun Phrases - defined as the number of noun phrases whose sole constituent is a pronoun, per 20 consecutive noun phrases.

14. Frequency of Complex NP premodification - defined as the number of NPs whose head is premodified more than once per 20 consecutive NPs

15. Frequency of NP Postmodification - defined as the number of NPs

whose head is postmodified, per 20 consecutive NPs

16. Frequency of clausal NP postmodifiers - defined as the number of NP postmodifiers which are clauses per 20 consecutive postmodified NPs.

17. Frequency of non-finite clausal postmodifiers - defined as the number of NP postmodifiers which are non-finite clauses per 20 clausally postmodified NPs.

18. Frequency of embedded postmodifiers - defined as the number of NP postmodifiers which are themselves postmodified per 20 consecutive NPs

19. Frequency of noun phrase quantifiers - defined as the number of NP quantifiers per 20 consecutive NPs

#### VERB PHRASE (VP) STRUCTURE

20. Frequency of passive verbs - defined as the number of passive marked VPs per 20 consecutive VPs

21. Frequency of agentless passives - defined as the of number passive marked VPs which lack an explicit agent per 20 consecutive passive marked VPs.

22. Frequency of past tense marking - defined as the number of VPs

marked for past tense (the -ed form) per 20 consecutive VPs.

23. Frequency of aspect marking - defined as the number of VPs marked for perfect or progressive aspect per 20 consecutive VPs

24. Frequency of modal verbs - defined as the number of VPs containing a modal verb per 20 consecutive VPs

25. Frequency of modal 'must' - defined as the number of modals which are 'must' per 20 consecutive modal verbs.

26. Frequency of modal 'may' - defined as the number of modals which are 'may' per 20 consecutive modal verbs.

27. Frequency of modal 'can' - defined as the number of modals which are 'can' per 20 consecutive modal verbs.

28. Frequency of modal 'will' - defined as the number of modals which are 'will' per 20 consecutive modal verbs.

29. Frequency of modal 'shall' - defined as the number of modals which are 'shall' per 20 consecutive modal verbs.

#### MEASURES OF PERIODICITY

30. Average paragraph length - defined as the average length of 10 consecutive paragraphs, in standardised lines of 10 words long.

31. Reading Test Score - defined as the average reading test score for 2 100 word samples using the FOG measure reported by Harrison (1980).

32. Average line length - defined as the average length in words (counting words separated by a hyphen as two words) of a line, calculated from 10 consecutive lines.

#### FAMILIARITY

33. Word Unfamiliarity score - defined as the proportion of 20 words picked at random which could not be found in West (1953).

#### SENTENCE COHESION

34. Total textual ties per 100 words - text ties were defined by reference to Halliday and Hasan (1976).

35. The number of reference ties as a proportion of 34 (above) - 'reference' is defined in relation to Halliday and Hasan (op cit).

36. The number of lexical ties as a proportion of 34 (above) - 'lexical' is defined in relation to Halliday and Hasan (op cit).

37. The number of ellipsis ties as a proportion of 34 (above) - 'ellipsis' is defined in relation to Halliday and Hasan (op cit).

38. The number of substitution ties as a proportion of 34 (above) - 'substitution' is defined in relation to Halliday and Hasan (op cit).

39. The number of conjunction ties as a proportion of 34 (above) - 'conjunction' is defined in relation to Halliday and Hasan (op cit).

#### TYPE - TOKEN SCORES

40. The proportion of word types in a sample of 100 word tokens (groups 1 to 3 only).

### 2.2 The Texts

80 texts were chosen in all (see appendix), 20 from each of the 4 populations already identified as GIL (P.A.), GIL (I.A.), BBSIL, DMA. Please consult 2.1.4 above for explanation of these terms.

#### 2.2.1 Sampling

For Groups 1 and 2 of the I.V. (the 2 GIL populations) all the current information leaflets were obtained by post from the following government departments. These departments were chosen as those government departments who issue information leaflets to the general public.

Department of the Environment

The Home Office

Department of Employment

HM Land Registry

Department of Education and Science

HM Customs and Excise

Department of Health and Social Security

Office of the Health Service Commissioner

Court of Protection

Manpower Services Commission

Department of Transport

Any leaflets of too short a length to provide enough scores on all variables were discarded. The texts were initially sorted into 2 populations differing on the stylistic dimension of Status; those that used the pronoun 'you' to address the reader and those that never use this pronoun to address the reader.

The stylistic dimension of Status is described by Crystal and Davy (1969: p74) to be a complex one. It therefore seemed possible that a simple distinction on the basis of personality of address might not yield clear enough status distinctions between P.A. and I.A. GIL. It was decided to use another subjectively chosen indicator of formality to produce a finer gradation of status differences in the data.

Friendliness of presentation to the reader was conjectured to be a variable that would provide a useful, but easily measureable, additional indicator of status differences. The use or not of cover graphics, (excluding departmental logos, insignia and badges), was chosen as the most easily accessible textual property which might yield such a distinction.

4 populations had now been identified:

1. I.A. GIL without cover graphics
2. I.A. GIL with cover graphics
3. P.A. GIL without cover graphics
4. P.A. GIL with cover graphics

It was considered that populations 1 and 4 would be the best separated in terms of status distinctions, and these were therefore numbered and then sampled randomly by means of numbers drawn from a random number table. Samples of 20 I.A. texts and 20 P.A. texts were drawn.

Some government Departments were very much better represented than others in the samples, this is because these departments have a greater information leaflet output than others. It was decided not to correct this naturally occurring bias in the data, because such a biased sample is arguably the most authentic: Potential readers are statistically more likely to encounter a leaflet from a large output

department than from a small output one.

A complete list of the leaflets for I.V. groups 1 and 2 can be found in appendix A.

The texts for Group 3 of the I.V. (BBSIL) were randomly sampled from the total population of national Banks and Building Societies represented in Bangor, Gwynedd. These are:

Lloyds Bank PLC

Midland Bank PLC

Alliance and Leicester Building Society

Abbey National Building Society

Halifax Building Society

Barclays Bank PLC

Trustee Savings Bank

National and Provincial Building Society

National Westminster Bank PLC

As with groups 1 and 2 leaflets of insufficient length were discarded. The leaflets were numbered and 20 leaflets sampled on the basis of numbers drawn randomly from a random number table. The complete list of leaflets in the sample is to be found in Appendix A.

For Group 4 the individual texts chosen were whole editions of the



Daily Mirror Newspaper. Editions were pulled from a large pile of Daily Mirror Newspapers covering the the period from 30th October 1985 to 23rd December 1985, it is acknowledged that this was not true random sampling. But it is unlikely that this will have any effect on the result of the study. These editions are not included in appendix A.

### 2.3 Procedure

Scoring of each text was done manually by the author alone, no problems of establishing inter-rater agreement were therefore present. For each variable a random page number was selected for each text by drawing numbers randomly from a random number table. In the case of newspapers a random article was chosen by a similar procedure, articles being counted from the top left hand corner of the page selected.

For variables 3 - 29 a random unit of the type in which the unit to be counted has a potential of occurrence was located and 20 such potential occurrences counted from that point. A score was recorded each time a potential occurrence was realised. For variable 30 (paragraph length), a random paragraph was selected and the length of this paragraph in words (and that of the following nine paragraphs) counted. The mean paragraph length in words was then calculated, and divided by 10 to give paragraph lengths in standard 10 word lines.

Variable 31 (Reading Test Score) was calculated by selecting a random sentence number and counting off one hundred words from this point. The number of sentences was counted in this 100 word sample and then divided into 100. To this score was added number of words of three or more syllables. A second sample of 100 words was scored and the mean of this score and the first score calculated.

Variable 32 (average line length) was calculated by selecting a random line number and counting the number of words in this and in each of the following 10 lines. The mean line length in words of these 10 lines was then calculated.

Variable 33 (Word Unfamiliarity score) was calculated by dropping a pin on a randomly selected page 20 times and each time noting the word closest to the pin's point. The word unfamiliarity score was calculated by scoring one point for each one of the 20 words which does not occur in West's (1953) word list.

Variable 34 (number of cohesive ties per 100 words) was calculated by counting the number of such ties per randomly selected 100 word sample. Variables 35 to 39 were calculated as proportions of the ties of particular types within each sample.

Variable 40 was calculated by locating a random starting point to count a sample of 250 running words for each text in groups 1 2 and 3. The text was read through writing down each new word, and noting

any subsequent occurrences of any word. Inflected forms e.g. plurals past tense forms etc were all counted as the same word. The result was an alphabetical index of each word type in the 250 word sample and its corresponding frequency of occurrence. It was subsequently possible to calculate the number of word types in the sample and express this as a proportion of the total number of words (tokens).

This section completes the Method. The chapter which follows sets out the Results of the study.

## CHAPTER 3 - PART 2 - RESULTS

The Format of these Results will include presentation of the results of the statistical tests carried out, but a tentative initial assessment of their linguistic and stylistic significance.

The raw scores for each Text on each variable can be found in the Appendix B.

### 3.1 The Kolmogorov -- Smirnov Test

The full scores for this test can be found in Appendix C.

#### 3.1.1 Within Groups

The first use of this test was to test, separately, the null hypotheses that each group's scores on each variable could be considered as being drawn from a normally distributed population.

For each variable this meant that 4 tests needed to be carried out.

It will be recalled from section 1.11.3 that the reason for carrying out this particular test was to see if individual variables could be analysed using the Parametric One way ANOVA or whether the Kruskal -- Wallis test based on ranks was more appropriate to this task. The information of greatest interest here is therefore not the result of the K--S test for individual groups on variables, but whether there are any variables which contain any groups for which the null hypothesis has to be rejected. This is because even one non-normal group on any variable strictly rules out the possibility of using the one way ANOVA.

The word 'strictly' is used in the last paragraph with good reason, since one has to admit that many workers in practice go ahead with their studies anyway. This is done on the general principle that violating a few 'assumptions' will not invalidate the whole study. It was safer in the present study however to be cautious and take the option set out on the previous page.

The Results for this first use of the K-S test are as follows:-

14 variables contained at least one group for which the null hypothesis was rejected at the 5% level, these are:

Variables: 3,6,7,9,18,19,20,21,24,25,29,36,38,39

The implication is that for these variables the alternative hypothesis be accepted for at least one of their constituent groups, that the sample concerned is drawn from a non-normally distributed population at the 5% level. A one way ANOVA cannot be used to analyse these variables.

For 24 variables the null hypothesis could not be rejected for any group at the 5% level these are:

Variables:

5,8,10,11,12,13,14,15,16,17,22,23,26,27,28,30,31,32,33,34,35,37,40

The implication for these variables is that the scores for every

group can be regarded as being drawn from a normally distributed population at the 5% level. A one way ANOVA could therefore be used on these variables.

Rather more than a third of the variables are ineligible for testing with the one way ANOVA, this is too many to reject, therefore a uniform test based on ranks the Kruskal-Wallis test will be used. Were variance differences to have been examined as well, this would no doubt have lead to the rejection of many more variables. It seems pointless to test for homogeneity of variance however since so many variables do not appear to have normal distributions.

### 3.1.2 Pooled Groups

The second use of the K - S test was made to find out whether scores for the pooled group scores on individual variables could be considered as being drawn from normally distributed populations. This use of the K - S test, it will be recalled, from sect 1.11.3 was chosen as a task essential to making decisions about the use of parametric or non-parametric correlation co-efficients, and helping to clarify the validity of using Factor, and Discriminant Function Analysis on the data.

The Results for this second use of the K-S test are as follows:-

For 14 variables the null hypothesis was accepted at the 5% level, these variables are:

Variables: 4,8,10,13,14,17,23,24,27,28,30,34,35,36

The implications for these variables are that their scores, over all groups, can be regarded as being drawn from normally distributed populations. These variables could therefore be further investigated by using the Pearson Product Moment Co-efficient providing that a scatter plot of any two variables showed the relationship between them to be linear and monotonic. These variables could also be included in a factor analysis without violating assumptions about normality.

For the remaining 24 variables, the null hypothesis was rejected at the 5% level and the alternative hypothesis accepted. The variables which yielded this result are:

Variables: 3,5,6,7,9,11,12,15,16,18,19,20,21,22,25,26,29,31,32,33,37  
38,39,40

The implication is that observations on these variables cannot be considered to have been drawn from a normal distribution at the 5% level, and therefore must be considered as being drawn from non-normal distributions. These variables could only be tested for correlation using a non-parametric correlation co-efficient, e.g. Spearman's rho (providing the relationship was monotonic), and could not be included in a full Factor Analysis without violating the assumptions of these techniques about normality, and therefore the

validity of their results.

Because as many as two thirds of the variables in the study have to be considered as drawn from non-normally distributed populations, there are clearly too many variables to reject in order to carry out factor and cluster analyses without violating assumptions about normality. As suggested therefore in sect 1.11.3 above Factor and Cluster analysis will still be used but as an exploratory technique, with extreme caution exercised in their interpretation. To help clarify the status of these exploratory techniques, a correlation matrix for all variables and a distance matrix, both based on ranks, i.e. non-parametric techniques, will be processed by hand to provide an approximation to a Clustering or a Factoring technique.

### 3.2 Kruskall - Wallis Test

The Kruskal - Wallis test was used to test, at the 5% level, the null hypothesis that the scores for all four groups are drawn from the same population. Results significant at the 5% level indicate rejection of the null hypothesis and acceptance of the alternative hypothesis that the groups are drawn from different populations. This does not however provide information about which groups are responsible for the result. This test was carried out for each variable separately. The results are displayed in table 3.2 (following page)



Table 3.2 Kruskal-Wallis Test - Results significant at the 5% level

Variable	Description	Probability
3.	Frequency of Sentences with adverbial clauses	.0012 sig
4.	Frequency of Simple Sentences	.0001 sig
5.	Frequency of Compound Sentences	.0006 sig
7.	Frequency of Minor Sentences	.0000 sig
8.	Frequency of Conditional/Concessive adverbial clauses	.0000 sig
9.	Frequency of Complement Clauses	.0000 sig
10.	Frequency of Clauses with more than 2 adverbials	.0065 sig
11.	Frequency of Marked Clause order	.0000 sig
12.	Frequency of Time Adverbials	.0000 sig
13.	Frequency of Pronominal NPs	.0000 sig
14.	Frequency of complex premodification	.0001 sig

Table 3.2 Kruskal-Wallis Test - Results significant at the 5% level

(cont)

Variable	Description	Probability
15.	Frequency of Postmodified NPs	.0053 sig
17.	F'ncy of non-finite clausal NP postmodifiers	.0000 sig
18.	Frequency of NP postmodifiers containing embedding	.0000 sig
19.	Frequency of NP quantifiers	.0253 sig
20.	Frequency of Passive VPs	.0000 sig
21.	Frequency of agentless passives	.0000 sig
22.	Frequency of Past Tense	.0000 sig
24.	Frequency of Modal Verbs	.0000 sig
25.	Frequency of 'MUST'	.0004 sig
26.	Frequency of 'MAY/MIGHT'	.0000 sig
27.	Frequency of 'CAN/COULD'	.0016 sig

Table 3.2. Kruskal-Wallis Test - Results significant at the 5% level (cont)

Variable	Description	Probability
28.	Frequency of 'WILL/WOULD'	.0000 sig
30.	Average Paragraph length	.0000 sig
31.	Reading Test Score	.0000 sig
32.	Average Line Length	.0000 sig
33.	Word Unfamiliarity Score	.0000 sig
34.	Text ties per 100 tokens	.0048 sig
35.	Frequency of Reference ties	.0040 sig
36.	Frequency of Lexis ties	.0008 sig
37.	Frequency of Ellipsis ties	.0000 sig
38.	Frequency of Substitution ties	.0000 sig
40.	Type Token ratio	.0007 sig

The results above indicate that the for these variables the scores for each variable can be considered as coming from at least 2 different populations at the 5% level. The linguistic implications are that these variables are those that are involved in making stylistic distinctions in the data set. At least one style must be distinct from one other to produce this result.

Table 3.2 Kruskal- Wallis Test - Results not significant at the 5% level

Variable	Description	Probability
6.	Frequency of Mixed Major Sentences	.2583 n.s.
16.	Frequency of Clausally Modified NPs	.5449 n.s.
23.	Frequency of Aspect marked VPs	.0909 n.s.
29.	Frequency of 'SHALL/SHOULD'	.5351 n.s.
39	Frequency of Conjunction ties	.8744 n.s.

The above results indicate that the scores for these variables have to be regarded as all being drawn from the same population of texts at the 5% level. Linguistically, this means that the variables

involved do not make stylistic contrasts in the data set.

In contemplating these results, it must be remembered that the list of variables was not chosen randomly but was non randomly selected, being based on the outcome of a qualitative study. This selection of variables was therefore very heavily biased in favour of finding variables associated with stylistic context. This observation does not apply to those variables chosen more to make up sets of items, e.g. modal 'SHALL' (variable 29), and Mixed Major Sentences (variable 6).

It is not therefore surprising to note that the null hypothesis was rejected for all but 5 of the 38 variables at the 5% level. 3 of these non-significant variables, variables 6,29 and 23 ( Mixed Major Sentences, modal 'SHALL' and Aspectual marking) were included in the dataset to make up sets, one (variable 39 Conjunction ties) had not been qualitatively examined. The only surprise was the failure of clausal postmodifiers to produce stylistic distinctions, as this had been highlighted as stylistic in Part 1.

On the basis of these results a tentative observation can be made. This is that the quantitative analysis carried out here supports the assumption that qualitative assessment is an equally valid method of locating linguistic stylistic features. No observations however can be made about how effective the variables studied are at distinguishing different styles. For this the results of the Pairs

tests, conducted after the K - W test, need to be examined.

### 3.2.1 Pairs Tests - General

The null hypothesis for the Pairs test is similar to that for the K -W test: This is that the samples which form groups A and B, (where A may be any group from 1 to 4 , and B any group not chosen to be A) have been drawn from the same population. Tables 3.2.2 to 3.2.7, below, show the variables for which the null hypothesis was rejected for all six possible pairs of groups. Each table indicates with a '\*' the group with the highest mean rank on each variable to clarify which group scored was the highest scoring.

### 3.2.2 Pairs Test - Group 1 (P.A.) with Group 2 (I.A.)

Table 3.2.2 Pairs Test - Group 1 (Personal Address) with Group 2 (Impersonal Address)

The null hypothesis was rejected at the 5% level for:

Var.	Description	Highest	
		Mean Rank	
		P.A.	I.A.
4.	Frequency of Simple Sentences	*	

Table 3.2.2 Pairs Test -Group 1 (Personal Address) with Group 2 (Impersonal Address) (cont)

The null hypothesis was rejected at the 5% level for:

Var.	Description	Highest	
		Mean Rank	
		P.A.	I.A.
9.	Frequency of Complement Clauses		*
10.	Frequency of Clauses with more than 2 adverbials		*
13.	Frequency of Pronominal NPs	*	
17.	Frequency of non-finite clausal NP postmodifiers		*
20.	Frequency of Passive VPs		*
21.	Frequency of Agentless Passives		*
30.	Average Paragraph length		*
31.	Reading Test Score		*
32.	Average line length		*

\*

These results indicate that for variables 4 and 13 Personal Address texts scored significantly higher than Impersonal Address texts at the 5% level; and that for variables 9,10,17,20,21,30,31 and 32, Impersonal Address texts scored significantly higher than Personal Address texts at the 5% level.

In linguistic terms these results indicate that:-

at sentence level: GIL Impersonal Address texts have significantly fewer Simple Sentences than do Personal Address texts.

at Clause level: Impersonal Address texts have the adverbial component of clause structure filled by 2 or more adverbials significantly more than do Personal address texts.

at Noun Phrase level: Personal Address texts are significantly more likely to have a phrase consisting only of a pronoun than Impersonal Address texts, and Noun Phrase Postmodifiers in Impersonal Address texts are significantly more likely to be non-finite clauses than in Personal Address texts.

at Verb Phrase level: Impersonal Address texts are significantly more likely to use passive verbs, than Personal Address texts. Passive verbs are significantly more likely to have no explicit agent in Impersonal Address texts than in Personal Address Texts.



for other variables: Lines and Paragraphs are significantly longer in I.A. than in P.A. texts

In stylistic terms these results show those linguistic variables which are associated with the situational differences between I.A. and P.A. These differences have already been established as those of status. A cautious interpretation of these variables would therefore be that they are stylistic variables of status in GIL.

None of these results are unexpected: In Part I of this thesis the qualitative observation was repeatedly made that the status differences between I.A. and P.A. appear to result in I.A. making fuller use of the structural possibilities of English at Sentence, Clause and Phrase level than does P.A.

A number of variables might have been expected to show significant results (on the basis of the Part I analysis), but did not. It might have been expected that I.A. texts would have had significantly more of: complex sentences (v.3), conditional and concessive adverbials (v.8), complex NP postmodification (v.15), embedded postmodifiers (v.18), unfamiliar words (v.33), word types (v.40) than D.A. texts. There are two possible reasons for this: Either the original analysis was erroneous (which seems unlikely), or that the stylistic differences observed were too small to show up as significant on the test. The latter explanation seems more likely since in each case the I.A. group had a higher mean rank than the D.A. group. This is a tendency which occurs repeatedly in the K-W test results.

It is possible that had an ANOVA been used instead of the K-W test then some of the results might have then been significant, since the ANOVA is more sensitive than the K-W test. Also if lower probabilities had been set again some results would have been significant. Even so, the use of the ANOVA would not strictly speaking have allowed inferences to be made about the parent populations, and the lowering of significance levels would have made errors far more likely.

### 3.2.3 Pairs Test - Group 1 (P.A.) with Group 3 (BBSIL)

Table 3.2.3 Pairs Test - Group 1 (Personal Address) with Group 3 (Bank and Building Society Information Leaflets

The null hypothesis was rejected at the 5% level for:

Var.	Description	Highest	
		Mean Rank	
		P.A.	BBSIL
4.	Frequency of Simple Sentences	*	
8.	Frequency of Conditional/Concessive Adverbials	.*	
11.	Frequency of Marked Clause Order	*	
12	Frequency of Time Adverbials	*	

18.	Frequency of NP Postmodif's containing embedding	*
21.	Frequency of Agentless Passives	*
26.	Frequency of 'MAY/MIGHT'	*
40.	Word types as a proportion of word tokens	*

These results indicate that for variables 4,8,11,12 & 18 Personal Address texts score significantly higher than BBSIL texts at the 5% level; and that for variables 21, 26 and 40, BBSIL texts score higher than Personal Address texts at the 5% level.

In linguistic terms these results indicate that:-

at sentence level: BBSIL texts have significantly fewer Simple sentences, than do P.A. texts. This can be compared with the similar result above for I.A. and P.A. texts, where I.A. texts function in the same way as BBSIL texts. In addition to this, adverbial clauses, when they occur, are conditional or concessive, significantly more often in P.A. texts than in BBSIL texts.

at clause level: P.A. texts use a non-standard ordering of clause elements significantly more often than do BBSIL texts. Where the adverbial position in a clause is filled, this is significantly more likely to be a time adverbial in P.A. than in BBSIL.

at noun phrase level: Postmodified NPs are significantly more likely to contain embedding in P.A. than in BBSIL texts.

at verb phrase level: Passive verbs are significantly more likely to lack an explicit agent in BBSIL than in P.A. texts. This result can be compared with those for I.A. and P.A. above, where I.A. texts behave in a similar way to BBSIL texts. In addition, a modal verb, where present, is significantly more likely to be 'MAY' or 'MIGHT' in BBSIL than in P.A.

for other variables: BBSIL texts are significantly more likely to have a higher number of different word types than do P.A. texts

In stylistic terms these results show those linguistic variables which are associated with the situational differences between P.A. and BBSIL texts. It has already been established that both these text types are in a leaflet Modality, and are Personal in Address. A cautious interpretation of these variables might therefore be that they are stylistic variables associated with an aspect of Province; that is whether the leaflets are from Central Government or from Banking.

The qualitative study in Part 1 of this thesis did not compare BBSIL texts with GIL texts, so there were no prior expectations about these results. The results do however partly support the intuitive expectation that variables distinguishing province will be different to those distinguishing status. The exceptions to this are the

frequency of Simple sentences, and the frequency of Agentless passives. If these variables are status variables as the comparison of I.A. and P.A. above seems to suggest, then it might be wise to tentatively accept that there are status differences within Personal address style, and that BBSIL is marginally more formal than P.A.

Of the remaining variables, variable 8 (Conditional/Concessive adverbials, variable 12 (Time adverbials) and 26 (Modal 'MAY'), all seem to suggest that texts from Government sources appear to be concerned very much with stressing conditions, times, and granting permission, or indicating what is allowed, than are texts from Financial Institutions. Variable 11, (Marked Clause order) may also relate to this tendency if it is true that initially placed adverbials are responsible for the high scores of P.A. texts: It might be expected that important conditions would be placed in an early position in the sentence in order to emphasise them, and this tendency was noted in the qualitative work for part one.

The exact stylistic significance of the remaining 2 variables is less easy to interpret. The following comments are therefore largely conjectural. The more impoverished vocabulary of P.A., and the richer vocabulary of BBSIL indicated by their respective scores on variable 40 may be related to Province. The subject of P.A. texts, is essentially that of informing the General Public about how Government legislation affects them. It must be understood by everybody, largely readers who have no background knowledge in the area of legislation concerned, vocabulary must therefore be simple

and repetitive.

The subject matter of BBSIL texts is that of banking and investment services. Each Financial institution has to compete in a free market for customers, many of whom may bank with rival companies. Customers therefore may already have some background knowledge in the area, and therefore a simple repetitive vocabulary is not necessary. On the contrary, a richer vocabulary may produce a knowledgeable sounding text more capable of convincing the customer of the superiority of a particular service.

Variable 18, (Frequency of NPs containing embedding) can be considered as a way of including as much qualifying information in a noun phrase as possible. It can only be interpreted as part of the general trend observed in Part 1, for GIL texts to have heavily postmodified NPs.

Since BBSIL texts were not examined in Part 1 there were no prior expectations about what variables might have been expected to show significant results. It is impossible therefore to comment in this case on variables which failed to yield significant results.

#### 3.2.4 Pairs Test - Group 1 (P.A.) with Group 4 (DMA)

Table 3.2.4 Pairs Test - Group 1 (Personal Address) with Group 4 Daily Mirror Articles.

The null hypothesis was rejected at the 5% level for:(see pp323-324)

Table 3.2.4 Pairs Test - Group 1 (Personal Address) with Group 4 Daily Mirror Articles. (cont)

The null hypothesis was rejected at the 5% level for:

Var.	Description	Highest	
		Mean Rank	
		P.A.	DMA
8.	Frequency of Conditional/Concessive Adverbial clauses	*	
11.	Frequency of Marked Clause Order	*	
12.	Frequency of Time Adverbials	*	
13.	Frequency of Pronominal NPs	*	
14.	Frequency of complex premodification		*
17.	Frequency of non-finite, clausal NP postmodifiers		*
18.	Frequency of NP postmodifiers containing embedding	*	
22.	Frequency of Past Tense		*

Table 3.2.4 Pairs Test - Group 1 (Personal Address) with Group 4 Daily Mirror Articles. (cont)

The null hypothesis was rejected at the 5% level for:

Var.	Description	Highest	
		Mean Rank	
		P.A.	DMA
24.	Frequency of Modal Verbs	*	
26.	Frequency of 'MAY/MIGHT'	*	
28.	Frequency of 'WILL/WOULD'		*
30.	Average Paragraph length	*	
32.	Average line length	*	
36.	Frequency of Lexis ties	*	
37.	Frequency of Ellipsis ties	*	
38.	Frequency of Substitution ties		*



These results indicate that for variables 8,11,12,13,18 24,26,30,32,36 and 37, Personal Address texts scored significantly higher than Daily Mirror texts at the 5% level; and that for variables 14,17,22,28 and 38, Daily Mirror texts scored significantly higher than Personal Address texts at the 5% level.

In linguistic terms these results indicate that:-

at sentence level: In P.A. texts, adverbial clauses are significantly more likely to be conditional/concessive than they are in DMA texts

at clause level: P.A clauses are significantly more likely to be non-standard in order than are clauses in DMA texts. In addition, clause adverbials are significantly more likely to relate to Time in P.A. than in DMA.

at noun phrase level: Premodification is significantly more likely to be complex in DMA than in P.A., and postmodification by clauses significantly more likely to be by non-finite. However, postmodifiers are significantly more likely to contain embedding in P.A. than in DMA.

at verb phrase level: Verb Phrases are significantly more likely to be marked for past tense in DMA than in P.A., whilst in P.A. Verb Phrases are significantly more likely to contain a modal than in

DMA. Modal Verbs are significantly more likely to be MAY/MIGHT in P.A. than in DMA, the opposite is true for WILL/WOULD.

at the level of sentence connection: Ties between sentences are significantly more likely to involve lexis and ellipsis in P.A. than in DMA. In DMA ties are significantly more likely to involve substitution than in P.A.

for other variables: Paragraphs and lines are likely to be significantly longer in P.A. texts than in DMA texts.

In stylistic terms these results show those linguistic variables which are associated with the situational differences between GIL Personal Address texts and Daily Mirror Articles. These differences involve several stylistic dimensions. Those which can be clearly identified are Province; DMA is journalism whereas GIL P.A. is public administration, and Modality; DMA are news reports whereas GIL P.A. are information leaflets. A cautious interpretation of these variables would therefore be that they are Province and Modality variables. Further results may help clarify to which of these two particular dimensions individual variables belong.

It is not possible to discuss whether these results were expected or not as no qualitative analysis was made of newspaper English. However, it might be expected that a number of the variables which distinguish P.A. from BBSIL might also occur here. This is because

both comparisons involve a comparison of government with non-government provinces. This expectation is to some extent borne out in the results. Both the present comparison of P.A. with DMA, and the previous comparison of P.A. with BBSIL share 8,11,12 and 18 as significant variables. In both cases P.A. has the highest scores. None of these variables show significant differences between the two GIL styles. These facts strengthen the possibility that these variables are strong stylistic markers of a general GIL style.

3.2.5 Pairs Test - Group 2 (Impersonal Address) with Group 3 (Banking and Building Society Leaflets).

Table 3.2.5 Pairs Test - Group 2 (Impersonal Address) with Group 3 (BBSIL)

The null hypothesis was rejected at the 5% level for:

Var.	Description	Highest	
		Mean Rank	
		I.A.	BBSIL
3.	Frequency of Sentences with Adverbial Clauses	*	
9.	Frequency of Complement Clauses	*	
10.	Frequency of Clauses with more than 2 adverbials	*	

	I.A.	BBSIL
11.   Frequency of Marked Clause Order	*	
15.   Frequency of Postmodified NPs	*	
17.   F'quency of non-finite, clausal postmodifiers	*	
18.   Frequency of NP postmodifiers containing embedd-   ing	*	
20.   Frequency of Passive VPs	*	
25.   Frequency of 'MUST'	*	
26.   Frequency of 'MAY/MIGHT'	*	
27.   Frequency of 'CAN/COULD'		*

Table 3.2.5 Pairs Test - Group 2 (Impersonal Address) with Group 3 (BBSIL) (cont)

The null hypothesis was rejected at the 5% level for:

Var.	Description	Highest	
		Mean Rank	
		I.A.	BBSIL
28.	Frequency of 'WILL/WOULD'		*
30.	Average Paragraph length	*	
31.	Reading Test Score	*	
33.	Word Unfamiliarity Score	*	

These results indicate that for variables 3,9,10,11,15,17,18,20,25, 26,30,31, and 33, Impersonal Address texts scored significantly higher than BBSIL texts at the 5% level; and that for variables 27 and 28, BBSIL texts scored significantly higher than Impersonal Address texts at the 5% level.

In linguistic terms these results indicate that:-

at sentence level: I.A. sentences are significantly more likely to be complex and to have a complement which is realised by a clause than BBSIL sentences.

at clause level: I.A. clauses are significantly more likely to have a non-standard order, and to have any adverbial component consist of more than 2 adverbial groups, than do BBSIL clauses.

at noun phrase level: I.A. noun phrases are significantly more likely to be postmodified, and these postmodifiers are significantly more likely to contain embedding than are BBSIL noun phrases. Also, where noun phrase postmodifiers are clauses, these are significantly more likely to be non-finite in I.A. than in BBSIL.

at verb phrase level: I.A. verb phrases are significantly more likely to be passive than those of BBSIL. Where verb phrases contain modals, these are significantly more likely to be MUST, or MAY/MIGHT in I.A. than in BBSIL. The opposite is true of CAN/COULD, and WILL/WOULD.

for other variables: Paragraphs are significantly longer in I.A. than in BBSIL, Reading test scores, and word unfamiliarity scores are also significantly higher.

In stylistic terms these results show those linguistic variables which are associated with the situational differences between I.A.

and BBSIL. These differences have been already established as those of Province, and possibly Status (BBSIL like P.A. is personally addressed). A cautious interpretation of these results would be that they are stylistic variables of Province and Status.

There were no prior expectations about these results, as BBSIL was not included in the original qualitative study. However it might be expected, that BBSIL would differ significantly along with I.A. on in the same way, on similar variables. This expectation is borne out for variables 9,10,17,20,30 and 31; frequency of complement clauses, frequency of clauses with more than 2 adverbials, frequency of non-finite, clausal postmodifiers, frequency of passive VPs, and reading test scores.

3.2.6. Pairs Test - Group 2 (I.A.) with Group 4 (DMA)

Table 3.2.6 Pairs Test - Group 2 (Impersonal Address) with Group 4 Daily Mirror Articles.

The null hypothesis was rejected at the 5% level for:

Var.	Description	Highest	
		Mean Rank	
		I.A.	DMA
3.	Frequency of sentences with adverbial clauses	*	
4.	Frequency of simple sentences		*
7.	Frequency of Minor sentences	*	
8.	Frequency of Conditional/Concessive clauses	*	
9.	Frequency of Complement clauses	*	
10.	Frequency of clauses with more than 2 adverbials	*	
11.	Frequency of marked clause order	*	
12.	Frequency of time adverbials	*	



Table 3.2.4 Pairs Test - Group 2 (Impersonal Address) with Group 4 Daily Mirror Articles. (cont)

The null hypothesis was rejected at the 5% level for:

Var.	Description	Highest	
		Mean Rank	
		I.A.	DMA
14.	Frequency of complex premodification		*
15.	Frequency of postmodified NPs	*	
18.	Frequency of NP postmodifiers containing embedding	*	
19.	Frequency of NP quantifiers	*	
20.	Frequency of Passive VPs	*	
21.	Frequency of agentless passives	*	
22.	Frequency of past tense		*
24.	Frequency of modal verbs	*	

Table 3.2.6 Pairs Test - Group 2 (Impersonal Address) with Group 4 Daily Mirror Articles. (cont)

The null hypothesis was rejected at the 5% level for:

Var.	Description	Highest	
		Mean Rank	
		I.A.	DMA
26.	Frequency of 'MAY/MIGHT'	*	
28	Frequency of 'WILL/WOULD'		*
30	Average Paragraph length	*	
31	Reading tests score	*	
32	Average line length	*	
33	Word unfamiliarity score	*	
35	Frequency of Reference ties		*
36	Frequency of Lexis ties	*	
37	Frequency of ellipsis ties	*	

Table 3.2.6 Pairs Test - Group 2 (Impersonal Address) with Group 4 Daily Mirror Articles. (cont)

The null hypothesis was rejected at the 5% level for:

Var.	Description	Highest	
		Mean Rank	
		I.A.	DMA
38	Frequency of substitution ties		*

These results indicate that for variables 3,7,8,9,10,11,12,15,18,19, 20,21,24,26,30,31,32,33,36 and 37, Impersonal Address texts scored significantly higher than DMA texts at the 5% level; and that for variables 4,14,22,28,35 and 38, DMA texts scored significantly higher than Impersonal Address texts at the 5% level.

In linguistic terms these results indicate that:-

at sentence level: I.A. sentences are significantly more likely to be complex than in DMA. The reverse is true of simple sentences. The Complement and Adverbial components of I.A. sentences are significantly more likely to be clauses than in DMA sentences. Minor sentences are significantly more frequent in I.A. than in DMA.

at clause level: I.A. clauses are significantly more likely to have a non-standard order than DMA clauses. The adverbial component of I.A. clauses is significantly more likely to be composed of 2 adverbial groups, and to relate to time, than is the adverbial component of a DMA clause.

at noun phrase level: I.A. noun phrases are significantly more likely to be postmodified than DMA noun phrases. The reverse is true of premodification. Where a noun phrase is postmodified, this contains embedding significantly more often in I.A. than in DMA texts. Noun phrase quantifiers occur significantly more frequently in I.A. than in DMA.

at verb phrase level: verb phrases in I.A. are significantly more likely to contain modals and passives than are VPs in DMA. Passives are significantly more likely to lack an agent in I.A. than in DMA texts. Modal verbs are significantly more likely to be MAY/MIGHT in I.A. and significantly more likely to be WILL/WOULD in DMA. Verb phrases are significantly more likely to be marked for past tense in DMA than in I.A. texts.

at the level of inter sentence relationships: sentence ties are significantly more likely to involve 'reference' and 'substitution' in DMA texts than in I.A. texts. The reverse is true of 'lexis' and 'ellipsis' ties.

for the remaining variables: Paragraphs and lines are significantly longer in I.A. than in DMA texts. Reading test and Word Unfamiliarity scores are significantly higher in I.A. than in DMA texts.

In stylistic terms these results show variables which are associated with the situational differences between I.A. and DMA. Those situational dimensions which appear to be involved have already been highlighted as Province, Status and Modality. A cautious interpretation of the variables in these results is that they are stylistic variables of Province, Status, and Modality.

DMA was not included in the initial qualitative study therefore there were no expectations as to what results the pairs test for I.A. and DMA would produce. However, because of the relatively large number of stylistic dimensions involved in the contrast between these two styles it might be expected that they would be differentiated by a large number of variables. This is true; 25 out of 32 possible variables distinguish I.A. and DMA.

3.2.7. Pairs Test - Group 3 (BBSIL) with Group 4 (DMA)

Table 3.2.7 Pairs Test - Group 2 (Banking and Building Society Leaflets) with Group 4 Daily Mirror Articles.

The null hypothesis was rejected at the 5% level for:

Var.	Description	Highest	
		Mean Rank	
		BBSIL	DMA
5.	Frequency of compound sentences		*
7.	Frequency of minor sentences	*	
8.	Frequency of conditional/concessive clauses	*	
17.	Frequency of non-finite, clausal NP postmodifiers		*
21	Frequency of Agentless passives	*	
22.	Frequency of past tense		*
24.	Frequency of Modal verbs	*	
30	Average paragraph length	*	

Table 3.2.7 Pairs Test - Group 2 (Banking and Building Society Leaflets) with Group 4 Daily Mirror Articles. (cont)

The null hypothesis was rejected at the 5% level for:

Var.	Description	Highest	
		Mean Rank	
		BBSIL	DMA
32	Average line length	*	
35.	Frequency of Reference ties		*
36.	Frequency of Lexis ties	*	
37.	Frequency of Ellipsis ties	*	
38	Frequency of Substitution ties		*

These results indicate that for variables 5,17,22,35, and 38, DMA texts scored significantly higher than BBSIL texts at the 5% level; and that for variables 7,8,21,24,30,32,36, and 37, BBSIL texts scored significantly higher than DMA texts at the 5% level.

In linguistic terms these results indicate that:-

at sentence level: DMA sentences are compound significantly more often than BBSIL sentences. The reverse is true of minor sentences. When sentences contain an adverbial clause this is significantly more likely to be conditional/concessive in BBSIL than in DMA.

at clause level: none of the clause structure variables showed significant results at the 5% level.

at noun phrase level: where NP postmodifiers are clauses, these are non-finite significantly more often in DMA than in BBSIL.

at verb phrase level: BBSIL verb phrases contain modals significantly more often than DMA verb phrases. The reverse is true for past tense marking. Where VPs contain passives, these are agentless significantly more often in BBSIL than in DMA texts.

at the level of inter sentence relationships: BBSIL sentence ties involve Lexis and Ellipsis significantly more often than DMA sentence ties. The reverse is true of ties involving Reference and Substitution.

for other variables: BBSIL paragraphs and lines are significantly longer than in DMA texts.



In stylistic terms, these results show those linguistic variables which are associated with the situational differences between BBSIL and DMA. The two situational dimensions which have been established as involved in forming these situational differences are Modality, and Province. A cautious interpretation of these variables might therefore be that they are stylistic variables of Modality and Province.

Neither DMA nor BBSIL were included in the initial qualitative study, and therefore there were no expectations about what results would emerge from their comparison. However it might be expected that the three Leaflet groups I.A., P.A. and BBSIL, collectively sharing the same Modality type, might differ consistently from DMA, in terms of the variables which distinguish them all from DMA. The comparison of all 3 groups with DMA shows that they all share variables 7,8,22,24,30,32,36,37,and 38. This evidence gives a good deal of support to the stylistic notion of an information leaflet modality, the presence of all these variable together as a set possibly being an indicator of this style.

Of distinct interest too is the fact that the P.A./D.M.A. differences (see 3.2.4 above) are almost a subset of the I.A./D.M.A. differences. The variables of this subset are: 8,11,12,14,18,22,24, 26,28,30,32,26,38, and show I.A. and P.A. to jointly have more:conditional/concessive adverbial clauses, initial adverbials, time adverbials, embedded postmodifiers, modal verbs,modal 'May',

longer paragraphs, longer lines, lexical ties, and substitution ties than D.M.A. I.A. and P.A. have jointly significantly less: complex NP premodification, past tense marking, modal 'Will', and substitution ties. This set of variables occurring together in a text therefore seem indicative of a Government Information Leaflet Modality. The alternative view is that they are strong markers of the Daily Mirror's journalistic style or of a journalistic style in general. Were this the case though one would not expect so many variables to be shared by I.A. and P.A. texts. Further research in this area seems likely to be an interesting topic and may help to shed more light on the nature of the relationships between these text types.

### 3.3 Spearman Correlation Co-efficients.

The Spearman measure of Correlation was not used to test a specific hypothesis, but to investigate the correlational structure among all the variables studied. All possible pairs of variables were compared using this test, including those with non-significant results on the Kruskal-Wallis test, but excluding variable 40 which does not have a full set of scores (values missing for group 4).

The resulting multiple correlation matrix is not displayed. It is difficult to interpret and present because of its huge size (some 2' 6" sq), and the fact that given the number of pairs in the comparison, nearly every result is 'significant' at the 5% level. As planned, the correlation matrix was processed 'by hand' to try and

discover groups of strongly mutually correlating variables, after the manner of factor analysis.

Attention was first paid to all those correlation co-efficients of .33 or larger. This figure was an intuitive and arbitrary choice, and this must be remembered when considering the results. The highest correlation was .66. These selected correlations were examined to find groups of mutually correlating variables. Complete mutual correlations were considered most desirable, but correlations for any variable which were lower than .33 with no more than one third of the variables in the rest of the group were tolerated. This condition was again an arbitrary choice, to reduce the groups produced to a manageable and interpretable level. Groups consisting of less than 5 variables were discarded.

The result was one group of 6 completely mutually correlating variables (see Fig. 3.3.1 below), and a second larger group of 18 partly mutually correlating variables (see Fig 3.3.2 below).

Fig. 3.3.1 1st Group of mutual correlations (GROUP 1)

Brief variable description

Embedded post modifiers	18		-									' - ' indicates a negative correlation
Passive VPs	20		.53	-								
Frequency of 'WILL/WOULD'	28		-.58	-.41	-							
Paragraph length	30		.64	.46	-.46	-						
Reading test score	31		.45	.51	-.36	.49	-					
Word Unfamiliarity	33		.36	.36	.33	.41	.33	-				
			-----									
			18	20	28	30	31	33				

Fig 3.3.2 2nd Group of mutual correlations (GROUP 2)

( ) indicate correlation below .33  
 ' - ' indicates a negative correlation  
 no sign indicates a positive correlation  
 key to variables on next page

7	-																			
8	.41	-																		
14	-.36	-.39																		
18	.35	.61	-.37	-																
22	-.55	-.50	.34	-.34	-															
24	.40	.61	-.46	.43	-.46	-														
28	-.43	(-.31)	.36	-.58	.48	(-.30)	-													
30	.39	.48	(-.23)	.64	-.52	.59	-.46	-												
32	.43	.51	(-.29)	.41	-.50	.48	-.42	.66	-											
37	.46	.57	-.45	.35	-.46	(.28)	(-.21)	(.32)	.34	-										
38	-.57	-.33	(.16)	(-.23)	.43	-.42	(.19)	-.46	-.40	-.40	-									

---

7	8	14	18	22	24	28	30	32	37	38
---	---	----	----	----	----	----	----	----	----	----

Key to variables for 2nd group of correlations

- 7 - Frequency of Minor Sentences
- 8 - Frequency of Complement Clauses
- 14- Frequency of Complex NP Modification
- 18- Frequency of NP postmodifiers with embedding
- 22- Frequency of Past Tense
- 24- Frequency of Modal Verbs
- 28- Frequency of 'WILL/WOULD'
- 30- Average Paragraph length
- 32- Average Line Length
- 37--Cohesive ties involving ellipsis
- 38- Cohesive ties involving substitution

Notes on Figs 3.3.1 and 3.3.2

A correlation between two variables can mean either one of two things. Either one variable causes the other, or both variables share a common factor. Certain Linguistic relationships can be expected to correlate in the former way, the relationship between the proportion of Simple and non-simple sentences for example, where a rise in one proportion will result in a decrease in the other. This type of relationship is of little interest to the stylistician who wishes to search for common situational factors affecting variables. However, apart from the different sentence type, variables (variables 3 to 7) none of the other variables in this study are

expected to show a direct causal relationship.

The fact that at least two groups of mutually correlating variables appear to be present in the data appears to fulfill the expectations of this study, that linguistic variables are not all measuring different things but group together to produce intuitively perceivable stylistic effects.

For both Group 1 and Group 2, the results appear to indicate that the variables in the matrix are to some extent conditioned by a common factor. The influence of this factor is only partial because the correlations between variables are not perfect (+ 1 or -1 indicates a perfect correlation). In stylistic terms, the results would appear to indicate that the scores on variables in the 2 groups are being influenced each by at least one stylistic dimension, 'Status' for example. The possible identity of these dimensions needs careful consideration and this is done below.

The Correlations in Group 1 (Fig 3.3.1) are composed of some positive and some negative correlations. It is variable 28 (Frequency of Will/Would), that attracts the negative correlations. The variables of Group 1 seem to indicate then a dimension on which texts tend to have:-

frequent embedding

frequent use of passive verbs

long paragraphs  
high reading test scores (indicating possible reading difficulties  
for some readers)  
unfamiliar vocabulary.

They also tend not to use the modal forms 'WILL' and 'WOULD'.

The non-use of WILL and WOULD are difficult to interpret but may possibly indicate a tendency not to refer to future time, as tends to be true of Government Information Leaflets. More interesting though are the positively correlating variables which seem to indicate a dimension orientated towards the contrast between formal and informal style. A tentative interpretation of the variables in Group 1 might therefore be that they are Status variables.

The correlations of Group 2 are again a mixture of positive and negatively correlating variables. Variable 28 again consistently attracts negative correlations as do variables 14 (Frequency of complex premodification), 22 (Frequency of Past tense), and 38 (Frequency of substitution sentence ties). The variables of Group 2 then appear to indicate a dimension on which texts tend to have:-

frequent use of minor sentences (possibly due to frequent use of  
side headings)

frequent use of conditional/concessive clauses

frequent embedding in NP postmodifiers



frequent use of modals  
long paragraphs  
long lines  
and frequent use of ellipsis ties

and tend not to use

complex premodification

past tense

modals: 'WILL/WOULD'

and substitution ties.

These correlations are very difficult to interpret. The positive correlations seem to indicate text which is more concerned with static informational content, and which is perhaps technical. The negative correlations seem to indicate a more descriptive text that makes fuller use of the time/tense system, opportunities for adjectival premodification and the potentialities for variety introduced by using substitution as a cohesive device. This group of variables seems to suggest a contrast between Information leaflets and Journalism, and might therefore be a complex of the stylistic dimensions of both Modality and Province.

### 3.4 Factor Analysis

A Factor Analysis was conducted on the data (excluding variable 40) for the same reason as the Spearman Correlation Test above. the difference this time was that the further analysis of a correlation matrix (this time using Pearson's correlations) was done by a statistical technique rather than by hand.

The first output of this technique was a list of factors, the percentage of variance in the data (accounted for by each factor) and the eigenvalues of the factors. A table of these initial statistics, table 3.4.1, can be found on the next page. The eigenvalues were plotted on a 'scree plot', for which see Fig 3.4.2 (on page 351). Visual observation of the scree plot and Table 3.4.1 reveals that the first 2 factors have relatively large values and account for 21.8 and 11.4% of the shared variance, respectively. This still of course leaves about two thirds of the remaining variance, quite a lot. The remaining factors all account for only small amounts of the shared variance. Some workers do use a rule of thumb that any factor with an Eigenvalue above 1 is worth keeping. This practice has not been followed here. It is deemed that very little would be gained by examining these factors, on the contrary a great deal of conflicting and misleading material might result.

These first two Factors extracted were then subjected to a Varimax rotation in order to ease interpretation. The final rotated solution

TABLE 3.4.1

FACTOR ANALYSIS

INITIAL STATISTICS:

FACTOR	EIGENVALUE	PCT OF VAR	CUM PCT
1	8.07641	21.8	21.8
2	4.22980	11.4	33.3
3	2.24287	6.1	39.3
4	1.85220	5.0	44.3
5	1.84260	5.0	49.3
6	1.70841	4.6	53.9
7	1.62487	4.4	58.3
8	1.39647	3.8	62.1
9	1.33109	3.6	65.7
10	1.25745	3.4	69.1
11	1.07714	2.9	72.0
12	1.01672	2.7	74.7
13	.92162	2.5	77.2
14	.80837	2.2	79.4
15	.72905	2.0	81.4
16	.68677	1.9	83.2
17	.65078	1.8	85.0
18	.63725	1.7	86.7
19	.54192	1.5	88.2
20	.51611	1.4	89.6
21	.46908	1.3	90.9
22	.45831	1.2	92.1
23	.41453	1.1	93.2
24	.35479	1.0	94.2
25	.29735	.8	95.0
26	.26291	.7	95.7
27	.25043	.7	96.4
28	.23254	.6	97.0
29	.20841	.6	97.6
30	.19168	.5	98.1
31	.15803	.4	98.5
32	.14610	.4	98.9
33	.12767	.3	99.2
34	.10760	.3	99.5
35	.07325	.2	99.7
36	.06058	.2	99.9
37	.04083	.1	100.0

FIG. 3.4.2.: 'SCREE' PLOT

FACTOR ANAL

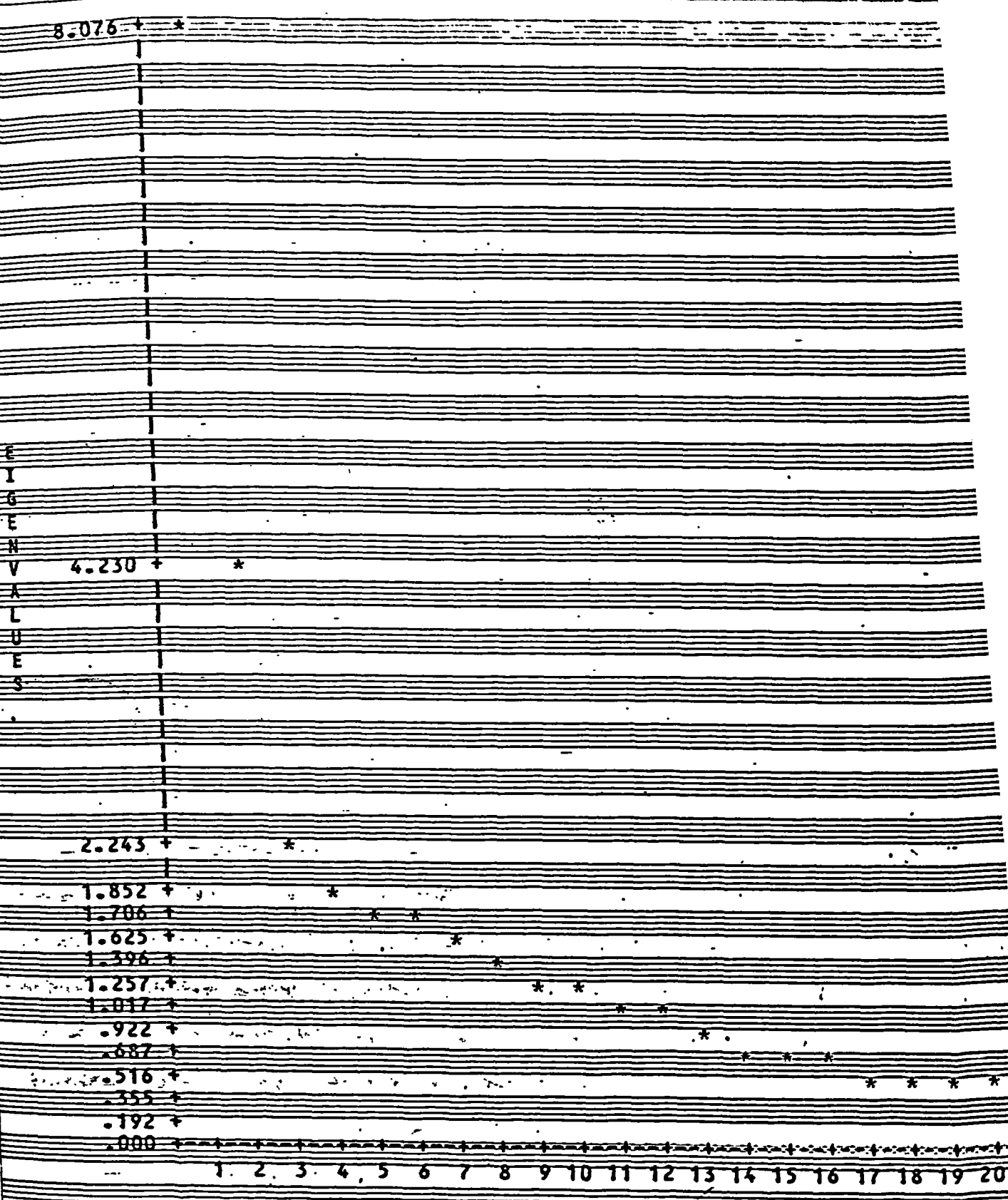


FIG 3.4.2 - 'SCREE' PLOT (CONT)

FACTOR ANALYSIS

1 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 3

TABLE-3.4.3:

F A C T O R   A N A L Y S I S

ROTATED FACTOR MATRIX:		
	FACTOR 1	FACTOR 2
C3	.25724	.30556
C4	-0.13578	-0.41582
C5	-0.51939*	.16062
C6	.05758	.26057
C7	.53281*	-0.04067
C8	.83243*	.02146
C9	.05708	.57925*
C10	.16980	.37098
C11	.39472	.45066*
C12	.47758	-0.09236
C13	.48253	-0.49547
C14	-0.51215*	.00756
C15	.14684	.36520
C16	-0.03718	.12896
C17	-0.20932	.57107*
C18	.44373	.65044*
C19	.25317	.21354
C20	.26396	.76286*
C21	.33471	.40575
C22	-0.75363*	-0.06727
C23	-0.23470	-0.04211
C24	.68911*	.14710
C25	.25740	.29329
C26	.36697	.29582
C27	.16074	-0.42007
C28	-0.53673*	-0.31147
C29	-0.04587	.32275
C30	(.55706*)	.60802*
C31	.24929	.72482*
C32	.57152*	.35782
C33	.11542	.69963*
C34	.10430	-0.39803
C35	-0.28894	-0.15971
C36	.38734	.04788
C37	.51145*	-0.03512
C38	-0.47725	-0.09635
C39	.11726	-0.14686

is presented in table 3.4.3. (see page 352). This table lists each variable in the Factor Analysis and its loading on (correlation with) each factor extracted.

The results of the factor analyses suggest a partial structuring of the variables into two main factors, though there is a very large amount of variance (65.7%) which is unexplained by any large factor structures. This means that to a fairly large extent, the variables included in the study are all measuring rather different things and are only partially explainable in terms of a factor structure.

In Linguistic terms, the two main factors extracted can be seen as dimensions affecting the use of a whole group of structural items. In Stylistic terms the factors may be synonymous with stylistic dimensions like Status, Modality etc. A preliminary interpretation of the factors in stylistic terms is done below.

Because of the exploratory nature of the study, and the non-normal distributions underlying many of the variables included in the Factor Analysis, no decisions were made as to which factor loadings were significant and which were not. Instead, interest was paid to only the higher factor loadings, those marked with an asterisk in table 3.4.3. 'Higher' was defined arbitrarily as being above .5. This arbitrary decision must be taken account of when interpreting the higher factor loadings.

### Composition of the Two Factors

POSITIVE loadings above .5 are obtained on the following variables for FACTOR 1

7. Frequency of minor sentences
8. Frequency of conditional/concessive adverbial clauses
24. Frequency of Modal Verbs
30. Average paragraph length (This variable has a higher loading on Factor 2)
32. Average line length
37. Frequency of ellipsis ties

and NEGATIVE loadings on the following variables

5. Frequency of compound sentences
14. Frequency of complex premodification
22. Frequency of past tense
28. Frequency of 'WILL/WOULD'

This first factor shares a large number of variables with Group 2 of the manual analysis of correlations above. The only variable not encountered in Group 2 but encountered here is variable 5. Variables 18 and 38 are encountered in Group 2 but not here, though their loadings; .44373 and -.47725 are very close to the cut off point used here of .5. All the variables attracting negative correlations



in Group 2 have negative loadings on Factor 1, the same is true for the positive loadings. In view of the strong agreement between this factor and Group 2, it seems appropriate to place the same tentative interpretation on this factor as on Group 2, namely that of the contrast between static explanatory text and more varied descriptive text. It will therefore be referred to as the Explanatory v Descriptive text factor.

POSITIVE loadings above .5 are obtained on the following variables for FACTOR 2:-

9. Frequency of complement clauses
17. Frequency of non-finite, clausal post-modifiers.
18. Frequency of NP postmodifiers containing embedding
20. Frequency of passive verbs
30. Average paragraph length
31. Reading test score
33. Word Unfamiliarity score.

All these variables behaved similarly in the Kruskal-Wallis results where they showed in most cases I.A. texts to be significantly different from all the other groups. The two exceptions were variable 17 which failed to show a significant difference with D.M.A., and variable 18 which failed to show a significant difference with P.A.

There are no negative loadings above .5 on Factor 2. As was the case

with Group 2 and Factor 1, the agreement between Factor 2 and Group 1 is very close. The only variables encountered in Group 1 and not encountered in Factor 2 is variable 28. The only variables encountered in Factor 2 and not in Group 1 are variables 9 and 17. Again positive loadings pair with the attraction of positive correlations. The same tentative interpretation therefore seems appropriate to Factor 2 as to Group 1; that it is related to formality contrasts. It will be referred to as the Formal v Informal Text factor.

### 3.5 Manual Analysis of Rank Order Based Distance Co-efficients

Having now considered the results of tests which try to locate patterns of similar variables (Manual correlation based analysis and Factor Analysis), the results of those which search for patterns of similar texts will be presented (Manual Analysis of Distance co-efficients and Cluster Analysis). The two types of test in each of these pairs can be compared in similar ways.

The initial stage in the manual analysis of distance co-efficients was to compute the distance Co-efficients themselves. To do this a Basic program written by Mr P J Scholfield of UCNW was used. This programme first ranked the score for each text over on each variable. The use of ranks (rather than raw scores) was a 'conservative' procedure, given doubt as to the 'interval' nature of the scores. The Euclidean distance co-efficient (the usual one used)

was computed based on the distance between the ranks of individual variables for each text, adjusted for tied ranks. These distance co-efficients, one for every possible pair of texts, was printed in order of increasing distance as output to the program.

The formula for calculating the distance co-efficient is:

$$D = \sqrt{\sum_{ab} (x_a - x_b)^2}$$

i.e. for each pair of texts a & b take variable 1, subtract the (rank) scores, square the figure, then do the same for variable 2 and add the squared result. When all the 'squared differences' have been added, take the square root. This is the distance coefficient for text a & b.

The number of pairs of texts compared by the program was 3,160. There are therefore far too many distance co-efficients to present either here or in any appendix. Because the distance coefficients are printed in order of increasing distance texts which are very similar will appear at the top of the list, and texts which are very dissimilar at the bottom of the list. The central portion of the list will be a transitional area between these two extremes.

Because of these factors, the favoured outcome for a manual analysis of the co-efficients has to be one where pairs in the topmost

section of the list, where pairs of texts are very similar have members which both belong to the same group, e.g. Impersonal Address. If a particular text is frequently paired with members of groups other than its own, this is an indication that the text has possibly been misclassified by the researcher. The pairs in the bottom section of the list, where pairs of texts are very dissimilar should ideally contain texts from differing groups. If a text is frequently paired with its fellow group members at this point in the list a misclassified text is again indicated.

The main problem with this analysis was what to consider as the top and bottom sections of the list. It was decided to look at only a quarter of the distance coefficient output, that is the top and bottom eighth of the list. This involved examining the first and last 395 pairs of texts. The results were as follows:-

(see following page)

Table 3.5.1 Misclassified Texts indicated by the top 395 Rank Order  
Distance coefficients (texts ordered from most  
to least misclassified)

Text No.	Times misclassified	Member of:	Most often misclassified as:
48	7	BBSIL	DMA
5	5	P.A.	BBSIL
13	5	P.A.	BBSIL
6	4	P.A.	BBSIL
55	4	BBSIL	P.A.
42	3	BBSIL	P.A.
1	2	P.A.	BBSIL
14	2	P.A.	BBSIL/I.A.
17	2	P.A.	BBSIL
18	2	P.A.	BBSIL

Table 3.5.1 Misclassified Texts indicated by the top 395 Rank Order Distance coefficients (texts ordered from most to least misclassified) (CONT)

Text No.	Times misclassified	Member of:	Most often misclassified as:
53	2	BBSIL	P.A.
56	2	BBSIL	P.A.
57	2	BBSIL	P.A.

Note: Texts misclassified only once where considered trivial cases and not included.

The only text which appeared to be unsatisfactorily classified when the bottom 395 distance co-efficients, was text 2 which had a high distance score with members of its own (P.A.) group twice.

The major outcome of this analysis is that BBSIL texts are frequently misclassified as P.A. and vice-versa. A further point of note is that there is one BBSIL text very frequently misclassified as DMA.

In Linguistic terms, the results show that Personal address GIL texts and BBSIL texts have very similar distributions of the linguistic items studied on the variables tested in this research. In stylistic terms, the results appear to indicate that P.A. and BBSIL are close stylistic neighbours, though there is not enough evidence to suggest that they are one and the same style. This result is not unexpected since both P.A. and BBSIL are written in a leaflet Modality, and are relatively informal in addressing the reader as 'you' rather than 'the investor', 'the applicant' etc.

The very badly misclassified text 48 as DMA, suggests a leaflet whose style is nearer that of Journalism than leaflets. This misclassification can be explained on further perusal of the leaflet (see Appendix D). The leaflet concerned is unusual in that it concerns a layman's introduction to how share trading takes place, by Cliff Michelmore, with injunctions to the potential investor to buy shares in a unit trust. The major part of the leaflet is in the form of a descriptive article rather than seeking to provide direct information about banking services. The text therefore seems also intuitively journalistic as well as quantitatively so.

### 3.6 Cluster Analysis

Cluster Analysis was applied to the data for all the individual texts, across every variable but 40, which has missing values. The overall purpose of cluster analysis is similar to the manual

analysis of distance co-efficients carried out above. Cluster analysis works on a matrix of distance or similarity co-efficients, searching for groups of similar cases (in the present case these are texts). The Introduction to Part 2 (sect 1.9.1) revealed that there are a variety of clustering techniques available, all of which may yield slightly different results with the same data.

The two clustering techniques used in the present study were both Hierarchical Clustering Techniques: Single Linkage, and Ward's Method. Using Euclidean distance matrices these techniques were chosen as those being best understood by the researcher, relatively quick to run, and available to the author on the CLUSTAN computing package.

Both clustering methods start by finding the two texts with the smallest distance co-efficient and fusing them into a cluster. This cluster is then considered as a single unit. The program now looks for the next most similar pair, but considers both the individual texts and the new cluster. Texts and Clusters are paired together until all the texts have been joined in one large cluster. The program then finishes.

Single Linkage and Ward's method differ in the way in which they fuse texts and clusters. Single linkage simply fuses two clusters if they each have individual texts which are close neighbours. The other texts in the cluster are not considered. Single Linkage is therefore apt to find long thin clusters, and to 'chain' clusters



together which have only superficial similarities. Ward's Method takes all the individuals in clusters into account when fusing.

Therefore any new cluster member has to be compared with all members of the existing cluster. Not surprisingly Ward's method tends to find more spherical shaped clusters. Everitt (1980) notes that Single Linkage Clustering is often favoured by some researchers because it has the most desirable mathematical basis.

The output of both these programs is large, and is not included in any part of this thesis. It consists of the co-efficient values at which pairs of texts, or pairs of texts and existing clusters were fused together to form new clusters. The most useful option with Single Linkage and Ward's Method on the Clustan Package is a visual representation of the hierarchical clusters formed in the form of a dendrogram. The dendrograms for the Single Linkage and Ward's Method Clustering solutions are presented on pages 363 and 364 (following).

The dendrograms have the distance co-efficients plotted up the y axis and the individual text numbers across the x axis. A vertical line is drawn vertically above each text or cluster until it is fused with another text or cluster. The fusion of the two is shown by a horizontal line. Because the distance co-efficients are plotted up the vertical axis, the vertical lines on the dendrogram give a visual representation of how close to each other clusters are.

1.814

1.656

1.499

1.342

1.185

363

1.028

0.871

0.714

0.557

0.399

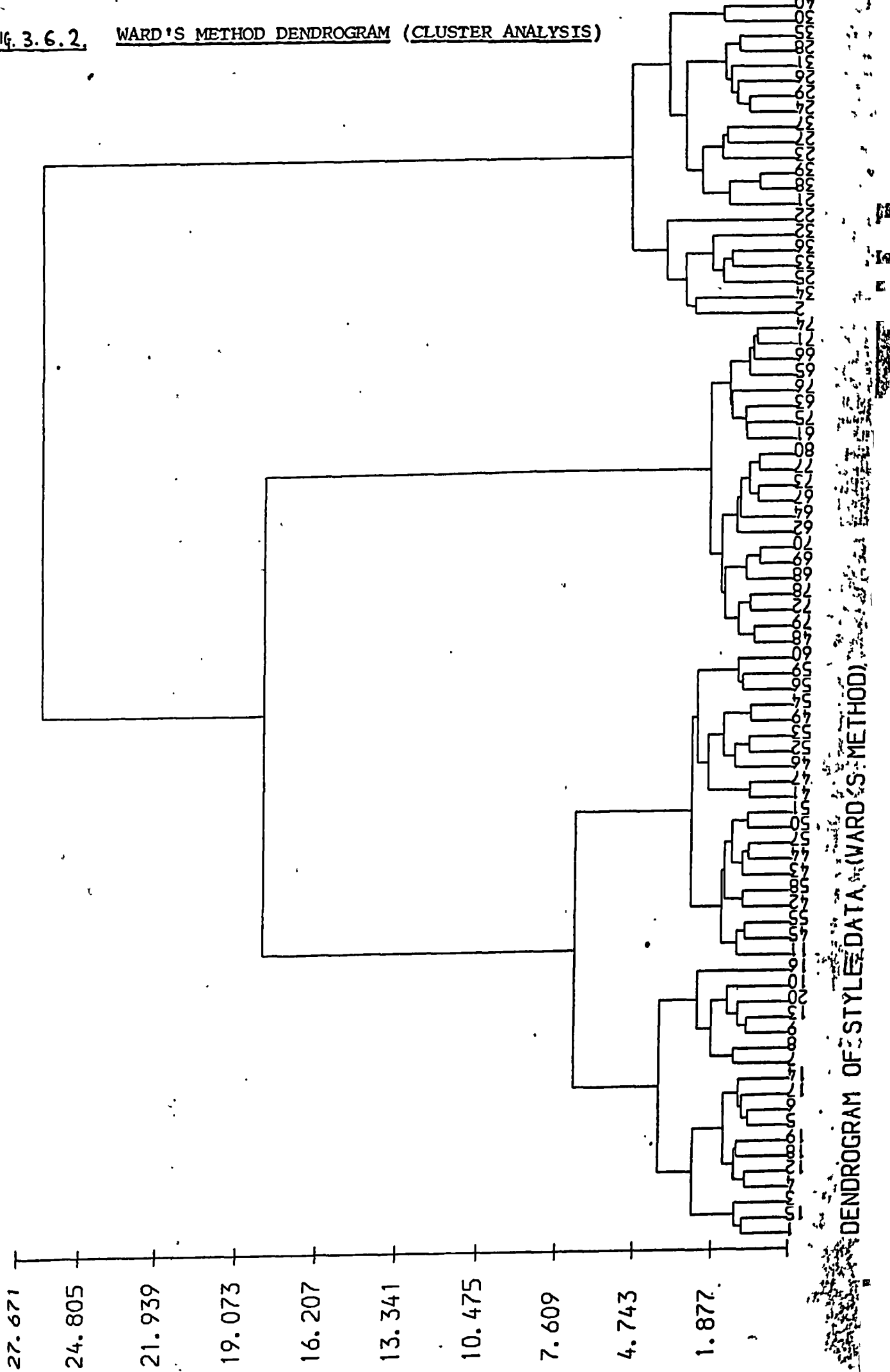
Fig. 3.6.1 SINGLE LINKAGE DENDROGRAM (CLUSTER ANALYSIS)



DENDROGRAM OF STYLE DATA (SINGLE LINK)

FIG. 3.6.2.

WARD'S METHOD DENDROGRAM (CLUSTER ANALYSIS)



### 3.6.1 Initial Interpretation of Single Linkage Dendrogram (fig. 3.6.1.)

The Single Linkage Dendrogram is uninterpretable; no vertical lines of any appreciable length separate clusters, and the tendency has been for the program to add texts one at a time to a big cluster. If anything, this dendrogram seems to suggest that there are virtually no patterns among the texts. This does not agree at all with the manual analysis of distance co-efficients whose results are reported in 3.5 above. The evidence of section 3.5 suggests that I.A. and DMA are distinct homogeneous categories of texts, and that P.A. and BBSIL are distinct but very close neighbours. The Single Linkage Cluster analysis therefore is to be rejected as a true analysis of the data. The chaining tendency of this technique appears to have constructed a long thin cluster which cannot be interpreted as such. What it does suggest however is that any clusters present are not very separate, but perhaps overlapping.

### 3.6.2 Initial Interpretation of Ward's Method Dendrogram (fig. 3.6.2.)

Examination of the Ward's Method dendrogram shows that the data is composed of four distinct homogeneous groups. The two groups on the left (from text 1 to 16, and from 11 to 60 along the bottom) are the closest related. Together these groups are only very distantly related to the third group from the left (from text 48 to 74). These three groups together are well separated from the group furthest to the right (from text 2 to 40)

Analysis of the composition of each group so far identified in the dendrogram reveals the following:-

The group furthest to the left contains:

all the texts numbered 1 to 20 except text 2 and text 11.

The second group from the left contains:

all the texts numbered 41 to 60 except text 48 but also including the misclassified text 11.

The third group from the left contains:

all the texts numbered from 61 to 80 but also including the misplaced text 48.

The fourth group from the left contains:

all the texts numbered from 21 to 40, but also the misclassified text 2.

The first cluster (of 18 texts) on the left of the Ward's Method dendrogram is virtually the same in composition as the first group on the Independent Variable (variable 1). That is texts 1 to 20. This corresponds to texts sampled from the GIL Personal Address

Population. The only texts which are unexpectedly absent from the cluster are texts 2 and text 11. Text 11 is re-classified by the cluster analysis as BBSIL. It is not however one of the texts misclassified in the manual analysis of distance co-efficients. Text 2 is re-classified as I.A. by the cluster analysis. The manual analysis of distance co-efficients also highlighted this text as atypical of P.A.

The second cluster (of 20 texts) from the left is virtually identical in composition to the third group on the Independent variable (texts 41 to 60). This cluster corresponds therefore with texts sampled from the BBSIL population. The only text which is unexpectedly absent from this cluster is text 48. Text 48 is re-classified by Ward's Method as DMA, exactly the same result as obtained in the manual analysis of section 3.5. Text 11, not misclassified in the manual analysis, unexpectedly occurs in this cluster.

A perusal of text 11 does not reveal any features which appear to make it more appropriately classifiable as BBSIL than P.A. . The most likely explanation for the re-classification in Ward's method of text 11 would appear to be that the two clusters involved are close neighbours anyway: In the dendrogram the cluster corresponding to P.A. and that corresponding to BBSIL are the first two really large clusters to be fused. This observation of close proximity is supported by the manual analysis in section 3.5 above. It is not

surprising therefore that at least one text has been re-classified from the one I.V. group to the other.

The third cluster (of 21 texts) from the left corresponds almost exactly with group 4 on the Independent variable (texts 61 to 80). This cluster therefore relates to the sample taken from the population of Daily Mirror Articles. The only additional (re-classified from the BBSIL cluster) text (48) has been well commented on already.

The fourth cluster (of 21 texts) from the left, corresponds almost exactly with group 2 on the Independent Variable (texts 21 to 40). This cluster therefore relates to the sample taken from the population of GIL Impersonal Address texts. The one additional text (re-classified from P.A.) in this cluster would appear to be quantitatively more like I.A. than P.A. texts. An intuitive assesment of text 2 does seem to indicate that it is more formal than the other texts in its group and therefore perhaps better classified as I.A. This is not a very satisfying explanation though, as the address of the leaflet is personal which seems to make it intuitively less formal than most I.A. texts.

Ward's Method cluster analysis seems to suggest greater independence of BBSIL and P.A. from each other than the manual analysis of section 3.5 does. This good agreement between the manual analysis of distance co-efficients and the Ward's Method Cluster Analysis lends

added weight to the validity of the Ward's method result.

Having looked at the results of statistical analyses designed to discover structure among variables (Manual analysis of Rank Order Correlation Co-efficients and Factor Analysis), and those designed to discover structure among texts (Manual analysis of Rank Order Distance Co-efficients and Cluster analysis); the last two sections of the results report on two statistical techniques which reveal something about both these topics (Factor Score Plots and Discriminant Function Analysis)

### 3.7 Factor Score Plot

This plot is to be found on page 370. It resulted from the 2 factor factor analysis solution and was produced by the following method.

After obtaining rotated factor solutions for two factors as described in section 3.4, estimates of scores for each text were computed on each factor for every text, as though the factors represented new variables. These new variables are linear functions of the original variables. The scores obtained were the factor scores. The co-ordinates for any text on any two factors can be plotted by plotting one factor on the x and one on the y axis. In this way the plot for Factors 1 and 2 was obtained.

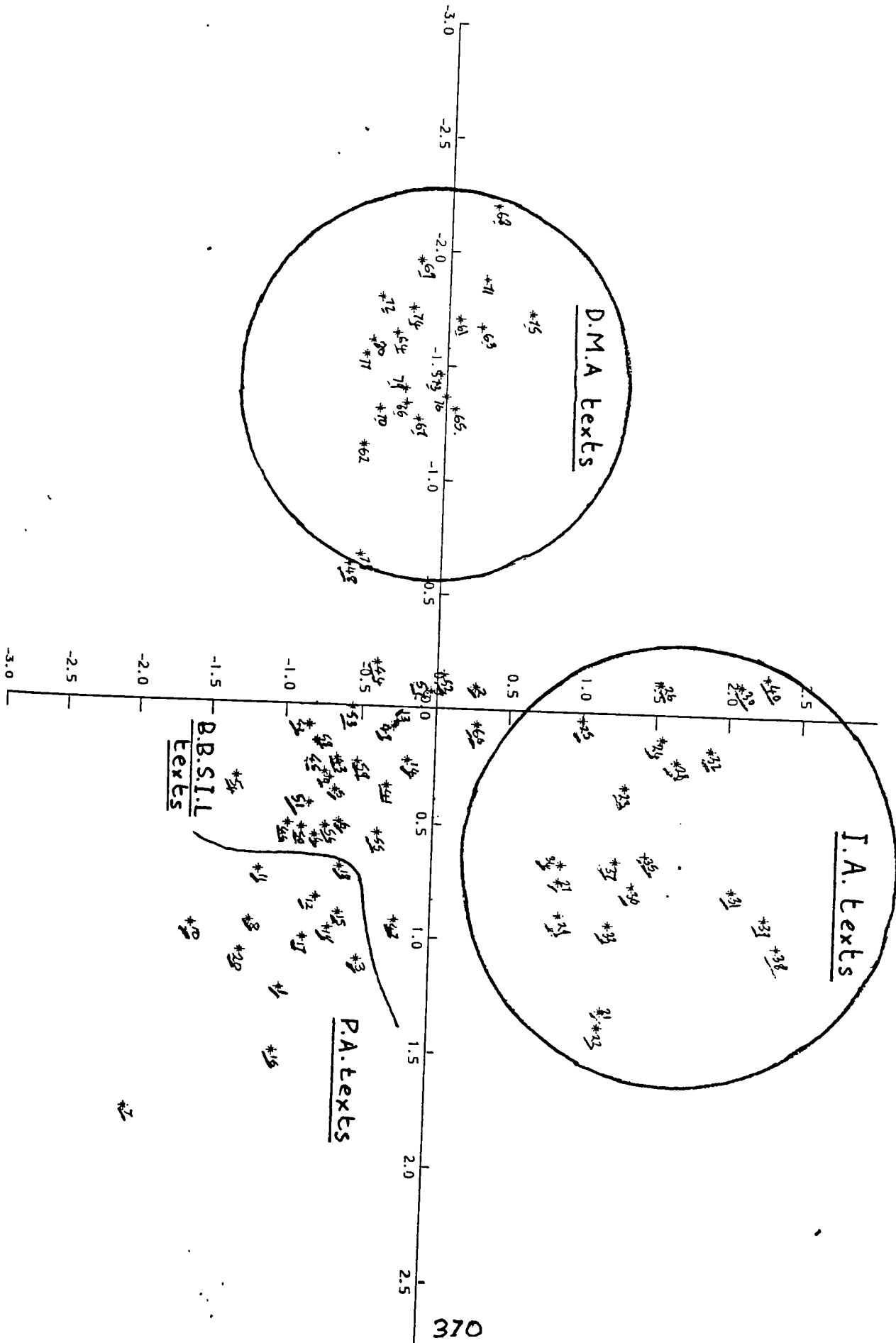
The aim of plotting Factor Scores in this way was two-fold. Firstly to see how well each Factor differentiated the 4 previously



WARD 5 DATA PLOTTED ON TWO MAIN FACTORS

FACTOR 2

FACTOR 1



established text groups, and secondly to create a visual impression of which groups load positively and which negatively on what factors. To facilitate this goal the identity of all the co-ordinates (texts) on the Factor Score plot was ascertained and added to the plot.

As can be seen from the plot of Factor Scores 1 and 2, the co-ordinates corresponding to Group 2 (I.A.) and Group 4 (DMA) form 2 tight homogeneous clusters clearly separated from all other groups. The co-ordinates corresponding to group 1 (P.A.) and group 3 (BBSIL) are jumbled together, though these two groups are separated from from I.A. and DMA. Factors 1 and 2 taken together therefore distinguish DMA, I.A. from each other but fail to distinguish P.A. from BBSIL.

In terms of positive and negative loadings I.A. has positive loadings on both Factor 1 (Explanatory v Descriptive Text) and Factor 2 (Formal v Informal Text), suggesting that I.A. texts are a formal, explanatory type of text. DMA texts have strong negative loadings on the Explanatory v Descriptive Factor, and some moderate positive and negative loadings on the Formal v Informal Factor. These results suggest DMA to be a descriptive text of lesser formality than I.A. . Both BBSIL and P.A. have moderate positive loadings on the Explanatory v Descriptive Factor and moderate negative loadings on the Formal v informal Factor. This suggests that these styles jointly are less formal than I.A., but explanatory in a similar way to I.A.

In overall significance, when considering what if anything it reveals about text clusters, the factor score plot appears to provide additional support for the manual analysis of distance co-efficients rather than the Ward's Method Cluster Analysis. This is because the factor score plots seem to suggest that P.A. and BBSIL are not well separated as text categories, as is also suggested by the manual analysis. The Ward's Method Cluster Analysis it will be recalled gives better separation of these two varieties. In stylistic terms, therefore the factor score plot tends to reinforce the tentative assumption that BBSIL and P.A. are stylistically very close neighbours.

### 3.8 Discriminant Function Analysis

The Final batch of the results are the output of a discriminant function analysis. Discriminant Function Analysis is described in detail in the introduction to Part 2. Discriminant Function Analysis is a statistical technique which enables the researcher to investigate how well a set of individuals fit into previously established categories. It therefore performs a similar but not identical function to Cluster Analysis. This method is not identical, because Cluster Analysis does not assume prior groupings of texts.

The version of Discriminant Function Analysis used was that available on SPSS-X. The most useful aspects of the output for the

FIG. 3.8.1 RESULTS OF DISCRIMINANT ANALYSIS

CASE SEQUIM	ACTUAL GROUP	HIGHEST PROBABILITY		2ND HIGHEST GROUP	DISCRIMINANT SCORES		
		P(D/G)	P(G/D)		1	2	3
1	1	0.5726	1.0000	2	2.1771	4.8039	-4.8545
2	1	0.0707	1.0000	4	3.1191	3.2559	-1.3583
3	1	0.4275	1.0000	4	2.3492	4.6952	-2.6338
4	1	0.9812	1.0000	4	2.8395	4.0943	-3.6664
5	1	0.5828	1.0000	4	2.4778	3.5201	-5.2716
6	1	0.7030	1.0000	4	3.0574	2.9830	-4.7930
7	1	0.6496	1.0000	4	3.5382	3.8943	-4.9975
8	1	0.3252	1.0000	4	3.4834	2.9719	-5.4841
9	1	0.7857	1.0000	4	3.5150	3.0701	-4.2269
10	1	0.9548	1.0000	4	2.8886	3.2371	-3.7669
11	1	0.3430	1.0000	4	4.2281	3.0280	-3.0541
12	1	0.8476	1.0000	4	2.7519	4.3754	-3.2749
13	1	0.1715	1.0000	4	2.5001	2.0606	-2.5439
14	1	0.9032	1.0000	4	2.6583	4.2769	-3.3996
15	1	0.2573	1.0000	4	1.9222	3.2255	-2.2310
16	1	0.5377	1.0000	4	2.9280	4.9753	-5.3100
17	1	0.9133	1.0000	4	2.9969	4.2993	-4.4193
18	1	0.9421	1.0000	4	2.3230	3.5902	-4.2727
19	1	0.2925	1.0000	2	2.3020	5.3529	-4.9301
20	1	0.9423	1.0000	4	2.3401	3.8411	-4.3368
21	2	0.4607	1.0000	1	-5.3945	3.7719	1.1303
22	2	0.8753	1.0000	1	-6.3584	4.3495	1.9692
23	2	0.6594	1.0000	1	-5.3663	3.7825	1.5427
24	2	0.1824	1.0000	3	-6.9168	2.7978	4.3314
25	2	0.5751	1.0000	1	-5.1299	4.7409	3.0787
26	2	0.0616	1.0000	3	-5.1104	1.3244	2.4238
27	2	0.4683	1.0000	1	-5.6452	5.2719	2.0251
28	2	0.6310	1.0000	1	-5.0616	4.6504	2.4694
29	2	0.2838	1.0000	3	-5.2305	2.6230	1.3259
30	2	0.7695	1.0000	3	-6.8154	3.4770	3.2650
31	2	0.4902	1.0000	1	-7.0100	5.0350	2.9622
32	2	0.7099	1.0000	1	-5.2770	3.5493	3.3650
33	2	0.8542	1.0000	1	-6.7312	4.4545	2.6396
34	2	0.5709	1.0000	3	-6.9662	2.9761	3.2763
35	2	0.7198	1.0000	3	-7.2310	3.8238	2.7965
36	2	0.4455	1.0000	1	-6.4681	4.7769	1.2854
37	2	0.6610	1.0000	3	-5.3392	3.1722	1.8207
38	2	0.1108	1.0000	3	-8.2645	3.9632	3.7132
39	2	0.6467	1.0000	3	-6.8705	3.7353	3.5957
40	2	0.4463	1.0000	1	-4.7108	4.6335	2.4504

Fig. 3.8.1 (cont)

CASE SENUM	ACTUAL GROUP	HIGHEST PROBABILITY		2ND HIGHEST GROUP	DISCRIMINANT SCORES			
		P(U/G)	P(G/D)		1	2	3	
41	3	0.4463	1.0000	2	0.0000	-3.4959	-5.5107	-0.0717
42	3	0.2194	1.0000	1	0.0000	-2.0074	-5.1424	-1.5657
43	3	0.5815	1.0000	2	0.0000	-3.6074	-7.0155	-2.5672
44	3	0.9921	1.0000	2	0.0000	-3.8886	-5.9356	-1.2539
45	3	0.5495	1.0000	2	0.0000	-4.5094	-4.8028	-1.3442
46	3	0.4673	1.0000	1	0.0000	-3.4251	-6.2979	-3.0601
47	3	0.5227	1.0000	2	0.0000	-4.5411	-4.8376	-2.0107
48	3	0.0090	1.0000	4	0.0000	-0.7957	-7.0821	-0.4347
49	3	0.7959	1.0000	2	0.0000	-3.4073	-5.3326	-1.0866
50	3	0.2302	1.0000	2	0.0000	-4.9411	-4.9425	-2.8068
51	3	0.6399	1.0000	2	0.0000	-4.7546	-6.8830	-0.9381
52	3	0.6513	1.0000	2	0.0000	-4.2448	-4.9487	-1.1298
53	3	0.8782	1.0000	2	0.0000	-3.3307	-6.3739	-2.0956
54	3	0.0235	1.0000	2	0.0000	-3.4554	-8.9708	-2.5744
55	3	0.3024	1.0000	2	0.0000	-5.5924	-5.3789	-1.9724
56	3	0.7642	1.0000	2	0.0000	-4.6422	-6.3499	-2.2541
57	3	0.8018	1.0000	2	0.0000	-4.8756	-6.1154	-1.5059
58	3	0.5157	1.0000	2	0.0000	-3.5388	-7.3809	-0.8134
59	3	0.4043	1.0000	2	0.0000	-5.3604	-6.9031	-1.4258
60	3	0.1144	1.0000	1	0.0000	-3.1499	-5.7853	0.7417
61	4	0.6659	1.0000	1	0.0000	8.3322	-1.7712	2.5661
62	4	0.8764	1.0000	1	0.0000	7.7337	-1.6963	2.3463
63	4	0.2973	1.0000	1	0.0000	5.9493	-0.9614	4.2983
64	4	0.3607	1.0000	1	0.0000	7.1047	-3.3051	3.0572
65	4	0.0014	1.0000	1	0.0000	7.5628	0.6557	2.7237
66	4	0.1763	1.0000	3	0.0000	4.1784	-3.9984	2.1406
67	4	0.1515	1.0000	1	0.0000	5.1609	-2.4595	2.2456
68	4	0.2625	1.0000	1	0.0000	7.7209	-0.3840	4.4526
69	4	0.9286	1.0000	1	0.0000	7.5477	-1.9219	2.5385
70	4	0.5318	1.0000	1	0.0000	7.7477	-2.8528	2.6404
71	4	0.7265	1.0000	1	0.0000	6.1827	-1.4147	3.5092
72	4	0.6163	1.0000	1	0.0000	8.0159	-2.2057	2.1493
73	4	0.5109	1.0000	1	0.0000	7.2111	-0.1225	2.3190
74	4	0.0376	1.0000	1	0.0000	8.8848	-1.0224	5.1920
75	4	0.3261	1.0000	1	0.0000	5.9427	-0.4807	3.8462
76	4	0.3930	1.0000	1	0.0000	8.3822	-0.7119	2.0078
77	4	0.5299	1.0000	1	0.0000	8.4238	-0.8651	2.5035
78	4	0.6227	1.0000	1	0.0000	7.1492	-1.3032	1.6006
79	4	0.6267	1.0000	1	0.0000	6.1736	-1.3371	2.0342
RU	4	0.5743	1.0000	1	0.0000	7.6612	-2.2066	4.0345

FIG. 3.8.2.

POOLED WITHIN-GROUPS CORRELATIONS BETWEEN CANONICAL DISCRIMINANT

FUNCTIONS AND DISCRIMINATING VARIABLES

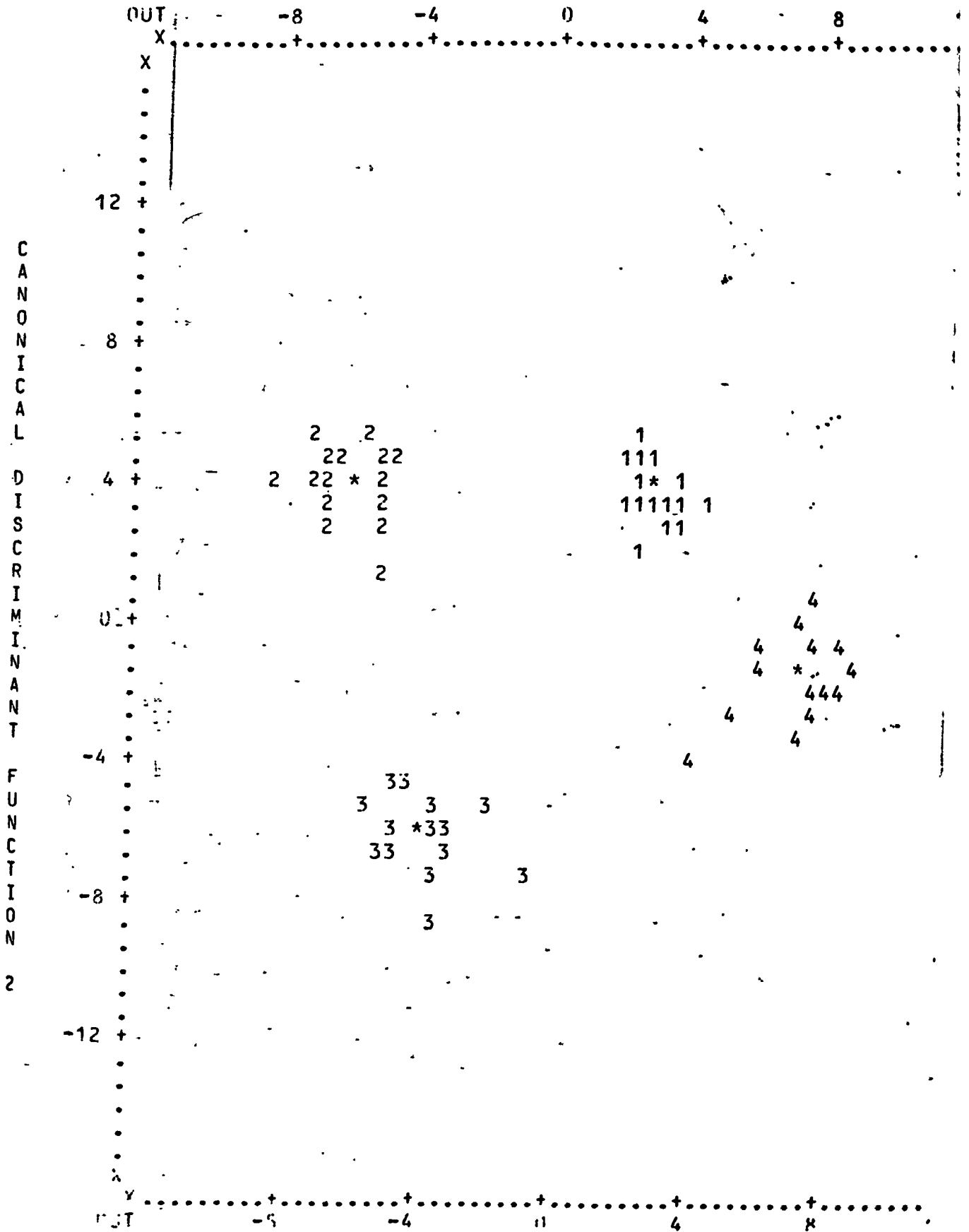
VARIABLES ARE ORDERED BY THE FUNCTION WITH LARGEST CORRELATION AND THE MAGNITUDE OF THAT CORRELATION.

	<u>FUNC 1</u>	<u>FUNC 2</u>	<u>FUNC 3</u>
C21	-0.57600*	-0.27874	0.16119
C30	-0.20892*	0.15839	0.05000
C32	-0.17124*	0.05712	-0.02772
C38	0.13401*	-0.00474	0.10587
C4	0.10368*	0.01387	-0.08213
C35	0.06314*	-0.02177	0.05208
C19	-0.06193*	0.04667	0.03401
C18	-0.14377	0.23342*	0.10086
C11	-0.08298	0.22163*	0.02709
C26	-0.05253	0.13974*	-0.02338
C28	0.10077	-0.13619*	0.02999
C25	-0.03227	0.11477*	0.02277
C3	-0.06437	0.08091*	0.02208
C15	-0.04267	0.07826*	0.04934
C8	-0.15179	0.20987	-0.31398*
C17	-0.03720	0.04672	0.27629*
C22	0.18465	-0.04767	0.26458*
C13	-0.00738	0.03372	-0.25218*
C31	-0.17882	0.12976	0.21265*
C33	-0.10587	0.09866	0.20774*
C9	-0.07782	0.09103	0.18097*
C20	-0.14071	0.13647	0.17648*
C12	-0.02350	0.13572	-0.15558*
C5	0.05980	-0.01425	0.15254*
C37	-0.06563	0.03314	-0.14734*
C34	0.03050	-0.03080	-0.13965*
C14	0.05287	-0.08459	0.13904*
C27	0.00308	-0.08228	-0.12740*
C24	-0.11323	-0.10493	-0.11846*
C7	-0.07550	0.02858	-0.10967*
C36	-0.06889	0.01699	-0.07812*
C6	-0.03628	0.03447	0.07040*
C29	-0.01779	-0.02549	0.06505*
C23	0.04135	-0.02738	0.06039*
C10	-0.05008	0.05451	0.05663*
C39	-0.00291	0.00238	-0.03858*
C16	-0.02690	-0.00103	0.03514*

FIG. 3.8.3.

ALL-GROUPS SCATTERPLOT - \* INDICATES A GROUP CENTROID

CANONICAL DISCRIMINANT FUNCTION 1



present study are displayed on pages 374 to 377. These results begin with a list of all the texts from text 1 to text 80. The actual group membership of each text is displayed and then the group to which discriminant analysis calculates that it should be assigned, with the highest probability. The following column indicates the second most probable group to which the text could be assigned. This section of the Discriminant Analysis' results gives an indication of how accurate the original text groupings are.

The last three columns indicate the scores for each text on the 3 discriminant functions. Discriminant functions are functions of the original variables in the study, calculated by Discriminant Analysis to give the best differentiation between the established text categories. This section of the results is not very interpretable in itself but as a plot of the first two functions (see below).

The third page of output for the Discriminant Analysis gives the pooled within groups correlations between the discriminant functions and the original variables in the study. The correlations are printed in descending order of magnitude, i.e. from the more perfect to less perfect correlations. In each column, the variables which correlate best with a particular function are marked with an asterisk in the column for that particular function.

The final page of output from the discriminant analysis is a plot of the co-ordinates for each text on discriminant function 1 and function 2. This stage of the results gives a visual indication of



how well separated the different text groups are on the individual discriminant functions, and, in addition to this an idea of the homogeneity or otherwise of the groups. In this respect the plot has a similar but not identical function to a factor score plot.

The actual results for this study show that in every case, Discriminant Analysis was able to correctly assign texts to their original groups. That is, on the plot on the previous page, 1 indicates P.A. texts, 2 indicates I.A. texts, 3 indicates BBSIL texts and 4 indicates DMA texts. This 100% correct assignment is of course made on the basis of the three discriminant functions, whose purpose is to give the best differentiation between the groups.

The results of this stage of the analysis seem to suggest a rather clearer differentiation of text categories, and therefore of styles, than was the case with the manual analysis of distance co-efficients and the Ward's Method Cluster Analysis. This rather neat outcome must be treated with suspicion in view of the violation of Discriminant Analysis' assumptions about normality which were present in this study (see section in this chapter on the result of the K - S test), and also the fact that the Discriminant Analysis was given the original groupings as a starting point.

Some interpretation of the grouping of variables into functions 1 and 2 needs to be made here in order that the Discriminant Analysis and the Factor Analysis can be compared. Prima facie, functions 1 and 2 of the Discriminant Analysis do not look very like Factors 1

and 2. None of the variables correlating highly with function 1 appear in either of factors 1 and 2 on pp.354-355. Function 2 has variable 28 in common with factor 1 but variable 18 with factor 2. This is not necessarily surprising since what distinguishes individual cases (texts) well (i.e. factors) may well not be the same as what distinguishes the 4 predetermined text groups well (i.e. Discriminant Functions). What is interesting to ask however is which approach is stylistically more useful, Factor Analysis or Discriminant Analysis?

Function 1 has its highest negative correlations with: agentless passives (v.21), long paragraphs (v.30), long lines (v.32), and noun phrase quantifiers (v.19); and its highest positive correlations with: substitution ties (v.38), simple sentences (v.4) and reference ties (v.35). Of all the correlations, agentless passives (v.21) are the most important with a negative correlation of .576. The remaining correlations are very small (below .21), and ought to be disregarded. Function 1 distinguishes P.A. texts and D.M.A. texts from the other 2 text types and to some extent P.A. and D.M.A. texts from each other (2 cases (texts) overlap).

Function 2 has its largest negative correlation with modal 'Will' (v.28), and its largest positive correlations with: embedded postmodifiers (v.18), initial adverbials (v.11), modal 'May' (v.26), modal 'Must' (v.25), complex sentences (v.3), and NP postmodification (v.15). None of these variables have acceptably high correlations with function 2, though the highest .23 is the positive correlation with embedded postmodifiers (v.18). Function 2

serves to distinguish P.A. and D.A. texts together from the other 2 text types and D.M.A. texts from all 3 others.

The only information then of potential use to the stylistician is the fact that the presence of agentless passives appears is the key variable characterising the I.A. texts and BBSIL texts as a group, and that their absence characterises P.A. texts and D.M.A texts as more or less separate groups.

Although Discriminant Analysis seems to have the potential to pinpoint the key variable distinguishing a group of texts, this is not the same as distinguishing a style. The grouping of BBSIL and I.A. texts as a style on the basis of their use of agentless passives does not have a great deal of intuitive support. These two text types read very differently, and though they may share one common feature this does not mean that they share the same style. Discriminant Analysis seems to have produced a neat partitioning of the texts into the required groupings and little more. This is not surprising since the program was given the text categories as a starting point.

Discriminant Analysis has therefore proved rather a disappointing research tool in this instance, though this does not mean that it should not be tried elsewhere: far more research needs to be done before this type of analysis can be rejected as unsuitable.

The results of the Discriminant analysis complete the Results

f

section. The Discussion which follows sets out to analyse the further and wider significance of these results.

## PART 2 - CHAPTER 4 - DISCUSSION

This chapter begins with a summary of the results, and an evaluation of the effectiveness of the statistical techniques used. This is followed by a comparison of these techniques with those used in earlier studies. The final part of the discussion is a consideration of the theoretical significance of the study's findings and their possible practical applications.

### 4.1 Summary of the Results & Evaluation of Techniques

The results of the K - S tests revealed an unavoidably large number of non-normally distributed variables in the data. This had been foreseen, and had the effect of ensuring that greater reliance would be placed on non-parametric techniques in analysing the data.

Examination of the results for I.A. and P.A. on individual variables for the K - W test provided extremely good support for the findings of the qualitative study. All the syntactic variables found to be significant in the pairs test for these two styles had been predicted as stylistic variables participating in I.A./P.A. contrasts in Part 1. What was completely unexpected however was the evidence presented on the next page in summarising the results of the Cluster Analysis, suggesting that I.A. and P.A. do not form a single GIL style.

The results of the K - W test generally showed that a high proportion of the variables selected for analysis yielded significant differences at the 5% level and could therefore be considered to be stylistic variables. These results therefore showed that the initial qualitative assessment of these variables as style markers was an accurate one. Further batteries of tests were then used to explore the data further for patterns of variables and groups of similar texts.

A manual analysis of distance co-efficients, a Cluster Analysis, and a plot of Factor Scores all revealed a remarkable degree of agreement on the homogeneity of two of the styles studied: GIL Impersonal Address and Daily Mirror Articles. However, the remaining two styles: GIL Personal Address and Banking and Building Society Leaflets, were shown by all the tests to be interlinked, and not entirely distinct styles. In addition, Cluster Analysis showed in its hierarchical ordering of the styles, evidence that I.A. and P.A. are not both members of the same GIL 'Bureaucratic' style as was originally expected, but only more distantly related to one another. P.A. being far closer to the BBSIL and DMA styles. Cluster analysis, and the other 'cluster type' techniques used proved essential in providing a check on the reality of what were in fact only 'supposed' styles.

A manual analysis of rank order correlation co-efficients and a two factor analysis yielded strong agreement on the existence of two

large groups of variables which may be interpreted as providing evidence that two main stylistic factors are at work in providing some differentiation of the styles in the data. One of these factors is particularly strongly associated with the presence or absence of conditional and concessive adverbial clauses in a text. The high frequency of this feature in a text appears to be a strong marker of information leaflet MODALITY, as it covers BBSIL as well as P.A. and I.A. texts.

A second factor however, shows this initial picture to be rather too simple: It is a factor strongly associated with the presence or absence of passive verbs in a text. The high frequency of this feature in a text seems almost certain to be connected with formal/informal STATUS. A factor Plot including this factor shows that it makes a strong separation between I.A. (formal) and P.A. & BBSIL (informal), supporting the Cluster Analysis evidence that there is indeed no homogeneous GIL PROVINCE.

Throughout all the tests used, strong evidence emerged that manual analyses of correlation and distance matrices can produce very similar results to Factor and Cluster Analysis. This seems to indicate that the lack of normally distributed variables in a stylo-statistical study need not hamper the exploration of patterns of variables and texts at all. The non-parametric statistics used, in these manual methods are statistically less complex, easier to understand, and applicable to both normal and non-normally

distributed data. The question therefore arises as to whether such complex techniques as Factor and Cluster Analysis proper are really necessary in stylo-statistical work at all.

A final test; Discriminant Analysis, showed 100% classification of the 80 texts in the sample into their original categories. Whilst such accurate partitioning of the data must not be given too much credence, (due to the violations of normality assumptions in the data), the discriminant analysis provides yet more support for the notion that the study is dealing with four distinct style categories. In this, discriminant analysis tends to agree with the manual analysis of distance co-efficients, cluster analysis and the plot of factor scores.

#### 4.2 Comparison with Earlier Studies: .1 Preliminary

Direct comparison of the results here with those of earlier studies is in fact impossible. The present study is the first one (to the author's knowledge to have made a specific stylo-statistical study of Administrative/Bureaucratic Language. That of Chiu (1973) sets out to examine Administrative English, but in fact confines itself to a study of lexical items in Canadian English Correspondence. Two studies which initially seem more promising as regards direct comparison of results are Biber (1985) and Markworth and Baker (1979). This is because they both include in their analyses texts from the Brown Corpus which come under the heading 'Government



Documents'. There is no indication in either study of exactly what sort of text 'Government Documents' are, but it seems likely from the use of the word 'documents' that they are more legal than bureaucratic texts. Even if the text types are comparable, the system of counting and scoring variables used in these studies definitely is not: Very few of the variables studied by Biber (op cit) and Markworth and Baker (op cit) are used in this study, and those that are, are counted as absolute frequencies per portion of text rather than as realised potential occurrences as is done here.

A possible criticism of this study is that it should have used the same variables as either Biber or Markworth and Baker, and measured these using their methods to assist comparison. This would not have been at all desirable. Both Markworth and Baker include 'Government Documents' in large macro studies to establish stylistic differences between spoken and written, or fiction and non-fiction styles. Part 1 of this study was a thorough preparation for part 2 in which precisely those variables which would be stylistic in the Administrative/Bureaucratic language context were chosen. This selection was therefore far better suited to the task in hand.

It can also be maintained that counting potential occurrences is a superior method of scoring variables for stylistic purposes. This is because absolute frequencies are related simply to the saturation of a feature in a portion of text. This information is not directly stylistic since it takes no account of how many structural

opportunities an author had in which to use a particular linguistic item, though it does admittedly reflect how often a reader would meet a relevant item when reading (something which the present author's method does not). It is maintained here however that Style is a matter of partial and constrained selection from the complete structural system available in English. Any method of scoring variables to yield stylistic information needs take account of this, hence the bias of the present study.

If the results of previous studies cannot be used to show possible failings in the present one, neither can they be used as supporting evidence. The findings here have to remain initial and tentative therefore. However, since the use of statistical techniques has been as yet relatively little used in linguistic stylistics, it is very important here to make a thorough assessment of the effectiveness of the techniques used here compared with those of earlier studies. This is the main subject of this section. It opens up comparisons with all stylo-statistical studies, not simply those which analyse Administrative/Bureaucratic text.

The remaining subsections of section 4.2 compare the techniques used in this study with that of earlier studies. Each earlier study will be the subject of a separate sub-section. The studies are discussed in date order from the oldest to the most recent.

#### 4.2.2 Carroll's (1969) study

Carroll set out with the rather ambitious intention of using Factor Analysis to obtain a picture of the major stylistic dimensions

operating in English. The texts chosen for study were drawn from a wide variety of sources from novels to scientific papers, and included; essays, newspaper articles, biographies, textbooks, speeches, sermons and school essays.

Carroll took the unusual step (for a stylo-statistic study) of including raters' qualitative judgements of texts as well as syntactic variables in his study. The qualitative variables were scores on sliding 7 point scales of the semantic differential type. There were 29 of these in all, and 38 syntactic variables. There do not appear to be any lexical variables.

A Factor Analysis of an unspecified type, and with the type of rotation used left unreported, was claimed to yield seven factors. No information at all is included on whether these factors exhausted all the variance in the data, and if they did not how were they selected from the remaining factors. Only the first six factors were interpreted. The loadings for most variables on Factors 6 and 7 are rather low (mainly below .3). The impression gained from this is that some rather trivial factors have been included. However as no information is provided on the amount of shared variance for which each factor accounts, it is not possible to substantiate this claim.

Carroll takes as significant, loadings on his factors which are above .25 (ibid: p149), this seems remarkably low when compared with the present study. To talk of factors as being 'significant' as Carroll does presupposes that scores on his variables come from a

normally distributed population. It seems intuitively unlikely that judgements of the type used by Carroll would show normal distributions, and the present study has shown that the normality assumption is often violated for linguistic variables.

Because Carroll pays attention to very small loadings (under .3), his factors are often said to be loaded on large numbers of variables: Factor 1 is loaded on 17 for example. These loadings vary in magnitude from figures very near to 1 to .25. The higher loadings are likely to be far better indicators of particular style dimensions than the very low loadings, for this reason it seems likely that Carroll may have included a good deal of stylistically irrelevant material in his factors.

Carroll's first factor and strongest factor (this factor must account for more variance than later factors), has high loadings on variables connected with subjective judgements. It is therefore of no use to the linguistic stylistician, who must be essentially concerned with the relation of qualitative judgements about style to linguistic variables. The remaining factors have loadings on both subjective and linguistic variables, but are interpreted in the light of the subjective loadings. Having obtained and interpreted his factors, Carroll does not go on to see how well they distinguish the texts in his sample; sermons, scientific papers, essays etc, but instead uses them to produce profiles of the works of two literary authors (ibid:pl54).

It is not surprising that Carroll does not go back to see if his factors make clear distinctions among his sample texts as is done in the present study: The title of Carroll's article is "Vectors of Prose Style", and therefore leads one to believe that Carroll does not see his factors as style dimensions capable of isolating particular styles, as portrayed in the present study. They are on the contrary intended as general stylistic attributes which cannot distinguish particular styles on their own, but together, distinguish a writer's style by a profile obtained over all the factors (Factor Analysis is of course well suited to this work since its tendency is to differentiate cases (texts) well). Although not specifically stated by Carroll, this does give the impression that he sees styles as unique properties of individual writers.

In addition, the two sample style 'profiles' put forward on p154 of Carroll's paper are of novels, thereby giving the impression that Carroll is more concerned with literary rather than, linguistic stylistics. Carroll appears to be most interested by what Crystal and Davy (1969:p76-77) refer to as the 'Singularity' dimension of situational constraint, i.e. the element of literary texts not explainable by relation to other style dimensions like 'status' etc.

As a measure of of true Singularity, Carroll's six factor's are very misleading. This is because the factors produced in Carroll's study based as they are on mainly non-literary texts, are bound to pinpoint elements of style dimensions such as 'status', 'modality' and so on. These dimensions are undoubtedly present in all literary

texts to different degrees in that many different real life situations are portrayed in literature. The general acknowledgement however, that literary styles have their own unique properties, apart from those which distinguish non-literary styles is not accounted for in Carroll's work; precisely because Carroll fails to distinguish the two when interpreting his factors. It is indeed inconceivable to imagine how the two could be distinguished when the present state of knowledge about non-literary styles is so low.

In the opinion of the present author, the further research of non-literary styles, on their own is a necessary prerequisite to the far more complex task of investigating what linguistic features make a style 'literary' or distinguish between literary styles.

It is not only Carroll's lack of a sound theory of stylistic variation which undermines the theoretical significance of his study, but his neglect of providing essential details about why he chose to extract the number of factors he did, and how much variance they account for. Without such information it is virtually impossible to assess the practical relevance of Factor Analysis to Stylistic study.

#### 4.2.3 Moerk's (1973) Study

Moerk, like Carroll made use of Factor Analysis as a tool in Stylistic Research. He is not as ambitious as Carroll and sets out only to discover the stylistic dimensions present in a sample of 30 short stories written by students.

Moerk's method of obtaining his sample appears exemplary; He obtained texts under experimental conditions by providing a sample of students with the first few lines of a particular story and asking them to continue it. Moerk reports that in his experience the opening sentence used tends to make sure that the content of the stories obtained is not too diverse. Moerk does not provide any direct evidence for this.

Moerk's concern to avoid subject matter as a source of variation in his sample texts can be compared to Carroll's (1969:p 155) anxiety on this matter. Carroll expresses the wish (ibid) to separate style and content, but regrets that the two are very closely linked. The framework of stylistic description used in the present study assumes that content constrained differences in texts are stylistic and are part of the Province dimension of Style.

The wish of both Carroll and Moerk to separate style and content seem to be due to a preoccupation with literary style phenomena, it seems odd therefore that Moerk made no attempt to exclude other

style dimensions such as Status, Modality, etc from his study, as these are also clearly not exclusively connected with literary style. Moerk's preoccupation with the style content problem therefore seems to result from the use of a rather unworkable theory of style.

In relation to the statistical test used; Factor Analysis, Moerk's design seems rather likely to be unreliable. This is because, whilst he uses only 30 subjects, he chooses to analyse an enormous 55 linguistic variables. Among the general heuristics suggested when considering designs suitable for Factor Analysis, it is held that a minimum of twice as many subjects as variables are needed.

Moerk reports the results for 11 factors but interprets only 4 of these. If factors 5 to 11 each account for large percentages of the shared variance in the data this is rather serious, as it seems to suggest either that these factors are concerned with something other than that which Moerk was looking for, i.e, Style dimensions, or that Moerk lacked the knowledge to interpret them. Any attempt to make an independent assessment of the meaning of factors 5 to 11 in Moerk's study is frustrated by the generally weak loadings for all variables on them. This tends to suggest that the factors contain little of real stylistic import.

Moerk appears to have stopped extracting factors when they failed to



yield significant loadings, (ibid:p52). This is a very unsatisfactory method, as even if a factor has some significant loadings, it may account for no more than a very trivial amount of the shared variance in the data. No information is forthcoming on how much of the shared variance in the data each factor extracted accounts for.

Moerk observes in his discussion (ibid:p57) that whilst factor 1 has a vast number of significant loadings, the remaining factors have very few. Moerk tenders a possible explanation for this, which is that some dimensions of style are very 'broad' whilst others are 'narrow'. From a statistical point of view it seems all too likely that only Moerk's first factor accounted for an appreciable amount of variance in the data and therefore was the only viable factor. It seems all too likely that there was very little correlation between Moerk's variables, and that most of them were measuring different things rather than representing factors. This assumption is to some extent supported by the uniformly low loadings on Moerk's later factors.

The same problem occurs to a some extent in the present study, where only two strong factors could be extracted, accounting together for only 33% of the shared variance. None of the remaining factors account for more than a trivial amount of variance. This is clearly a problem in the use of factor analysis in stylo-statistical research, since it tends to cast doubts about the ability of

linguistic variables to pattern very strongly in response to style dimensions. The answer to the problem appears to be to be very careful in selecting only those variables likely correlate strongly with particular style features. This appears to produce better success with Factor Analysis, as the study by Sandell (below) shows

#### 4.2.4 Sandell's (1977) Study

Sandell's study represents a far more thoughtful design than either of the previous two studies. Sandell set himself the very specific aim of isolating those linguistic features which correlate highly with an intent to persuade (ibid:pl15). Sandell believed that this intent should be strongest in texts identified as advertising texts. The aim of his study therefore was that of finding out what linguistic feature or features were prominent markers of an advertising style.

In deciding on a study based on natural observation, Sandell was careful to try and eliminate unwanted variables, and considered texts from two carefully delimited time periods, ten years apart, (ibid:pl15). It seems strange however that he should not have chosen to control for other unwanted variables however.

Three text groups were considered for analysis by Sandell, all from editions of the same newspaper, and all supposedly varying in their

intent to persuade, (ibid:pl21). These were; advertisements, home care editorials, and foreign news telegrams. Only 12 variables were used, all linguistic, and these were selected by Sandell as a result of discussing with advertising professionals what variables were most likely to be markers of advertising language, (ibid:pl15). Sandell's use of this method may have been an important factor in the success of the Factor Analysis he carried out (see below).

In scoring his variables Sandell uses the absolute frequency method. This method was rejected in the present study as inappropriate to stylistic work, (see page 279 above). The two factors of intention to persuade and time of publication were to be measured for significant effects in a multivariate ANOVA.

Sandell, like the two previous authors discussed uses a Factor analysis to analyse his data, but alongside a battery of other tests; One and Two way ANOVAs, Discriminant weights, and Correlation Co-efficients. In this way Sandell is able to accumulate considerable evidence from different sources to show a relationship between the linguistic features chosen, and persuasive intent. The One way ANOVA showed the majority of Sandell's variables to be stylistic, and the multivariate ANOVA a strongly significant intent to persuade, but no effect from time of publication.

In examining the mean scores for each of the three groups of texts on the significant variables, Sandell found that they were

frequently ordered with respect to what he considered to be intent to persuade: That is with mean scores for each of the three groups of texts, Sandell found that half the variables were ordered with respect to what he considered to be intent to persuade (ibid:pl22): that is with Advertising texts first, followed by Home Care Editorials and then News Telegrams. This strengthened Sandell's opinion that the variables chosen were associated with an intent to persuade.

Parametric Correlation Co-efficients were used on Sandell's data although no check was made on the normality of the variables involved. Sandell obtained two correlation matrices for all the variables. One ignoring group differences (termed between-groups), and one taking account of between group differences (termed within groups) (ibid:pl23). It is not at all clear from the text how Sandell could obtain a single matrix for the latter when there were 3 groups of texts.

The correlations for the two matrices were compared, and where correlations observed in the within-groups analysis were even higher in the between-groups analysis, this was to be held to be due to persuasive intent (ibid). The reasoning behind this is not easy to understand. Even if it is accepted, there are still problems. This is because, of the many other unmeasured variables which might have been the cause of between group differences, Sandell only controlled for time, the increase in magnitude of correlations might

therefore have been due to a different style dimension or dimensions.

A similar use is made of Factor Analysis, which was done on the between, and the within group scores. Again only one solution is produced for within-groups. Sandell uses Kaiser's method to decide on the extraction of 3 factors for both Factor Analyses conducted. Again increases in factor loadings from within to between groups was held to be an indication that a factor was concerned with the expression of persuasive intent. The first factor, which accounted for 37% of the variance (in the between-groups analysis), and the third factor (accounting for 15% of the variance) were pinpointed as being persuasive by using this method, (ibid:ppl26-127).

What is perhaps surprising about Sandell's between groups Factor Analysis, when compared with the present study, is that most of the shared variance in Sandell's data (81%) is explained by 3 factors. The present study compares very unfavourably with this, where only 2 strong factors emerged, accounting for no more than 33.3% of the shared variance. Sandell offers no clear explanation for the very successful outcome of his factor analysis, but it seems likely that Sandell's careful selection of variables, made after discussion with advertising professionals may have resulted in variables which were strongly related, thereby giving rise to a strong factor structure.

No attempt was made to consult writers of Government Leaflets in the present study. The supposed degree of unintelligibility of Official Prose has become a very sensitive issue, and it was therefore not thought possible that official writers would be particularly objective if approached for information about their style of writing. (Instead, the literature of pressure groups like the Plain English Campaign was studied for clues as to what might be strong marker of official prose, and a qualitative study<sup>made</sup> to pinpoint possible style markers. It was hoped that the present author's experience as a Welfare Rights Officer might also help in the intuitive selection of possible style markers.

It may be that the process of initial research on what variables to use, employed in this piece of research was therefore mistakenly orientated towards the worst examples of official prose, rather than what was typical, whilst the texts themselves were a true, random, and therefore typical sample. More typical style variables in Official Leaflets may not be those that attract public attention, but one would expect them to be stronger markers of an official style.

It could therefore still be that there are variables which provide evidence for a unitary Government Leaflet Style; something for which no real evidence has been discussed in this study. On the other hand, the selection of variables used in this study was broad and relatively large (compared to Sandell's for example), with some

variables added for the sake of closure. The likelihood of important style markers being missed therefore seems small.

The only real solution to this problem is to recommend that the study be re-run with more variables, and a greater number of texts.

To summarise Sandell's study, it seems that his use of factor analysis shows how successful this technique can be in analysing style. It highlights a possible bias in the design of the present study.

#### 4.2.5 Markworth and Baker's (1980) study

Like Sandell, Markworth and Baker used several tests to analyse their data. But the sample of texts used in this case was intended to cover a very wide area of stylistic variation, rather than a narrow band as with Sandell (above). The texts types sampled by Markworth and Baker covered Fiction, Learned journals, Government Documents (these last 2 considered to be formal), Popular Journals and Newspaper Reports (considered informal), (ibid:p236). Scoring on the 36 syntactic variables used was in terms of absolute frequency per 100 sentences, (ibid), though many of the items counted vary in potential for occurrence depending on sentence type.

ANOVAS were run on all variables to see how well they differentiated:

- a) fiction from other styles
- b) formal from informal styles
- c) the two formal styles from each other
- d) the two informal styles from each other

No actual check was made on the assumptions about normality which ANOVA requires. As this study has shown, normality cannot be relied upon for all linguistic variables. Even so thirteen variables had such obviously non-normal scores that they had to be excluded from the analysis, though they could have been analysed using a non-parametric technique. The present study has shown non-parametric techniques to be effective in stylistic work. Had Markworth and Baker used non-parametric techniques universally on their data they would not have had to ignore over a third of their data, nor to have taken the risk of obtaining unreliable results by the use of possibly undetected 'non-normal' variables. In this respect, non-parametric techniques are clearly to be preferred to parametric ones, for stylistic work.

Of the 23 remaining variables available for analysis only sixteen were shown to participate in stylistic contrasts. Most of the variables (12), served only to differentiate fiction from non-fiction styles, (ibid:pp238-239).

Markworth and Baker also used a Discriminant Function Analysis on their data, as has been used here. Four discriminant functions must



necessarily have been produced, but only a plot for the first two is shown (ibid:p239). This is a default on the SPSS-X statistical package, which Markworth and Baker may have used. The same procedure was followed with the present study. Unlike the present study Markworth and Baker only show the group centroids (mean discriminant scores) for each separate text group. In the present study, the co-ordinates for each text are shown, as well as the group centroids, therefore giving information about how much groups overlap, and how many cases (texts) cannot be assigned to their original categorisations. This is the main purpose of Discriminant Analysis. In the present study there is no overlap of groups and all texts were classified correctly.

It is odd that Markworth and Baker do not make use of the prime function of Discriminant Analysis (see Ehrenberg 1975:p272), to check up on the homogeneity of their text/style categories, and instead try to interpret the functions in a similar way to that in which a factor analysis was interpreted in the present study, especially as any such stylistic interpretation of the functions obtained must rest on the premise that the texts form distinct and independent styles.

Although, individual text co-ordinates are not shown on Markworth and Baker's plot (ibid:p239), the group centroids of all non fiction varieties seem too close to preclude overlapping text categories. This in effect means that non-fiction styles are clearly not stylistically distinct from each other, on the information in the

two strongest discriminant functions in Markworth and Baker's analysis. The only really clear separation of text groups is into fiction and non-fiction styles on the basis of discriminant function 1. Markworth and Baker do however re-run the discriminant analysis without the fiction text (ibid:p242), to give what appears to be a slightly better separation of the text/style categories.

Markworth and Baker's findings are surprising, in the light of the very clear results achieved in the present study with Discriminant Analysis. They are so surprising that further and more detailed comparison is necessary. Markworth and Baker report the correlations of 8 variables with Discriminant function 1 of their analysis, (ibid:p240). Of interest is the fact that past tense marking and adverbial clauses are characteristic of fiction but not of non-fiction, and 'passive constructions' characteristic of non-fiction but not of fiction. These variables were measured in the present study, and measure the same items, though not in exactly the same way. They produce rather weak correlations in the discriminant analysis, but are notable for the way in which they distinguish Newspaper style from the Information leaflet styles. This can be clearly seen in the mean rank scores output from the K-W test in Table 4.2.5, below:

Table 4.2.5

VARIABLE	MEAN RANKS			
	DMA	BBSIL	I.A.	P.A.
adverbial cls frequency	29.63	34.63	56.93	40.82
past tense frequency	68.43	26.33	33.70	33.55
passive verb frequency	30.15	31.90	66.35	33.60

Note: high mean ranks indicate high frequency of occurrence

The results of the present study and that of Markworth and Baker do not necessarily contradict one another. On Markworth and Baker's plot of discriminant function 1 Newspapers are closer to fiction than other text groups (this does not necessarily imply that they are fictional!). A second discriminant analysis conducted by Markworth and Baker, (ibid:pp242 ff), excluding fiction texts reveals past tense to be characteristic of newspapers and popular journals but not of government documents, the reverse is true for adverbial clauses. Though I.A. and P.A. are not directly comparable to Markworth and Baker's Govt Documents, a similar trend is observed above, in table 4.2.5 .

Only in this very small area of comparison is it possible to compare

the results of Markworth and Baker's study with those of the present one. There is rough agreement in this sphere however. The remaining variables distinguishing Government Documents from Newspapers in Markworth and Baker's Analysis might therefore be useful to include in a re-run of the present study, of the type suggested when discussing Sandell (1977) above. These are; inclusion of direct discourse, verbal contractions, transitive verb frequency (infrequent in Govt Doc's), partially reduced relative clauses and noun adjuncts (frequent in Govt Documents).

To summarise this discussion of Markworth and Baker (1980), their study shows parametric tests to compare rather unfavourably with non-parametric types for stylistic analysis. The results of their discriminant analyses are partly supported in the present study, and highlight possible fruitful variables for further study. Markworth and Baker do however fail to use the results of their discriminant analyses for its prime purpose of investigating the homogeneity or otherwise of previously established groupings. In this case 'groupings' refer to text/style groupings. A Factor Analysis, as was used by Biber (1985) below, might have been a more appropriate method of investigating the patterning of variables in their data, which Markworth and Baker obviously required.

4.2.6 CLUSTER ANALYSES: a) McEntegart and Le Page's (1982) study &  
b) Bailey (1979)

McEntegart and Le Page's study is only briefly mentioned here because it is Sociolinguistic and not Stylistic. The main reason for mentioning it is that it employs Cluster Analysis as a research technique. There are very few linguistic studies that do this, and no true linguistic stylistic studies. Bailey (1979) is a report of an authorship study.

McEntegart and Le Page applied Cluster Analysis to data consisting of phonological feature variables and informants from two speech communities in Belize and St Lucia. The technique was used to try and discover if informant clusters based on the phonological variable scores were related to particular socio-economic factors, (ibid:pl16).

The study is unusual in the literature because it reports a failure to obtain satisfactory results. McEntegart and Le Page had not expected to find true typologies (discrete clusters), but administrative clusters (overlapping clusters), because of awareness that the speech communities sampled were in a state of flux (ibid). The authors however found administrative clusters in only the Belize data. This was produced by Ward's method (as used in the present study, and pointed to the existence of 4 clusters. McEntegart and Le Page performed several Cluster Analyses and found that Canonical

Variate Plots were very helpful in deciding which Clustering solution to use. Another method used by the authors to help make their decision, and roughly paralleled in the present study by the use of Factor Analysis, was a Principal Components analysis, (ibid:pl16). McEntegart and Le Page do not report on this.

The present study agrees with McEntegart and Le Page's in finding Ward's method to give a good partitioning of the data. In the present case the clusters obtained were very nearly true typologies and not merely administrative clusters. It is however not possible to draw the conclusion from this that Ward's method is the best Clustering technique for linguistic research, there are too many other Clustering techniques available to make generalisations like this. Further comparison of Ward's Method with a variety of others needs to be done in future studies.

All other approaches to the clustering problem used in the present study - manual analysis of distance co-efficients, factor score plots, and discriminant analysis - tended to agree very well with each other, though there were small differences in Cluster Membership. McEntegart and Le Page's difficulty in this sphere, in particular in not finding Principal Components Analysis helpful, may well have been due to the rather closely overlapping nature of their clusters.

In contrast to McEntegart and Le Page's lack of success with Clustering techniques, the present study shows them to be very

useful in investigating stylistic data. This is no doubt due in part to the fact that the style categories investigated had been qualitatively identified beforehand as such, and were therefore apparent even without using statistical techniques. This may therefore mean that while Cluster Analysis provides a useful check on the homogeneity of style/text categories, it would not be as useful on data in which a stylistician could discern no clear groupings, i.e. as a means of discovering styles. Far more research however needs to be done in this area before any strong conclusions can be drawn. The present study merely points towards these areas of potential useful research.

The study reported by Bailey (below) is similar to the present one in that it was used by defence lawyers in the trial of Patricia Hearst (Bailey 1979:pl1) to prove that texts allegedly written by the defendant did not cluster with other texts known to be by her, but with the political writings of members of the Simbianese Liberation Army who kidnapped her.

The evidence, displayed in the form of a dendrogram does show that Hearst's own writings do not share the same cluster as the texts allegedly by her. as Bailey points out however, this is not proof of authorship and could be due to other factors; for example the alleged Hearst texts have a political content whilst other known samples of her writing are from a travel diary and university examinations.

The same problem exists of proving what style factors are responsible for the clusters obtained in this study of course. This problem cannot be overcome by any direct validation technique, as none exists, (ibid:pl1). What can be said about the status of the clusters in the present study is that the same clusters show strong stylistic distinctiveness when the results of the K - W test are considered alongside them. There is therefore good evidence though not proof, to presume that the clusters are style related categories.

A rather different problem which has to be faced in this study is that of that of whether the Ward's Method Cluster Analysis used here is the most accurate portrayal of the 'true' clusters of texts in the data, when other the possibilities are considered: There is fairly good comparability of the Ward's Method solution with the other 'cluster type' (though not Cluster Analysis) solutions; the manual analysis of distance co-efficients, discriminant analysis and Factor Score Plots. All these methods agree on the homogeneity of I.A. and DMA. Where they tend to disagree slightly is in how much P.A. and BBSIL overlap. In the Discriminant Analysis these two categories are mutually exclusive, in the Cluster Analysis there is some slight confusion of the two, this becomes even greater in the manual analysis, and very great overlap is observed in the factor score plots.

There is no easy answer to this problem: None of the results are



accompanied by significance levels like those which accompany the K-W test. It is therefore not possible to say that any one solution is more significant than another. The assessment must therefore be qualitative and based on what seem to be the most likely clusters to occur.

Considered in this way, the solution obtained from the discriminant analysis seems too neat, and intuitively unlikely to be accurate. The Factor Score plots seem to confuse P.A. and BBSIL rather more than might be expected, and again seem intuitively less likely to be totally accurate. There is however extremely good agreement between the Ward's method Cluster Analysis and the manual assessment of distance co-efficients. Both of these could equally well be accurate, it seems better however to accept the solution of the manual analysis as more reliable than that of Ward's method in view of the violations of assumptions about normality which were necessary when using the latter technique, but to add to this the observation that Ward's method provides extremely good support for the manual analysis

#### 4.2.6 Biber's (1985) study

Biber's (1985) study (also re-reported in a later article; see Biber 1986), is an important study in linguistic stylo-statistical work because it is (at the time of writing) the most recent item in the literature. For this reason it seems particularly important to look

at Biber's general aims and method, as well as examining the success of the statistical techniques used.

Biber's study is a large computer counted study, and rather similar to that of Markworth and Baker (above) in that it takes a very broadly based sample of texts. This is why Biber terms his study a 'Macroscopic,' (ibid: p337): His aim was to quantitatively study different styles, not for their own sake as in the present study, but to see what they revealed about the nature of the written/spoken medium of situational constraint.

Care must be taken in the use of the term macroscopic; in Niber's definition this is not necessarily connected to study size, but to the nature of the analysis, (ibid:p339). The nature of macroscopic stylistic study according to Biber (ibid) is:

'the identification of underlying textual dimensions in a set of texts, enabling an overall account of linguistic variation among those texts...'

whilst the nature of microscopic study is:

'... to pinpoint the exact communicative functions of individual linguistic features'

The present study too is macroscopic, in that it tries to identify

underlying textual dimensions, though in a rather more restricted set of texts than that used by Biber. It aims to do this by the use of Factor Analysis and a manual analysis of correlation co-efficients to produce a factor-like solution. However it also seeks to suggest the communicative functions of individual linguistic features, firstly by making a qualitative analysis of how features function to provide stylistic information on status, modality etc, and secondly to use statistical techniques to provide further supporting evidence.

It can be seen therefore that the aims of this study are much wider than that of Biber's, and in fact involve another area of research not acknowledged at all by Biber, which is a check on the homogeneity of text/style groupings by Cluster Analysis.

The fact that the aims of this study are much broader than that of Biber's could be considered a defect, in that too many aims may mean that none of them are really researched properly. The alternative view is taken here, that aims like Biber's are too narrow and, that an integrated picture of a particular style cannot be developed unless aims include both features which function to distinguish that style in particular, and those which show what stylistic dimensions it has in common with other styles. These observations also have to be based firmly on evidence which shows that texts sampled randomly from a group of texts, all supposedly written in the same style, really do form a group which is discrete, and are not simply

spurious style categories.

The sample of texts covered by Biber's study is composed of a written and a spoken sample, Biber (1985:p343). The written sample is, like that of Markworth & Baker's (1980) study, drawn from the Brown Corpus (Kucera and Francis 1979). It is composed of 408 texts, in 10 text categories of unequal sizes. The categories range from 'Press' to 'Professional letters', and include: 'Editorial letters, Skills and Hobbies, Popular Lore, Government Documents, Academic Prose, Belles lettres, General Fiction, and Romantic Fiction'. The written sample is heavily biased in favour of 'Academic Prose' and 'Belles Lettres', which together account for approximately two fifths of the sample.

The spoken corpus is taken from the London - Lund Corpus of Spoken English (Quirk and Svartvik (1980). It includes 6 text categories of unequal sizes and consists of 137 texts. The categories range from 'Face-to-face conversation' to 'Planned Speeches' and include; 'Telephone conversations, Interviews, Broadcast, and Spontaneous Speeches'. The sample is heavily biased in favour of 'Face-to-face conversation' which accounts for almost half the texts in the sample, Biber (ibid).

Whilst the large size of Biber's sample ought to have ensured greater statistical reliability, this advantage is very largely offset by:

a) the lack of attention paid to the definition of text categories,

and

b) the failure to ensure that the samples obtained were random samples from the written and spoken populations which Biber claims they represent.

and

c) obvious sample bias.

The labels applied to Biber's text categories are not further explained, and it is not at all clear for example how a speech was categorised as 'planned' or 'spontaneous' since planning is not necessarily manifested by the existence of written notes but could involve a speech being committed to memory. The heavy sample bias obvious in Biber's study severely undermines his assumption that they are representative samples of Written and Spoken English, Biber (1986).

In the present study, great care was taken to ensure that the populations of texts used were clearly defined and therefore open to check. Care was taken to standardise sample size, and to make true random and therefore representative samples from the populations identified. It is only by following this type of procedure that

stylo-statistical study can produce meaningful, and replicable results.

The aim of Biber was to isolate those dimensions involved in the differentiation of spoken and written texts, and to this end he used a factor analysis. Normal distributions underlying all his variables are tacitly assumed. The results of the K-S tests in this study show that normality cannot be assumed for all linguistic variables, when only 80 observations are made per variable. Biber may have circumvented this problem through use of a large sample size, where the central limit theorem might be presumed to operate. The Central Limit theorem says that as sample size increases, so does the likelihood that distributions underlying variables will be normal. Biber makes no checks on normality however and does not show that he is aware of this possible source of unreliability.

Biber performs only one analysis on his data, a Factor Analysis, and does not seek to show that his findings are reliable by supporting them with other techniques. In the present study many techniques with similar purposes were used to provide greater certainty that the findings produced were valid.

Biber decided on the number of factors to extract from his analysis using a scree plot. 5 factors were extracted, and rotated with a Promax rotation, but no information about the percentages of variance they account for are presented. Even so, it seems clear

from the scree plot provided by Biber (1985:p349) that only two really strong factors exist. Biber is unable to interpret the fifth factor, and further evidence for the weakness of the third and fourth factor is the scarcity of high loadings on them (only 3 out of 42 possible variables have loadings of above .35 on either factor).

Biber (1985:p354) calls his first factor 'Informational versus Edited text'. This interpretation is based on the negative loadings this factor has on syntactic features involved in 'packing informational content into a text', (ibid:p353) for example, 'prepositional phrases, type/token ratio and adjectives', and the positive correlations on variables showing 'a lack of concern with explicit informational content', (ibid:p354) for example, 'general hedges' and 'general emphatics'.

Biber's interpretation of his first factor is not wholly convincing; other positive loadings include, 'that' clauses, 'if' clauses, and 'wh' clauses which all would appear to have the same function the negative variables in increasing information bearing content. Biber does acknowledge this possible inconsistency, and reports evidence from one source that these variables are really characteristic of non-edited spoken texts. This problem exemplifies the subjectivity present in the interpretation of factor analyses by stylisticians. It is a problem from which the present study is not exempt either. Researchers must obviously be prepared to offer as much external

evidence as possible to support the interpretation of their factors

Biber's second factor is more clearly interpretable. It involves positive loadings on variables measuring; nominalization, passives and prepositions, and negative loadings on variables measuring time and place adverbs. Biber's interpretation of this factor as relating to abstract versus situational content seems intuitively correct.

Biber goes on to calculate factor scores for all his texts on the factors he obtains, but fails to make use of them in the way done by the present study to check up on the homogeneity of text categories. Given the vagueness of Biber's text categories referred to above, this would seem to have been an advisable thing for Biber to have done.

Biber does not report all his factor scores, but instead plots the mean factor scores for each of his text categories on one factor only, (ibid:p356). Biber loses a great deal of information in this way, and makes the rather questionable assumption that his text categories are homogeneous. What the plot does appear to show is that factor 1 does produce a fairly clear spoken/written text distinction.

To sum up Biber (1985) and Biber (1986), it seems that whilst they represent the most recent works in stylo-statistics, they are not at all well designed as far as sampling is concerned, and this makes an



assessment of the status of the findings produced difficult. In common with the vast majority of stylo-statistic studies, Biber makes no check on the homogeneity of the text samples used in his study and this in turn makes the validity of the labels used for these categories difficult to assess. A general difficulty with Factor analysis emerges when both Biber's and the present study are considered, this is the problem of the subjectivity involved in interpretation of the factors extracted.

The next section discusses the possible theoretical significance and practical implications of the present study.

#### 4.3 Theoretical Significance & Practical Implications of the findings

The main significance of this study is in the sphere of practical implications for the conducting of stylo-statistical research, and in contributing to specific knowledge about official writing in English, not in contributing to Style theory. This is because it did not set out to test a particular theory of style, but merely used one of the many theories available as a tool for practical analysis. Neither did the study set out to prove the superiority of a particular syntactic description over others as a tool in stylistic analysis.

This is not to say, however that the present study has no

significant contribution to make to the field of Style theory. The view adopted here was one of Style as situationally constrained language variation, and the suitability of this view is supported in the results of the study. That is, in finding styles that are not individually based but which can be related to specific non-linguistic situational dimensions. Evidence for at least two main dimensions in the data is provided in the results of the factor analysis. Both dimensions reveal different attributes of the text categories studied, for example Groups 1, 2 and 3 are very similar when considered on factor 1 but not on factor 2.

These dimensions appear to fit some of the categories described by Crystal and Davy (1969), namely; Modality, and Status. Apparent support for this is provided by the fact that factors 1 and 2 of the Factor Analysis referred to in the last paragraph appear to reflect previously (qualitatively) identified Modality and Status differences respectively. These two dimensions show that I.A. P.A. and BBSIL text groups share the same modality attributes, but that I.A. is more formal in status than either the other two.

However care must be taken in drawing the latter type of conclusion: The interpretation of the factors was made on a purely subjective basis, and they could perhaps be equally well interpreted differently, even within the method of description provided by Crystal and Davy (ibid); factor 1 may perhaps represent the general Province of Information giving, or factor 2 a 'personality'

dimension, not specifically identified by Crystal and Davy (ibid). but perhaps an element of Status. The results of the study must therefore be modestly appraised as providing further support for the Crystal and Davy's (ibid) particular stylistic description, but not conclusive proof that it is the only description possible.

What this study is able to do is to challenge the hitherto widely held assumption, tacit in linguistic and non-linguistic literature; Charrow (1982), Campbell and Holland (1982), Crystal and Davy (1981), Quirk (1982), Gowers (1962), Chase (1955), Bolinger (1980), Chiu (1973) and many others, that there is a Province constrained style which can be collectively referred to as 'officialese' or 'bureaucatese'. Even a sample of a well known type of 'officialese'; the leaflets of government departments, reveals two widely differing styles, one of which (labelled P.A. for convenience) has far more in common with the information leaflets put out by financial institutions for advertising purposes, than with the other 'official' style (I.A.) discovered.

The results of the present study have also important practical implications for future stylo-statistical studies. Firstly in terms of the type of statistics used. It appears that underlying normal distributions for linguistic style variables cannot be relied upon, as has been done in previous studies, and that by implication the use of statistical methods based on parametric statistical techniques in stylistic work cannot be relied on. Such techniques

should therefore either be discarded in favour of non-parametric techniques, or additional work done to provide support for the results of parametric techniques used.

The Parametric techniques used in the literature are frequently badly reported, neglecting to provide details on percentages of variance in Factor Analyses for example. This points to a possible lack of knowledge on the part of those using these techniques. Though to be fair, journal editors may be partly to blame for not allowing sufficient space. Non-parametric tests are generally easier to perform and understand than the equivalent Parametric tests, and are therefore more suitable in this respect as well.

Non parametric techniques are shown by this study to reveal results which are perfectly acceptable for stylistic work, and moreover, manual analyses of rank order based correlation and distance co-efficients are shown to yield results which compete well with those of Factor and Cluster Analysis.

The implication therefore is that non-parametric statistics could be used far more fruitfully in stylo-statistics.

A second practical <sup>a</sup>implication arising from the present study, is that it is perfectly possible for stylo-statistic studies to make a check on the suitability of previously determined text/style categories, this can be done using the Cluster Analysis technique, making a plot of factor scores, or more simply, by

a manual processing of (non-parametric) distance co-efficients. It is argued here that it is theoretically unsound to make statements about the attributes of a particular style without checking what texts make up that style. It cannot be assumed that previously determined text/style categories are necessarily the best way of partitioning the data.

The format of this study in undertaking an initial qualitative study, followed by a quantitative study to test the hypotheses generated about the nature of official writing and further explore its nature, has shown the value of qualitative analyses to be high. Several workers in Stylistics, among them Crystal and Davy(1969) and Sandell (1977), have maintained that qualitative research should not be looked on as inferior to quantitative research. This study has actually demonstrated the remarkable agreement which can exist between qualitative and quantitative assessments of what variables are stylistic in official writing, and therefore the tendency to accurate native speaker intuitions about what linguistic features are stylistic. Future researchers in Style should not therefore feel compelled to always provide quantitative support for their findings.

Unqualified praise cannot however be bestowed on qualitative analysis. The long-standing view quoted earlier that there is a single identifiable 'official' style has been shown in this study to be inaccurate. Two techniques of quantitative research were

extremely important in aiding this discovery; Factor & 'factor type' analyses, and Cluster & 'cluster type' analyses. Factor Analysis provided the basic information that the two supposed GIL 'sub styles' were indeed very different. Cluster analysis and 'cluster type' techniques added the information that whilst one of these 'sub-styles' (I.A.) was truly independent, the other (P.A.) was really closely related to non-official texts from private sector institutions. A hierarchical cluster analysis further clarified this relationship by showing that P.A. and BBSIL texts were both linked to the same branch of the hierarchy produced, whilst I.A. was not.

This practical implications of this information for future stylistic research appear to be that whilst significance tests like the Kruskal Wallis test or One way ANOVA, do little more than confirm what is already known about the behaviour of style variables in making distinctions, data reduction techniques like Factor and Cluster Analysis provide insight into patterns among variables and texts that might otherwise go undiscovered. It is these techniques, hitherto very little used in Stylistics, that are most likely to contribute most to future stylistic research. Not simply to further explore dimensions of situational constraint as Biber (1985) showed could be possible in his work with the spoken/written medium dimension, but in helping to clarify the existence of and relationships between actual styles.

The full possibilities of Factor Analysis in stylistic research are

demonstrated by the present study, which as well as examining patterns among variables by the extraction and interpretation of Factors, used Factor Scores to explore the homogeneity of the text categories studied and examine the relationships between them. Such a visual plot provides in addition a very clear representation of how patterns of variables differentiate the styles.

Whilst accepting Factor analysis as a very useful technique in stylistics, it is disturbing to note that in the present study only a relatively small proportion of the variance in the data is explainable by a strong factor structure. Comparison with other studies is hampered by the common failure to report percentages of variance alongside factors extracted. There is some small evidence however, which seems to imply (in Sandell (1977)), that selection of variables for study needs to be carefully made to select was is typical of the styles to be studied.

This section completes the Discussion. The Conclusion which follows seeks to summarise the findings of this study, highlight its shortcomings, and make suggestions for future research.

## CHAPTER 5 - PART 2 - CONCLUSION

This conclusion will sum up the work of the thesis by briefly restating its main findings and the final conclusions to be drawn from them. The shortcomings of the present study will then be discussed before concluding with a final section suggesting possible areas for further investigation.

### 5.1 Final Summary of Findings

The findings of the study can now be summarised under the following 11 points:

1. An initial qualitative study of one example of 'official writing': GIL, produced the hypothesis that GIL was indeed a distinct style, distinguished at all levels of syntactic structure, but particularly in terms of lack of NP premodifiers and the high frequency of complex NP postmodification, the use of conditional and concessive adverbial clauses, initial placing of adverbials, the use of Modal Verbs, especially 'MAY' which functions to express rights under the Law, and the non-use of past tense.

2. Further qualitative assessment of GIL revealed two sub-styles whose main contrasting feature was the use of the second person pronoun. This gave rise to the use of the labels of Impersonal and Personal Address to refer to these styles. These two sub styles also appeared to be distinct in other ways in particular in their use or



non-use of simple and non-simple sentences, passive verbs and unfamiliar vocabulary.

3. The Qualitative Study lead to the formulation of hypotheses to the effect that certain linguistic variables, including those mentioned above were significant in distinguishing between GIL and other styles and between GIL sub-styles. These variables along with others included for the sake of closure, were included in a set of 38 variables. It was proposed to score these variables on 20 I.A. and 20 P.A. texts, as well as 20 leaflets from the Private Sector and 20 Daily Mirror Editions.

4. Initial use of the Kolmogorov - Smirnov test for normality, showed that the hypothesis that each variable was drawn from a normally distributed population, for both individual and pooled group data, did not hold for large numbers of variables. The conclusion to be drawn from this was that the only appropriate tests for uniform use across all variables were non-parametric. Parametric tests if carried out could be possibly unreliable. A further conclusion is that normality cannot be relied on in linguistic data, as is often tacitly assumed by stylo-statistical studies. In spite of these conclusions, certain parametric techniques were used, merely as exploratory techniques from which no certain conclusions could be drawn.

5. The results of Kruskal-Wallis tests showed that nearly all of

the variables predicted as being significant in distinguishing GIL from other varieties were significant at the 5% level (though the number of variables participating in this distinction was rather small), and therefore stylistic. Exactly the same was discovered for variables where significant differences were predicted between I.A. and P.A. Here the contrasts involved a much larger number of variables. The conclusion to be drawn from this is that there is good evidence to suggest the existence of two stylistically distinct type of Government Information Leaflets. There is insufficient evidence to conclude that a single GIL style exists. This does not mean that such a style does not exist as only a sample of all possible linguistic variables was studied.

6. The results of a manual analysis of non-parametric correlation co-efficients showed no very strong correlations between variables, but a clear structuring into two groups of many of the variables involved in the strongest correlations. Closer examination of the variable descriptions revealed one group to be apparently concerned with STATUS and the other with PROVINCE and MODALITY contrasts. The conclusion must be drawn that it is the variables involved in these two groups that are the most important in making stylistic contrasts within the data.

7. The results of a 2 group factor analysis paralleled the results of the manual analysis of correlation co-efficients very closely and added further support to the conclusions of (6) above. A further

conclusion was drawn that the manual analysis produced results which were comparable with those of the factor analysis and theoretically more reliable in the present case in view of its avoidance of assumptions about population parameters.

8. A plot of the the factor scores calculated from the 2 group factor analysis revealed clearly that the previously established PROVINCE and MODALITY group of variables distinguish all leaflet varieties from Newspaper articles. STATUS variables, serve to distinguish Impersonally Addressed GIL from all other styles. Several conclusions can be drawn from this. The first is that an I.A. style is much more formal in style than other styles studied whilst predictably, the modality of a journalistic article and an information leaflet as well as the provinces of journalism and leaflet writing produce strong stylistic differentiation.

The second conclusion is that BBSIL and P.A. texts do not give rise to a distinct style but share a number of common features. I.A. texts are very clearly not part of this group in view of their much more formal nature. There is therefore no support for the commonly held view that Government produced texts have a distinct 'official' style.

9. The results of a manual analysis of distance co-efficients, support the conclusion immediately above: they show that the distance between texts from the BBSIL and P.A. groups is often very

small. At the same time they show I.A. and DMA texts to be highly independent styles, and confirm that these two styles are indeed represented by the texts supposed to have been written in those styles.

10. The results of a Ward's method Cluster analysis parallel those of the manual analysis of distance co-efficients closely. The hierarchical ordering of the 4 styles shows moreover that I.A. does not come under the same node in the hierarchical tree (dendrogram) as P.A. , further strengthening the conclusion that there is in fact no 'official' style.

11. A final analysis, discriminant function analysis, appeared to show that the initial classifications of texts into 4 distinct varieties can be fully substantiated through a 100% correct allocation of all texts to their original groupings. This result is at variance with other results suggesting confusion of P.A. and BBSIL texts. Bearing in mind the possible non-reliability of these results, only tentative conclusions can be drawn. The possible conclusion is that whilst the main stylistic variables (i.e. those participating in factor (or factor like) groupings, do not clearly distinguish P.A. and BBSIL , the remaining variables do make a P.A./BBSIL distinction possible. These variables probably are the basis for the weak third factor discovered when carrying out the factor analysis, (not further explored in this study).

## 5.2 Shortcomings of the present study

In an exploratory study of this type, attempting to outline the nature of a hitherto unresearched style, it is inevitable that the information obtained should be of a limited nature. Limited in the sense that no very firm conclusions can be drawn, due to the scarcity of hypotheses to be tested. These are shortcomings necessary in conducting any exploratory research however and should not be seen as defects per se. The final section of this chapter attempts to point out areas in which useful confirmatory research could be undertaken.

Much of the exploratory nature of this work was in part 'enforced' because many of the techniques desired would only provided reliable confirmation when applied to normally distributed data. Any future confirmatory study will have to address the problem of non-normally distributed variables in studies of this type before progress can be made. This point is taken up in detail in 5.3 below

Shortcomings of the type discussed above are not the only faults in this work, there are several other areas in which even an exploratory work of this type could possibly have obtained better results. The faults of Part 1 are those of neglecting the possibilities of investigating various linguistic variables. The faults of Part 2 are design faults and come under the headings of General Design Type, Choice of Variables, Choice of Texts for

## Analysis, and Statistical Techniques used.

### 5.2.1 Shortcomings of Part 1

The current interest areas in linguistics concerned with text structure, for example Genre Analysis (Dudley-Evans: 1986), were completely neglected as possible sources of linguistic variability. These might have yielded interesting stylistic information. On first inspection, methods of investigating textual structure, of which Genre Analysis seems the most pertinent, do seem to be attractive descriptive frameworks for the stylistician. This point is now further discussed.

The most attractive feature of Genre Analysis is that unlike previous methods of Text Analysis it does not concentrate on general features common to all texts but on the analysis of particular text types, e.g. Journal Articles, (Dudley-Evans: 1986 pl).

The work of Swales (1981) on article introductions, and that of Dudley-Evans (op cit) on the Discussion sections in MSc dissertations, reveals text structure to be composed of a series of 'moves'. For example the 4 moves discovered by Swales (1981:p22) to be involved in article introductions, summarised here as: 1. Establishing the Field, 2. Summarising Previous Research, 3. Preparing for Present Research, 4. Introducing Present Research.

These moves are shown to follow the order given, but may be realised differently in individual cases. For example move 3 may involve demonstrating that there is a gap in present knowledge, by raising questions, or by extending already known findings. Swales (op cit) indicates that moves are identifiable mainly by lexical clues.

Dudley-Evans (1986:p7) suggests a more detailed set of six moves for *thesis* introductions, which were produced after examining 7 MSc dissertations. Dudley-Evans suggests that whilst lexical items signal some moves, other signals are to be found in paragraph breaks but also particularly in Discourse structure, where the change from one move to another is marked by the change from the description of a situation to the response to it, (ibid:p8).

Examination of Government Leaflets tends to reveal a text structure which might be interpreted in terms of moves similar to those suggested by Dudley-Evans (1986). The expediency of carrying out a detailed analysis for the purpose of identifying stylistic variables however seems very much in doubt. The weakness of Genre Analysis from the stylistician's point of view is that moves cannot be given a clear linguistic definition as traditionally accepted syntactic categories can, for example clause, noun phrase, and so on. Even the markers of moves have to be referred to items which do not as yet have clear linguistic definitions, e.g. paragraph and discourse pattern.

items which cannot be objectively classified in linguistic terms cannot be readily studied in the way that items have been researched in this thesis. Furthermore any Genre Analysis based description of government leaflets would not be referable to an established description of 'moves' (at the time of writing at least) and would not be easily interpretable by other workers in the Stylistics field. It has to be concluded therefore that it was right to restrict part 1 to the analysis of syntactic structures at sentence level though further research of AdE using Genre Analysis as a model should undoubtedly be undertaken. Such a method of analysis is likely to reveal far more than the present study about how information leaflets function as information giving texts.

The other source of linguistic variability very much neglected in Part 1 was Lexis. Lexis has long been a point of interest for those comparing styles, and Crystal and Davy (1969) suggest it as a possible source of stylistic variation. However, Lexis is rather less attractive to the stylistician than syntactic patterns. This is because whilst stylistically marked syntactic patterns need careful research to establish, lexical markers are often extremely obvious even to the casual observer, for example it does not need a stylistician to classify a text containing: 'whereas' and 'hereinbefore' as a legal text.

Though no one chapter was specifically devoted to lexis in Part 1, observations were made on specific items when they appeared to be associated with particular syntactic structures. Individual modal



verbs were also considered.

It was this, together with the fact that a weighty amount of syntactic information had already been accumulated, and the knowledge that some work had already been done on lexis in Bureaucratic English Chiu (1973), that led to the decision not to investigate lexis further. There is no doubt however that the failure to include a section on lexis was an omission, which should not have been made had time allowed.

#### 5.2.2 General Design Type

While the Independent subjects design used in the second part of this study yielded much useful information, it failed to make use of the fact that many Government Leaflets are currently being re-written, in 'Plain English'. This fact is well reported in the Government White Paper: 'Administrative Forms in Government' (Cmd 8504). Such rewriting would have allowed the direct comparison of leaflets before and after revision to see if a shift to 'Plain English' affected the linguistic structure of texts, and if so how? Research of this type would have required the pairing of texts in Part 2 in a repeated measures design. Hypotheses for testing could have been generated by qualitative analysis of texts in Part 1.

A study like the one just suggested would have yielded information rather different to that of the present one, and would also seem to involve the tacit assumption that Government Leaflets formed a

uniform style. It seems right that the present ~~Exploratory~~ research in the style of Government Information Leaflets and how they relate to other styles should have come first.

Given that Government Information Leaflets were shown not to display a single style, in the present study, the problem of what 'Plain English' is stylistically becomes more complex. For example, can the rather more informal Personal Address Leaflets be re-written in 'Plain English', or are they perhaps already written in this style? As leaflets are being continuously revised anyway another difficulty would be in identifying leaflets not in Plain English.

The investigation of the relation of 'Plain English' to either Impersonal or Personal Address Government Leaflet styles was clearly too complex to tackle here. Some points of departure are however suggested in the last section of this chapter on possible areas for further study.

The reasons for adopting an observational rather than an experimental design were well set out in Chapter 2. And there does not appear to be any reason to regret this decision. However the information now collected about the stylistic nature of Government Leaflets could be used to write experimental texts. This point is taken up in 5.3 below.

### 5.2.3 Shortcomings in the choice of variables

One possible criticism of this study is that it considered mainly linguistic variables at sentence level and not beyond. This was not altogether true, sentence cohesion, whilst uninvestigated at Part 1 was used as a source of linguistic variables. Sentence Cohesion however merely describes the way in which sentences link to form text, and not textual structure. The reason for omitting models of text structure, e.g. Genre Analysis, in Part 1 has already been discussed (in 5.2.1). The same objections to the use of text analysis models hold for Part 2.

Apart from this it has to be admitted that the number of variables studied was quite small, and that some stylistic variables may have escaped notice. The initial selection of variables was made qualitatively, and it is difficult to see how this could have been avoided. But this subjective selection could quite easily have resulted in omissions, another investigator might have made a different selection.

Measures were taken to make sure that variables were not missed by including some variables for the sake of closure, e.g. including all sentence types as variables, not only those thought to be stylistic. Notwithstanding this the subjectivity involved in selecting variables means that the study cannot hope to claim that it is comprehensive; those variables which are clearly stylistic have been pinpointed, but unmeasured variables cannot be presumed to be

non-stylistic.

A more valid selection of variables could perhaps have been obtained by consulting the writers of government leaflets directly as to what they considered to be typical markers of official style. This approach was used by Sandell (1977), but would obviously have taken considerable time to do.

#### 5.2.4 Shortcomings in the choice of texts

There are two possible sources of criticism for the choice of texts made in this study, the first is that not enough different styles were represented, and the second that there was possible sample bias.

Time and resources only allowed the investigation of 4 possible styles, and because of this it has to be admitted that the present study fails to give a complete picture of how Government Information Leaflets relate to adjacent styles. It is rather difficult to decide what adjacent styles might be. In Dialectology, it is quite clear that adjacent dialects must be geographically related. Stylistic space is far more complicated and involves many dimensions according to Crystal and Davy (1969).

Any exhaustive stylistic comparison would have to have taken into account time (comparing current leaflets with older ones), discourse medium (comparing written leaflets with spoken material from

Government departments) as well as the dimensions of Modality and Status which were investigated here. Legal English, the English of Instructions, of Advertising, and of departmental correspondence might all have been usefully compared with the language of Government leaflets.

Such multiple comparisons would have been possible using the same statistics as were used in this study, but the counting of variables involved would have been immense. It is clear that comparisons like the ones suggested would be best done by computer on tagged corpuses. However none of the corpuses currently available, list texts of the appropriate types for analysis. Until a suitable corpus becomes available tasks like comparing Government Leaflets with a wide range of other varieties seem well nigh impossible.

It seems that practical problems of this type will either result in many small studies like the present one being done, or in stylisticians limiting themselves to the rather broad subject headings, as style categories, available on the major computer stored language corpuses, (cf Markworth and Baker (1980), Biber (1985)). The latter may result in an undesirable narrowing of attention on what is available rather than what is interesting, and is reminiscent of the drunk who having lost his doorkey whilst fumbling with it at the door of his house proceeds to look for it, not at that spot, but out in the street under a lamp, where he can see more clearly.

There are two sources of sample bias in this study one more serious in its potential effects than others. The first source is that of the Government Departments themselves. This source of bias could have effected both the I.A. and P.A. samples. No attempt was made to ensure that the departments from whom leaflets were requested sent an example of every leaflet in their stock. It is possible that some departments may have sent out only what they considered to be their 'best' leaflets, i.e. the clearest and easiest to understand.

Since it is not at all clear to what extent the style of Government leaflets and ease of understanding are related it is difficult to envisage the possible result of such bias if present. Any replication of the present study should be careful to try and ensure an 'uncensored' corpus of leaflets from every department.

It must also be remembered that not all leaflets received from Government Departments were sampled, only those which met certain criterion, namely presence or complete absence of the pronoun 'you' and the presence or absence of cover graphics. In the strict sense therefore the samples obtained were not representative of all Government Information Leaflets, but of two types which were conveniently referred to as Personal and Impersonal Address.

Another source of bias is the fact that the fourth sample was not a representative sample of Newspaper Articles but wholly composed of Daily Mirror articles. This bias is non-serious and can be defended on the grounds that the sample required was of 'typical' tabloid newspaper text rather than a representative sample.

### 5.2.5 Shortcomings in the use of Statistical Techniques

The most contentious statistical techniques used were Factor Analysis and Cluster Analysis. The sample size used was almost too small for a reliable Factor Analysis with only twice as many observations as texts. Some statisticians recommend far higher numbers of observations to ensure reliability. In spite of this possibility it has to be observed that the agreement between the factor solution obtained and the manual analysis of correlation co-efficients tends to suggest quite good reliability in fact.

A second problem with the Factor Analysis was the undesirably large amount of common variance which could not be accounted for by either of factors 1 or 2. This problem might well have been avoided if variables had first been selected which it was believed represented the same factors, as Biber (1985) did. Much more research needs to be done however on the application of Factor Analysis to stylo-statistical work, in order to ascertain whether problems like co-efficients tends to suggest quite good reliability in fact.

A second problem with the Factor Analysis was the undesirably large amount of common variance which could not be accounted for by either of factors 1 or 2. This problem might well have been avoided if variables had first been selected which it was believed represented the same factors, as Biber (1985) did. Much more research needs to be done however on the application of Factor Analysis to stylo-statistical work, in order to ascertain whether problems like

the one encountered here really are cases of factors being poorly represented through bad choice of variables, or whether in fact the case is that such solutions are the norm. This indicating perhaps that style variables do not conform as rigidly to a small number of situational constraints as is often supposed, (cf the constraints listed by Crystal and Davy (1969)).

The problem with the Cluster Analysis used is that while the Ward's method solution used was intuitively acceptable, and supported by the results of a manual clustering solution, only one other of the many other clustering solutions available was tried, and this (single linkage) proved totally unsuccessful. Because there is no established procedure for finding out which is the best clustering solution in a particular situation, far more solutions should have been tried in order to try and replicate the results obtained.

### 5.3 Suggestions for further research

The main aim of further research must be to provide confirmation of knowledge about English used in Government Documents. The knowledge gained by exploratory research in this preliminary study. That is further work should be conducted using fresh random samples from the same populations used here to see if the results already obtained can be replicated. This could greatly add to the strength of the claims made here about the styles described. There are also possibilities for any new work to further extend the knowledge obtained here in confirmatory ways, and this is discussed in the next paragraph.



New research on the English of Official Documents could extend knowledge both about the set of style variables that distinguish the two styles which are suggested by the present thesis, as well as testing to see how far membership of the style categories concerned extends to other document types. The first aim would involve setting up hypotheses about the ability of new linguistic variables to distinguish between I.A. and P.A.

A group of variables which ought to be investigated in this way are the modals, which yielded rather dissapointing results in the present study. Much finer classification of the modals according to the type of modality expressed e.g. Necessity/Possibility, and Deontic/Epistemic/Dynamic seems likely to yield better results than those already obtained. This conjecture is made on the basis of qualitative work already carried out. Factor Analysis should be used to discover what part if any new variables play in the factor structure already discovered.

Knowledge about the membership of the style categories discovered ought to be obtained by testing hypotheses about the ability of new texts sampled from other sources to score on the variables already measured in the same way as those in the established styles. This would involve further use of Clustering techniques. Hypotheses which could be tested in this way are that other forms of official language e.g. correspondence, press releases, questionnaire forms as well as legal texts do not cluster with one of the two Government

Styles already discovered, but are again independent. Such research might strengthen the argument that there is no one 'official' style or might show alternatively that new text groups blur the picture, overlapping with the styles already discovered.

Mention was made in section 5.2.2 above, of the possibility of comparing the styles established in this thesis and texts written in 'Plain English'. Plain English is a form of communication suggested by the 'Plain English Campaign' (see Cutts and Maher 1980) to be eminently suitable for official documents because it facilitates easier understanding of their contents.

There are two points of interest here. The first is the relationship between I.A. , P.A. and Plain English. This is a stylistic matter and ought to be one to be addressed in further research. The contention that Plain English is easier to understand than Government Documents not written in Plain English is not a stylistic matter, but a psycholinguistic one, and well beyond the scope of the present thesis. It is however one in which experimental work might be done, as envisaged in section 5.2.2.

Experimental work would entail the production of texts which could be made either 'official' (I.A. or P.A.) or Plain English. These might be used in trials to see if one type of text facilitated faster comprehension, the design of such an experiment would need to be very carefully thought out however. As some of the stylistic

parameters of I.A. and P.A. are now known, experimental texts could possibly be written to fit those parameters. It is difficult to see how experimental texts could be written in Plain English though without further research as to its stylistic parameters. It would of course be possible to test the hypothesis that P.A. texts are more quickly understood than I.A. texts.

Investigating

The stylistic relationship between Plain English and I.A./P.A. would not be an easy task, and some possible difficulties were discussed in section 5.2.2. The main problem is that of identifying a suitable population of Plain English texts. One possibility would be to take leaflets which had been rewritten by the Plain English Campaign itself, or by other pressure groups. A possible hypothesis for testing is that Plain English and P.A. are closely related in the way in which they contrast with I.A. . If a random sample of Plain English texts were to be tested on exactly the same variables as those used in this study, the hypothesis that P.A. and Plain English samples were in fact from the same population could be tested, using the Mann-Whitney U test, (unless the P.E. texts were paired with P.A. texts, then it would be the Wilcoxon test).

Three style groups I.A. , P.A. and Plain English could be tested using the Kruskal-Wallis test. A test ought additionally to be run to see if Plain English texts clustered with P.A. texts, or overlapped with the already established P.A. cluster. Cluster Analysis is not however a 'test' in the true sense of the word in that it does not test a hypothesis. Its investigation of cluster

patterns still falls into the realm of exploratory research.

This section on suggestions for further research ends the concluding chapter. The following pages contain the Bibliography and Appendices.

APPENDICES

APPENDIX A ( LIST OF ALL THE LEAFLETS USED IN EACH SAMPLE )

SAMPLE 1: PERSONAL ADDRESS LEAFLETS.

1. DHSS (April 1980) Attendance Allowance, Leaflet NI.205
2. Office of Population Censuses and Surveys (April 1974) Census Matters - a guide to the facts and figures;
3. DEPT of Transport (1983) The Orange Badge Scheme.
4. DHSS (April 1982) NHS dental treatment - What it costs and how to get free treatment, Leaflet D.11
5. Inland revenue (January 1982) Income tax and school leavers, Leaflet IR33
6. DHSS (Nov 1980) Family Income Supplement - More money each week for families on a low income, Leaflet FIS.1
7. Dept of the Environment (1982) Condensation and Mould Growth in your home,
8. DHSS (Nov 1982) NHS Prescriptions, Leaflet P.11
  
9. Dept of Education and Science (July 1983) On from A levels - Choose your course
10. DHSS (Nov 1982) Cash help, Leaflet SB.1
11. DHSS (Nov 80) Your retirement pension - if you are widowed or divorced, Leaflet NP. 32A
12. DHSS (August 1981) Child benefit for people entering Britain, Leaflet CH.5
13. Dept of the Environment, Thinking of bringing back AN ANIMAL OR SKIN OR SHELL from a holiday or visit abroad?
14. Dept of the Environment (October 1983) Housing Association Rents - A guide for housing associations and their tenants, Housing Booklet Number 13.
15. DHSS (July 1982) National insurance Unpaid and late contributions, Leaflet NI.48
16. DHSS (1984) Protect your health abroad - This leaflet tells you how, Leaflet SA.35
17. DHSS (May 1983) How to appeal against a decision made by a social security office or unemployment benefit office, Leaflet NI.246
18. DHSS (Dec 1980) Child benefit for childre away from home, Leaflet CH.4
19. DHSS (April 1983) Social security - School leavers and students - What you pay and what you get, Leaflet NP.12

20. Manpower Services Commission (Aug 1982) Getting Back to Work, Leaflet EPL 89.

SAMPLE 2: IMPERSONAL ADDRESS LEAFLETS.

21. DEPT of Employment (1982) Employment Rights of the Expectant Mother Leaflet PL710.
22. DEPT of Education and Science (1982/83) Business Studies.
23. DEPT of Employment (1982) Unfairly Dismissed ?, Leaflet PL712.
24. Customs and Excise (March 1983) Import and Export Clearance Procedures Leaflet 483
25. DEPT of Education and Science (April 1983) The DES A Brief Guide, Leaflet ISBN 0-85522-121-6.
26. DEPT of the Environment (1977) Local Plans: Public local inquiries.
27. Home Office (1981) British Citizenship, BN1.
28. DEPT of Employment (Nov 1980) Picketing
29. DEPT of Employment (1983) Offsetting pensions against redundancy payments, RPL 1.
30. DHSS (Nov 1983) Supplementary Benefit and Trade Disputes, SB.2.
31. Immigration and Nationality Dept. Admission to the United Kingdom, Leaflet RON2.
32. Dept of Employment (1982) Time Off for Public Duties Leaflet PL702.
33. DEPT of Employment (1983) Rights On Termination of Employment, Leaflet PL707(1st revision).
34. DEPT of Employment (1982) Employment Acts 1980 and 1982, Leaflet PL709.
35. Office of the Health Service a Commissioner for England Health Service Commissioner for England.

36. Home Office (July 1983) Information About Registration as a British Citizen, Leaflet BN8.
37. Court of Protection (Sept 1982) Mental Patients Possessed of Property, Leaflet PN1
38. H.M. Land Registry (Feb 1983) Notes for the Guidance of Persons Concerned with the Protection of Rights of Occupation Under the Matrimonial Homes Act 1967, Explanatory Leaflet NO:4.
39. H.M. Land Registry (May 1983) Application for First Registration Made by an Owner in Person, Explanatory Leaflet NO:2.
40. Dept of Employment (Feb 1976) The Fair Wages Resolution, Leaflet 30M.

SAMPLE 3; BBSIL Leaflets

41. Lloyds Bank (Sept 1985) Lobby Service, Leaflet PL42.
42. Barclays Bank (Feb 1985) Personal Bank Charges,
43. Midland Bank (Nov 1985) Budget Account, Leaflet 1781-9.
44. Nat West (June 1985) Car Loans from Nat West, Leaflet NWB3616.
45. TSB Bank (Nov 1984) Personal Loans, Leaflet TSB85/607.
46. National & Provincial BLDG Society Special Share Account, Leaflet U-550-2.
47. National & Provincial Bldg Society Savers Share Account, Leaflet U-551-0.
48. TSB Bank (April 1984) Cliff Michelmore Explores the World of Unit Trusts with TSB, Leaflet TCL 1342.
49. TSB Bank (May 1985) The TSB Service, Leaflet TSB85/699.
50. Barclays Bank (March 1985) Savings -- Higher Rate Deposit Account, Leaflet BB18249.



51. Barclays Bank (Oct 1985) Stocks and Shares for Beginners,  
Leaflet BB18113.
52. Halifax Bldg Soc. (Oct 1985) Endowment Mortgages,  
Leaflet 1/204500-3.
53. Halifax Bldg Soc. (Oct 1985) Halifax Budget Plan,  
Leaflet 1/305600-9.
54. Abbey National Bldg Soc. (Nov 1985) Abbey National  
Chequesave, Leaflet D48.
55. Abbey National Bldg Soc. (June 1985) Share Account,  
Leaflet D05.
56. Alliance and Leicester Bldg Soc. (Nov 1985) Gold Plus,  
Leaflet L/GOLD2.
57. National Westminster Insurance Services Ltd. (May 1984)  
Insurance or Assurance? Leaflet NWB3776.
58. Midland Bank (Nov 1985) Save and Borrow Account, Leaflet  
9205-9.
59. Lloyds Bank (Sept 1985) Homeloans, Leaflet PL22.
60. Lloyds Bank (Dec 1984) Black Horse Householders' Insurance,  
Leaflet PL16.

Note: Leaflet numbers refer to text numbers in the tables of  
appendix B (raw scores) on pages 454 - 457.

APPENDIX A: (TEXT SAMPLES)

I.A. text.....451 From text 3, pages 2 and 3 of  
'Admission to the United Kingdom'  
(RON2) Home Office: Dept of  
Immigration and Nationality.

P.A. text.....452 From text 10 'Cash Help'  
(SB.1) Dept of Health and Social  
Security.

BBSIL text .....453 From text 41 'Lobby Service'  
(PL42) Lloyds Bank plc.

visa nationals in the United Kingdom prior to departure subject and according to Home Office conditions of stay endorsed in the passport. Enquiries regarding re-entry visas may be made to any branch of the Passport Office in the United Kingdom and to British representatives overseas.

#### **b. Home Office Letters of Consent**

Foreign nationals who are not visa nationals but who nevertheless require entry clearance should hold a Home Office Letter of Consent. They, and any other non-visa nationals who wish to ascertain in advance whether they are eligible for admission to the United Kingdom can apply to the entry clearance officer in the country in which they are living for a Home Office Letter of Consent or an application may be made to the Home Office on their behalf by someone in the United Kingdom.

#### **c. Entry Certificates**

Commonwealth citizens who are required to hold entry clearance or who wish to ascertain in advance whether or not they are eligible for admission to the United Kingdom can apply to the entry clearance officer in the country in which they are living for the issue of an entry certificate.

3. An application for an entry clearance should be made well in advance of the proposed date of travel as, if the application is made abroad, it may have to be referred to the United Kingdom for a decision. The application should be supported by such evidence as is relevant in the circumstances. An overseas national who is in any doubt about the need for an entry clearance should consult the nearest British Embassy, Consulate or High Commission. The holder of an entry clearance can still be refused admission to the United Kingdom on such grounds as criminal record, medical condition, change of circumstances since the issue of the entry clearance or if it was fraudulently obtained.
4. People subject to deportation orders will also be refused entry. They may apply for a revocation of the order after their departure and may appeal against a decision not to revoke their order.

### **ENTRY FOR TEMPORARY PURPOSES**

#### **Visitors**

5. Visitors, including those coming to stay with relatives or friends, are admitted if they satisfy the immigration officer that they intend to stay only for the period stated, can maintain and accommodate themselves and their dependants (or will be maintained and accommodated by their hosts) without working or recourse to public funds, and can meet the cost of their return journey. Visitors are usually admitted for six months but this may be extended to up to 12 months if the immigration officer is satisfied that they can support themselves and their dependants for a longer period. A visitor is normally prohibited from taking employment and may not remain longer than 12 months.

#### **Visitors for private medical treatment**

6. Persons who wish to enter the United Kingdom for medical treatment have to satisfy the Immigration Officer that they have been accepted for consultation or treatment as private patients, and must produce documentary evidence that adequate funds are available to pay for the treatment and for maintenance and accommodation while undergoing it. *People suffering from communicable diseases may be refused entry.*

#### **Students**

7. Overseas students wishing to enter the United Kingdom for a full-time course of study at a university, a college of education or further education, an independent school or a bona fide private educational institution are advised to obtain an entry clearance which is issued on evidence that the applicant has been accepted for a course of study. Students must be able to meet the cost of their studies and of their own maintenance and accommodation, and that of their wife and children under 18 if they wish to bring them with them. Students are admitted for an appropriate period depending on the length of their course of study and must leave the country at the end of it. Students (except those undertaking postgraduate studies) are generally restricted from taking employment but they may seek the consent of the Department of Employment to take employment in their free time or during vacations. Permission is given only if there is no suitable resident labour. Dependants of students admitted with them are free to take employment unless the student himself is prohibited from taking employment.

#### **'Au Pair'**

8. 'Au Pair' is an arrangement under which an unmarried girl aged 17 to 27 inclusive, who is without dependants and is a national of a Western European country, Malta, Cyprus or Turkey, may come to the United Kingdom to learn the English language and live for a time as a member of an English-speaking family. She is normally expected to help with household tasks in return for pocket money. A girl may be admitted as an 'au pair' for a period of up to 12 months but is prohibited from taking employment or from staying on in the United Kingdom in some other capacity. She may be allowed an extension of stay up to a total period of two years if the 'au pair' arrangement is satisfactory. If she has previously spent time in the United Kingdom as an 'au pair' she may be admitted for a further period up to a maximum of two years altogether. A girl who entered the United Kingdom for some other purpose may also qualify to remain as an 'au pair' if she meets the 'au pair' requirements.

### **ENTRY FOR EMPLOYMENT OR BUSINESS OR AS PERSONS OF INDEPENDENT MEANS**

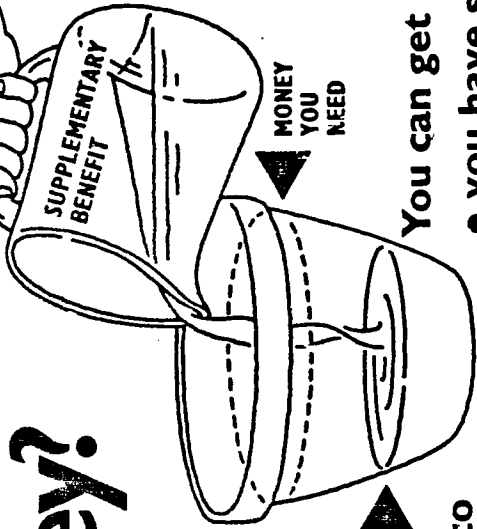
#### **Work Permits**

9. In general, people subject to immigration control coming to Britain for employment are required to hold work permits. Exceptions are certain permit-free categories, (see paragraph 12), Commonwealth citizens with a grandparent born in Britain, working holidaymakers, those intending to establish themselves as businessmen, or self-employed persons, and nationals of the European Community countries (see paragraph 27). Permits are issued for employment in England, Wales and Scotland by the Department of Employment and for employment in Northern Ireland by the Department of Manpower Services in Northern Ireland; the conditions for the issue of permits are the same throughout the United Kingdom. Work permits are available only for overseas workers holding recognised professional qualifications or possessing a high degree of skill or experience<sup>1</sup> and are issued by the employment departments for a particular job with a particular employer when there is no suitable worker in the United Kingdom or other European Community country. A work permit holder is admitted subject to a time limit on his or her stay and must obtain the approval of the department if he or she wishes to change jobs. Extensions of stay may be granted by the Home Office if the permit holder remains in approved employment. After four years permit holders can apply for the removal of the time limit on their stay. If the time limit is removed they are regarded as settled<sup>2</sup> in Britain and may then take any employment without obtaining approval. The wife of a permit holder and their children under 18 may be admitted for the same period as the head of the family, if he is able and willing to maintain and accommodate them without recourse to public

# Need more money?

If the money you have coming in is less than you need to live on,

MONEY YOU HAVE COMING IN

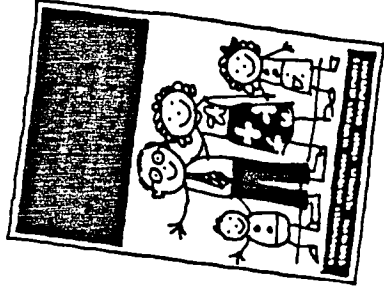


You can get it even if

- you have savings of up to £2,500.
- you own your own home
- you have not paid national insurance contributions.

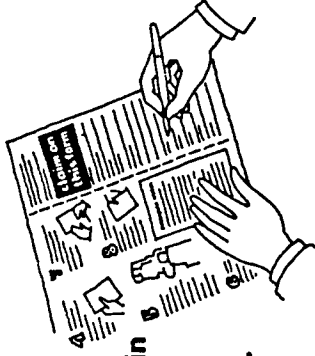
you may be able to get supplementary benefit. Claim even if you seem to have a little more than you need. Supplementary benefit is the difference between what you need and the money you have coming in.

You can't get supplementary benefit if you are in full-time work. But if you have children, you may be able to claim Family Income Supplement instead. Get this leaflet from a post office or social security office.



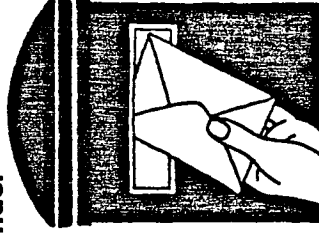
You may be able to get help with your rent and rates. Ask at your council offices.

## to claim

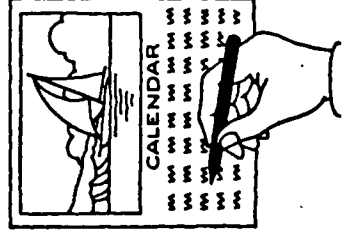


**1** Fill in the form on the right.

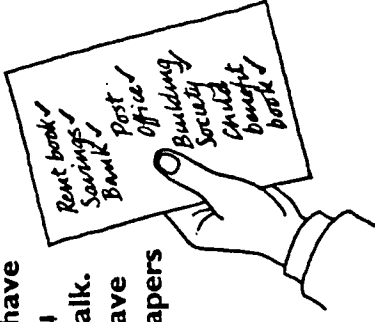
**2** Post the form to your social security office. You will find the address in the phone book under Health and Social Security, Department of. If you need money urgently, tell the clerk at the office.



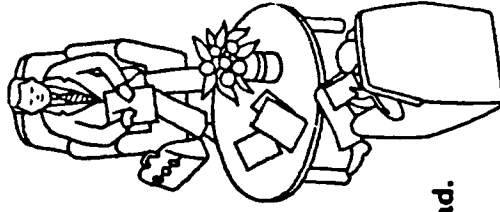
**3** You will have a talk with someone from the office to find out whether you can get supplementary benefit. The office will tell you when this will be.



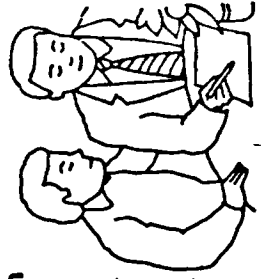
need to have with you for the talk. Try to have all the papers ready.



**5** The talk will usually be in your own home. If you are over pension age, and you don't want anyone to call, just tell the social security office. You can visit the office instead.



**6** The officer will write down what you tell him, and then check it with you. Tell him if he has not got it right.



## Your ATM automatically

# welcome, 17 hours a day.

The great thing about our LobbyService is that you can call in when you want to. Not just in normal banking hours. You can withdraw up to £100 in cash from your current or deposit account each day, deposit cash and cheques or request a new cheque book. All this in comfort, protected from the weather.

Our LobbyService provides a comprehensive and private facility based on our well-known and widely-used Cashpoint and Creditpoint services.

Indeed, you will need a Cashpoint card in order to gain entry to any LobbyService. As a customer, acquiring a Cashpoint card is as simple as completing the application form in this leaflet. There are no formalities such as a credit rating, unlike many other cash dispenser networks.

LobbyService, already at 170 locations, is open from 6.30am to 11.30pm, seven days a week (excluding Christmas Day). We are developing more of them all the time to create a countrywide network to which you can open the door 17 hours a day.

A handy, pocket-sized directory of LobbyService and other Cashpoint locations is supplied with each new Cashpoint card or you can obtain a copy at any UK branch of Lloyds Bank.



## Cashpoint – for instant cash

Cashpoint is the fast and easy way of getting cash. You just put your

Cashpoint card into the machine, tap in your personal number and you can withdraw up to £100 a day, providing your account can stand it. Unlike many other cash dispenser cards there is no pre-set weekly limit as to how much you may withdraw – even the daily maximum may be increased by special arrangement.

And you can also use Cashpoint to make withdrawals from a Deposit Account, with no loss of interest. It is an excellent way of saving and having direct access to your cash when you need it. Ask at your nearest branch for a leaflet and an application form to open an account.



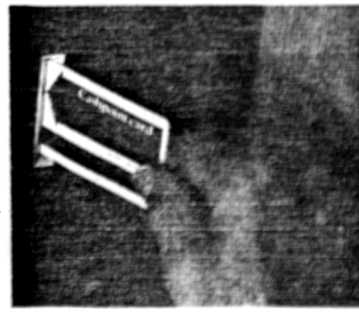
## Creditpoint – for instant paying-in

There's no waiting when it comes to paying-in cash or cheques, or when requesting a cheque book.

By using a Bank Giro credit slip, Lloyds Bank customers can also pay into their accounts quickly at over 1000 Creditpoints around the country.

Creditpoint takes care of it all. Paying-in slips and special envelopes are provided. You simply fill in the slip in the usual way, put the money in the envelope and place it in the Creditpoint.

Apart from being remarkably convenient, you can use this service with confidence because envelopes are opened and checked in the presence of two members of our staff.



APPENDIX B

TABLES OF RAW SCORES (BY GROUPS)

VARIABLES (GROUP 1 -- PA)

GP	TXT	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
1	1	4	13	2	1	4	16	3	5	6	7	10	4	4	5	1	5	0	2	2	0
1	2	3	12	5	0	1	12	5	5	2	3	2	4	11	2	1	9	1	10	10	0
1	3	9	11	0	0	3	9	2	5	4	9	11	4	2	5	3	3	0	2	2	0
1	4	7	8	3	1	1	14	1	2	7	5	9	6	4	4	2	5	0	0	0	0
1	5	8	9	3	0	1	13	0	3	3	5	7	7	4	6	1	4	0	3	2	1
1	6	7	10	2	1	1	14	3	6	5	7	6	3	3	3	0	2	2	3	3	1
1	7	2	17	1	1	11	20	0	3	3	4	11	1	5	4	0	1	2	2	2	0
1	8	5	11	2	0	1	19	0	0	2	4	10	2	12	6	2	5	0	2	2	0
1	9	4	11	5	0	0	11	1	1	3	8	9	4	6	7	1	1	1	1	0	2
1	10	5	11	2	1	9	15	2	2	4	15	12	3	3	7	4	1	0	0	0	0
1	11	7	10	2	1	3	16	0	2	3	2	10	3	8	10	2	2	1	2	2	7
1	12	11	6	1	3	0	18	1	4	5	4	8	7	1	4	1	1	1	1	1	2
1	13	5	8	4	1	1	11	0	5	2	6	9	5	6	11	4	1	3	3	3	0
1	14	5	14	1	1	1	12	2	5	8	9	5	5	5	10	4	3	1	7	4	1
1	15	6	13	1	5	5	17	0	7	8	8	9	5	3	5	1	1	0	4	4	1
1	16	1	16	4	0	16	11	0	8	4	13	8	0	20	5	3	2	0	3	2	1
1	17	10	10	0	0	3	14	5	0	6	10	8	0	3	4	2	2	0	4	2	1
1	18	15	5	0	0	2	17	0	4	9	6	4	8	3	1	0	3	1	3	3	0
1	19	11	7	2	1	0	14	1	4	7	2	8	2	3	2	1	2	2	0	0	0
1	20	5	10	3	2	1	12	1	2	2	13	15	2	8	3	3	2	1	0	0	0

VARIABLES (GROUP 1 - PA)

GP	TXT	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38
1	1	4	7	6	3	5	5	2	4,0	21,6	5,8	4	40	10	16	5	0	3	45,6
1	2	3	5	0	2	12	4	1	6,2	16,0	8,6	5	25	6	18	1	6	0	53,2
1	3	0	6	10	5	1	3	1	6,4	33,0	9,1	6	38	15	18	2	2	0	44,0
1	4	1	2	2	2	7	8	1	4,6	26,5	5,6	4	25	6	7	7	0	5	50,8
1	5	1	2	0	2	7	11	0	8,3	26,0	9,1	4	30	9	18	3	0	0	50,0
1	6	4	3	0	6	5	9	1	5,3	17,0	6,3	3	28	15	14	4	0	0	43,6
1	7	1	4	0	5	12	1	1	2,4	25,0	10,0	2	36	13	13	6	1	3	43,6
1	8	0	3	6	2	10	1	1	3,1	23,6	8,2	3	36	15	9	7	3	2	42,8
1	9	4	6	0	6	3	10	1	6,6	27,0	10,1	6	29	14	8	5	0	2	38,4
1	10	3	7	0	3	8	10	0	1,3	33,0	3,5	3	30	23	8	3	0	1	44,4
1	11	5	9	0	1	5	12	2	5,7	21,8	6,6	5	40	13	13	7	1	6	43,2
1	12	2	9	0	4	6	10	1	7,0	34,6	5,9	1	27	10	7	6	0	4	39,6
1	13	1	4	1	4	4	7	4	3,8	18,8	5,2	8	25	10	9	2	1	4	41,2
1	14	0	4	4	3	1	11	1	5,9	12,0	6,6	4	27	13	9	2	1	1	51,2
1	15	6	7	7	2	3	9	0	4,7	25,0	6,6	5	30	17	14	3	0	0	44,8
1	16	3	7	0	6	9	3	2	4,3	17,5	10,4	4	30	15	10	5	0	1	36,0
1	17	2	5	0	3	9	8	2	5,5	22,2	7,5	4	26	13	8	5	0	0	44,8
1	18	0	3	1	1	7	10	0	4,0	30,0	7,0	3	31	12	13	5	1	0	41,6
1	19	0	5	2	10	6	1	0	5,9	19,8	6,1	4	34	15	11	4	1	2	38,0
1	20	4	4	0	7	11	2	0	4,9	23,4	10,6	5	27	15	6	2	0	4	37,0

TABLE OF RAW SCORES BY GROUPS (CONT)

VARIABLES GROUP 2 - IA

GP	TXT	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
2	21	11	5	0	0	5	17	4	7	9	7	5	4	3	5	7	8	2	8	17	0
2	22	10	9	3	0	0	18	3	6	5	11	8	7	7	5	10	10	3	6	14	0
2	23	11	4	5	6	2	13	0	6	9	2	6	4	13	15	10	6	0	7	17	2
2	24	11	10	4	1	2	5	6	5	6	0	2	5	6	5	8	4	0	10	20	0
2	25	6	11	3	1	3	4	0	4	9	9	6	4	5	3	12	9	3	5	11	0
2	26	4	7	5	1	4	5	2	12	6	2	4	6	6	6	8	1	2	6	15	1
2	27	10	8	3	5	1	20	5	4	7	10	2	5	5	11	8	6	3	7	16	3
2	28	8	7	2	3	1	9	7	2	5	0	9	4	9	5	9	8	0	8	14	0
2	29	10	4	3	0	5	8	4	6	5	4	7	7	5	3	7	5	1	11	14	2
2	30	16	5	3	4	4	10	7	5	3	1	2	4	6	11	12	2	0	12	20	0
2	31	6	4	4	2	3	14	4	7	0	3	4	6	8	7	5	6	2	13	20	0
2	32	11	9	2	1	0	16	7	5	7	4	2	6	12	6	7	3	4	11	16	0
2	33	8	6	2	0	8	19	3	0	7	10	6	7	13	8	6	8	4	15	18	0
2	34	6	12	2	0	3	10	8	7	9	2	7	4	4	2	3	5	9	7	20	8
2	35	5	7	1	3	4	11	5	8	7	2	6	4	10	13	15	11	0	1	16	1
2	36	9	5	2	3	6	13	6	5	2	15	5	5	13	3	1	10	3	15	20	3
2	37	7	4	3	4	2	12	8	5	6	3	6	5	4	11	6	1	0	11	16	0
2	38	9	3	3	4	3	11	5	6	9	2	3	1	11	1	10	13	1	11	19	0
2	39	6	5	2	3	4	10	4	7	9	2	2	3	10	7	8	15	1	13	18	0
2	40	5	9	4	0	2	15	6	6	11	3	1	4	10	6	12	11	0	12	12	1

VARIABLES (GROUP 2 - IA)

GP	TXT	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38
2	21	5	7	4	3	1	8	5	9,5	42,0	11,2	9	23	4	7	9	0	3	47,2
2	22	0	10	9	2	8	5	1	16,4	27,0	11,6	11	12	19	37	1	2	3	47,6
2	23	8	7	1	4	8	9	0	9,7	37,0	12,0	6	24	6	12	3	1	2	47,2
2	24	2	4	3	6	9	1	0	7,5	71,0	10,1	9	28	12	14	2	0	0	46,0
2	25	1	1	0	1	1	0	1	7,0	53,0	9,8	9	21	12	7	0	1	1	44,8
2	26	2	4	6	5	1	7	1	7,3	65,0	8,2	10	22	8	10	4	0	0	51,2
2	27	1	8	3	2	3	11	0	7,7	57,0	8,8	4	37	14	16	3	0	4	44,4
2	28	4	4	5	11	2	2	0	10,5	40,0	7,6	11	20	13	7	0	0	0	49,2
2	29	0	10	5	5	8	2	0	8,7	38,0	10,2	9	31	13	16	1	0	1	53,6
2	30	0	5	0	2	1	6	11	11,6	56,0	6,2	5	27	14	7	5	0	1	40,8
2	31	3	4	5	13	2	0	0	11,8	84,0	13,4	12	31	9	15	3	1	3	35,6
2	32	1	9	0	10	0	10	0	7,4	40,0	8,4	12	6	25	9	3	0	1	46,0
2	33	4	4	1	5	11	3	1	10,8	58,0	7,2	8	11	22	8	2	0	1	53,2
2	34	4	4	0	0	1	4	0	5,4	40,0	12,0	9	25	4	14	8	0	0	38,4
2	35	3	6	4	5	5	4	1	9,1	48,0	9,8	11	32	13	14	3	0	2	40,8
2	36	3	3	1	3	0	3	0	5,1	63,0	7,2	7	23	9	11	3	0	0	53,6
2	37	0	5	1	5	7	1	3	7,2	37,0	20,6	10	28	12	6	5	1	4	40,4
2	38	2	6	1	5	3	6	4	12,9	53,0	20,4	11	26	8	9	7	0	2	42,0
2	39	2	6	2	5	4	3	5	10,6	50,0	19,3	10	26	9	13	3	0	1	45,4
2	40	0	5	6	0	1	3	13	5,9	46,0	7,5	9	25	14	7	2	0	2	46,0

TABLE OF RAW SCORES BY GROUP (CONT)

VARIABLES (GROUP 3 - BBSIL)

GP TXT	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
3 41	6	6	2	2	5	4	1	10	2	2	5	9	6	3	3	0	3	4	18	0
3 42	13	5	2	0	0	12	1	2	0	4	9	5	1	5	0	0	0	0	16	3
3 43	4	8	3	2	3	10	2	5	1	0	5	7	1	5	0	0	0	0	18	0
3 44	3	7	5	3	2	12	0	4	2	4	5	8	5	12	5	0	0	3	18	0
3 45	7	8	3	1	1	4	2	3	1	0	10	2	1	9	0	0	0	0	18	0
3 46	5	7	2	2	4	10	1	5	4	0	6	7	1	4	0	1	3	4	16	0
3 47	6	6	2	1	5	8	1	10	1	0	5	4	2	6	0	0	4	8	20	0
3 48	7	10	2	1	0	0	3	0	0	1	8	7	6	6	2	0	0	0	14	4
3 49	3	11	2	0	4	6	1	0	1	3	3	10	9	9	4	0	0	2	18	0
3 50	5	11	3	0	1	14	1	2	2	2	11	3	6	10	3	0	2	2	20	0
3 51	4	10	0	0	6	12	2	2	0	6	9	4	4	10	5	0	1	0	20	3
3 52	9	6	2	0	3	10	1	5	1	3	4	8	8	5	5	1	0	4	20	0
3 53	5	10	2	1	2	7	1	7	2	4	4	11	5	4	0	1	1	6	20	0
3 54	5	7	3	1	4	6	0	0	0	4	5	6	6	6	0	0	0	0	20	0
3 55	5	10	2	1	2	10	2	4	4	2	7	8	5	8	4	0	0	1	20	0
3 56	6	8	1	0	3	6	3	3	3	1	8	2	7	3	2	0	3	6	20	0
3 57	4	9	4	2	1	10	0	2	3	2	2	3	4	10	0	0	0	2	20	0
3 58	6	6	2	2	4	10	5	3	0	2	8	6	1	0	0	0	0	0	20	0
3 59	3	6	3	2	6	10	2	0	0	4	11	1	4	0	13	1	0	2	18	0
3 60	7	6	1	0	6	4	0	3	2	2	2	2	8	6	17	2	0	9	18	0

VARIABLES (GROUP 3 - BBSIL)

GP TXT	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38
3 41	2	5	0	1	15	4	0	5,2	24,3	7,4	4	26	10	13	2	0	1	51,2
3 42	3	3	0	3	8	9	0	2,3	19,5	12,1	5	25	15	8	1	1	0	56,0
3 43	0	4	0	0	8	10	0	5,3	16,0	10,4	8	29	11	7	9	0	0	47,2
3 44	5	4	0	0	9	9	1	3,4	23,6	10,5	4	29	16	7	5	0	2	51,2
3 45	2	6	0	3	7	10	0	10,1	18,8	7,8	5	23	15	7	2	0	5	47,2
3 46	1	5	0	0	10	7	3	3,6	21,7	6,5	4	49	21	19	3	0	1	46,4
3 47	0	8	0	0	12	6	2	7,2	26,8	6,8	6	27	7	11	9	0	0	45,2
3 48	0	3	0	0	4	12	3	4,4	20,2	7,3	4	28	14	12	0	0	2	52,4
3 49	3	3	0	4	11	4	1	6,6	44,5	9,4	6	30	9	19	0	0	2	56,8
3 50	0	4	1	0	8	13	0	3,9	27,6	12,6	4	24	9	10	0	2	2	48,4
3 51	0	3	0	0	7	13	0	3,2	18,4	12,4	10	26	12	12	0	0	2	48,4
3 52	5	2	0	4	2	14	0	4,7	33,8	7,8	8	34	14	15	2	0	3	45,6
3 53	2	1	0	2	6	12	0	5,6	32,8	7,8	2	35	19	12	7	0	2	46,0
3 54	4	1	0	2	17	4	0	5,9	28,2	7,2	1	35	16	17	1	0	3	45,2
3 55	3	7	1	2	6	11	0	5,2	26,4	13,8	2	35	12	16	2	0	5	45,2
3 56	2	4	0	1	7	8	4	3,7	35,6	7,8	2	24	8	11	6	0	1	48,0
3 57	2	5	0	4	5	8	2	4,1	40,0	8,2	7	24	10	11	1	0	2	52,4
3 58	8	4	0	0	7	12	1	6,4	19,1	10,0	2	24	12	9	3	0	2	49,6
3 59	5	5	1	0	5	10	3	4,8	29,2	5,2	4	29	11	13	4	1	0	50,4
3 60	3	2	1	1	8	8	2	2,6	43,0	5,6	5	31	11	10	10	0	0	51,2



TABLE OF RAW SCORES BY GROUPS (CONT)

		VARIABLES (GROUP 4 - DMA)																			
GP	TXT	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
4	61	3	8	6	4	0	1	1	6	1	0	4	6	5	4	5	0	2	1	1	3
4	62	4	11	3	1	0	1	1	2	0	0	5	5	4	5	6	1	0	2	5	6
4	63	5	5	8	4	0	0	4	1	3	0	6	6	6	6	4	0	2	5	2	10
4	64	4	14	2	0	0	2	1	6	0	5	0	12	5	7	6	0	0	2	7	8
4	65	5	11	3	1	1	1	3	5	6	2	5	11	4	1	4	0	0	10	1	9
4	66	8	10	2	0	0	3	0	1	2	0	4	7	10	6	6	0	0	5	10	8
4	67	4	14	2	0	0	0	3	4	1	2	5	9	2	6	6	0	0	3	4	4
4	68	2	14	4	0	0	1	6	2	2	0	1	6	8	13	5	0	0	3	3	17
4	69	5	9	6	1	0	0	1	1	3	3	5	5	5	5	5	0	0	3	1	15
4	70	6	10	4	0	0	0	1	5	1	1	3	7	2	6	5	0	0	2	3	15
4	71	8	4	6	2	0	2	0	3	5	1	5	11	9	5	6	0	1	1	2	10
4	72	4	9	3	4	0	1	3	1	1	1	5	7	3	8	4	0	0	1	1	4
4	73	2	14	3	1	0	3	1	6	2	0	4	6	7	6	6	0	0	0	1	8
4	74	6	8	5	1	0	0	1	2	2	4	5	9	4	3	8	0	2	3	0	11
4	75	7	6	7	0	0	4	4	5	1	2	1	10	3	7	8	1	0	1	3	7
4	76	3	11	4	1	1	1	1	6	3	0	3	5	5	4	10	0	0	0	0	1
4	77	5	12	3	1	0	2	3	2	1	0	6	7	5	4	7	0	0	1	0	13
4	78	8	12	3	1	0	2	1	3	2	3	2	3	3	5	8	0	1	2	1	2
4	79	7	9	4	0	0	0	6	1	0	2	3	10	3	6	7	0	0	2	2	6
4	80	4	11	4	1	0	1	1	6	2	0	3	9	4	5	9	1	0	0	1	13

		VARIABLES (GROUP 4 - DMA)																	
GP	TXT	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38
4	61	3	1	0	2	2	15	1	2,4	20,5	3,8	8	27	15	10	1	5	1	0
4	62	1	1	0	5	7	6	2	3,2	24,0	4,3	3	27	10	5	1	6	5	0
4	63	3	1	0	0	9	7	4	2,5	23,3	4,4	5	28	19	4	1	2	2	0
4	64	4	0	0	0	3	16	1	1,9	26,3	3,5	5	27	10	10	0	2	5	0
4	65	5	0	2	2	6	10	1	2,7	20,5	5,4	5	24	15	6	1	1	1	0
4	66	5	2	1	3	5	10	0	2,6	21,7	4,1	4	29	22	4	0	3	1	0
4	67	1	1	0	2	6	8	4	1,9	25,3	3,8	10	28	20	7	0	1	1	0
4	68	3	0	1	1	1	16	1	2,6	31,2	4,0	6	23	17	4	1	1	0	0
4	69	2	0	0	0	3	15	2	1,9	26,4	3,6	3	23	16	5	0	2	0	0
4	70	3	0	1	1	3	15	0	1,8	17,5	4,0	1	24	14	9	0	1	1	0
4	71	6	1	1	2	4	12	1	2,1	23,2	4,0	4	21	16	2	0	2	1	0
4	72	6	4	1	1	5	13	0	3,0	20,7	4,3	2	23	15	6	0	2	0	0
4	73	3	2	0	1	5	11	2	3,1	33,7	4,4	10	26	15	7	0	3	1	0
4	74	6	0	1	3	6	10	0	2,0	27,6	3,8	5	32	23	3	1	3	2	0
4	75	4	2	3	0	4	9	4	3,4	23,1	5,6	6	24	16	13	2	3	1	0
4	76	12	1	2	1	5	10	2	2,3	26,0	4,1	8	29	11	13	2	3	0	0
4	77	2	0	1	0	6	11	2	2,3	26,2	4,2	4	35	18	13	0	3	1	0
4	78	4	8	0	1	7	11	1	2,0	13,7	3,5	5	25	15	3	4	1	1	0
4	79	2	1	0	0	7	11	2	1,8	23,6	3,5	3	28	13	12	1	0	2	0
4	80	1	1	1	0	10	9	1	2,0	18,4	4,0	5	35	22	6	2	2	3	0

APPENDIX C

ALL GROUPS

VAR.	DISTRIBUTION TESTED	AGAINST D1	PROB
3	NORMAL	.1539	0.045*
4	NORMAL	.9462E-01	0.471*
5	NORMAL	.1842	0.009*
6	NORMAL	.2823	0.000*
7	NORMAL	.1984	0.004*
8	NORMAL	.1336	0.115*
9	NORMAL	.2234	0.001*
10	NORMAL	.1211	0.192*
11	NORMAL	.1766	0.014*
12	NORMAL	.2034	0.003*
13	NORMAL	.1317	0.125*
14	NORMAL	.1022	0.374*
15	NORMAL	.1727	0.017*
16	NORMAL	.1838	0.009*
17	NORMAL	.1057	0.333*
18	NORMAL	.2579	0.000*
19	NORMAL	.2953	0.000*
20	NORMAL	.2057	0.002*
21	NORMAL	.2198	0.001*
22	NORMAL	.2822	0.000*
23	NORMAL	.1317	0.125*
24	NORMAL	.1106	0.282*
25	NORMAL	.3148	0.000*
26	NORMAL	.1733	0.016*
27	NORMAL	.7841E-01	0.709*
28	NORMAL	.1188	0.209*
29	NORMAL	.2541	0.000*
30	NORMAL	.1093	0.294*
31	NORMAL	.1897	0.006*
32	NORMAL	.1168	0.225*
33	NORMAL	.1899	0.006*
34	NORMAL	.1363	0.102*
35	NORMAL	.1228	0.179*
36	NORMAL	.9328E-01	0.489*
37	NORMAL	.1650	0.026*
38	NORMAL	.2987	0.000*
39	NORMAL	.2046	0.002*
40	NORMAL	.2774	0.000*

THE HYPOTHESIS THAT THE SAMPLE IS FROM THE SPECIFIED DISTRIBUTION CAN BE REJECTED WITH THE INDICATED PROBABILITY (PROB) OF BEING INCORRECT.

\*\*\*\*\* KULNOGOROV-SMIRNOV TEST \*\*\*\*\*

GROUP 1

GROUP 2

DISTRIBUTION TESTED AGAINST				VAR.	DISTRIBUTION TESTED AGAINST			
VAR.		D1	PROB			D1	PROB	
3	NORMAL	.1700	0.610*	3	NORMAL	.1500	0.759*	
4	NORMAL	.1481	0.772*	4	NORMAL	.1920	0.453*	
5	NORMAL	.1890	0.472*	5	NORMAL	.1859	0.494*	
-6	NORMAL	.3338	0.023*	6	NORMAL	.2092	0.345*	
-7	NORMAL	.2690	0.111*	7	NORMAL	.1261	0.908*	
8	NORMAL	.1335	0.868*	8	NORMAL	.8563E-01	0.999*	
-9	NORMAL	.2367	0.213*	9	NORMAL	.1314	0.880*	
10	NORMAL	.1318	0.878*	10	NORMAL	.1919	0.453*	
11	NORMAL	.1679	0.626*	11	NORMAL	.1658	0.641*	
12	NORMAL	.1069	0.976*	12	NORMAL	.2567	0.143*	
13	NORMAL	.1746	0.576*	13	NORMAL	.1730	0.588*	
14	NORMAL	.1057	0.979*	14	NORMAL	.2064	0.361*	
15	NORMAL	.2230	0.273*	15	NORMAL	.1763	0.563*	
16	NORMAL	.1792	0.542*	16	NORMAL	.1679	0.626*	
17	NORMAL	.2275	0.252*	17	NORMAL	.1244	0.916*	
-18	NORMAL	.2446	0.183*	18	NORMAL	.1109	0.966*	
-19	NORMAL	.2645	0.122*	19	NORMAL	.1937	0.441*	
20	NORMAL	.2348	0.220*	20	NORMAL	.1664	0.637*	
21	NORMAL	.2356	0.217*	21	NORMAL	.1377	0.842*	
22	NORMAL	.3126	0.040*	22	NORMAL	.3066	0.047*	
23	NORMAL	.1883	0.477*	23	NORMAL	.1480	0.774*	
-24	NORMAL	.1498	0.761*	24	NORMAL	.1521	0.744*	
25	NORMAL	.2922	0.066*	25	NORMAL	.2150	0.313*	
26	NORMAL	.1941	0.438*	26	NORMAL	.2537	0.152*	
27	NORMAL	.9577E-01	0.993*	27	NORMAL	.1987	0.409*	
28	NORMAL	.1777	0.553*	28	NORMAL	.1659	0.641*	
29	NORMAL	.2700	0.108*	29	NORMAL	.3363	0.022*	
30	NORMAL	.8408E-01	0.999*	30	NORMAL	.1429	0.809*	
31	NORMAL	.9114E-01	0.996*	31	NORMAL	.1269	0.904*	
32	NORMAL	.1646	0.651*	32	NORMAL	.2155	0.311*	
33	NORMAL	.1890	0.472*	33	NORMAL	.2319	0.233*	
34	NORMAL	.2062	0.363*	34	NORMAL	.1518	0.746*	
35	NORMAL	.1976	0.415*	35	NORMAL	.2058	0.365*	
36	NORMAL	.1830	0.515*	36	NORMAL	.2258	0.260*	
37	NORMAL	.1647	0.650*	37	NORMAL	.2572	0.142*	
38	NORMAL	.3091	0.044*	38	NORMAL	.4503	0.001*	
-39	NORMAL	.1892	0.471*	39	NORMAL	.2119	0.330*	
-40	NORMAL	.1584	0.697*	40	NORMAL	.1004	0.988*	

THE HYPOTHESIS THAT THE SAMPLE IS FROM THE SPECIFIED DISTRIBUTION CAN BE REJECTED WITH THE INDICATED PROBABILITY (PROB) OF BEING INCORRECT.

GROUP 3

GROUP 4

VAR.	DISTRIBUTION TESTED AGAINST	DISTRIBUTION TESTED AGAINST	
		D1	PROB
3	NORMAL	.2188	0.323*
4	NORMAL	.1722	0.626*
5	NORMAL	.2669	0.134*
6	NORMAL	.1987	0.441*
7	NORMAL	.1343	0.883*
8	NORMAL	.2489	0.190*
9	NORMAL	.2460	0.200*
10	NORMAL	.1541	0.757*
11	NORMAL	.2017	0.422*
12	NORMAL	.1512	0.778*
13	NORMAL	.1963	0.457*
14	NORMAL	.1302	0.904*
15	NORMAL	.1673	0.662*
16	NORMAL	.1378	0.864*
17	NORMAL	.2452	0.203*
18	NORMAL	.4819	0.000*
19	NORMAL	.3747	0.010*
20	NORMAL	.1961	0.458*
21	NORMAL	.3065	0.056*
22	NORMAL	.5036	0.000*
23	NORMAL	.1644	0.684*
24	NORMAL	.1432	0.831*
25	NORMAL	.5054	0.000*
26	NORMAL	.2818	0.098*
27	NORMAL	.1957	0.461*
28	NORMAL	.1226	0.938*
29	NORMAL	.3080	0.054*
30	NORMAL	.1237	0.933*
31	NORMAL	.1106	0.974*
32	NORMAL	.2182	0.326*
33	NORMAL	.1824	0.552*
34	NORMAL	.2009	0.427*
35	NORMAL	.1521	0.772*
36	NORMAL	.1378	0.863*
37	NORMAL	.2009	0.427*
38	NORMAL	.4951	0.000*
39	NORMAL	.2465	0.199*
40	NORMAL	.1571	0.737*

VAR.	DISTRIBUTION TESTED AGAINST	DISTRIBUTION TESTED AGAINST	
		D1	PROB
3	NORMAL	.1515	0.749*
4	NORMAL	.1204	0.934*
5	NORMAL	.2233	0.272*
6	NORMAL	.3443	0.017*
7	NORMAL	.5274	0.000*
8	NORMAL	.2350	0.219*
9	NORMAL	.3290	0.026*
10	NORMAL	.2041	0.375*
11	NORMAL	.2243	0.267*
12	NORMAL	.2530	0.155*
13	NORMAL	.2172	0.302*
14	NORMAL	.1892	0.471*
15	NORMAL	.2229	0.273*
16	NORMAL	.2317	0.233*
17	NORMAL	.2091	0.346*
18	NORMAL	.5089	0.000*
19	NORMAL	.4521	0.001*
20	NORMAL	.2430	0.188*
21	NORMAL	.2130	0.324*
22	NORMAL	.9377E-01	0.995*
23	NORMAL	.1743	0.578*
24	NORMAL	.3138	0.039*
25	NORMAL	.2610	0.131*
26	NORMAL	.2244	0.266*
27	NORMAL	.1144	0.956*
28	NORMAL	.1839	0.508*
29	NORMAL	.2167	0.304*
30	NORMAL	.1741	0.579*
31	NORMAL	.1242	0.917*
32	NORMAL	.2037	0.378*
33	NORMAL	.2166	0.305*
34	NORMAL	.1421	0.814*
35	NORMAL	.1607	0.680*
36	NORMAL	.1692	0.616*
37	NORMAL	.2431	0.188*
38	NORMAL	.2108	0.337*
39	NORMAL	.3234	0.031*

THE HYPOTHESIS THAT THE SAMPLE IS FROM THE SPECIFIED DISTRIBUTION CAN BE REJECTED WITH THE INDICATED PROBABILITY (PROB) OF BEING INCORRECT.

# "This is where it all begins"

The Stock Exchange in the City of London is at the very heart of Britain's business life.

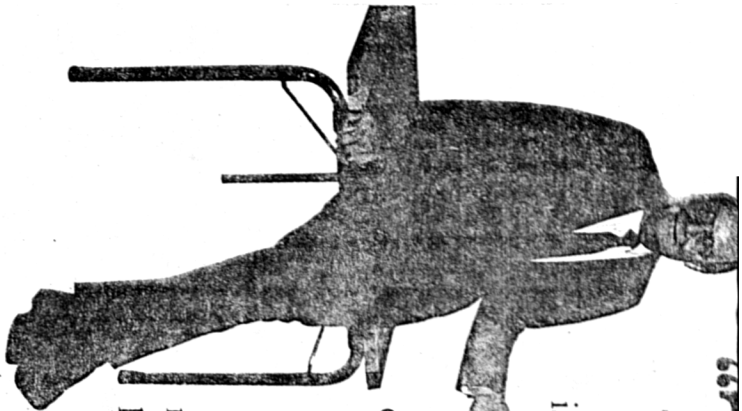
It's also at the heart of the unit trust business.

Here, on its busy trading floor, the shares of companies of every size and kind are bought and sold every working day of the week. Sometimes the trading will be slack; sometimes it will seem to be frantic.

Sometimes, because a company's news is a bit gloomy and its prospects look a little bleak, shares in that company will lose favour and their price will decline as people sell them.

At other times – and this is quite often the case – because a company's news is good and its prospects look rewarding, shares in that company will become popular and, as more and more people want to buy them, their price will go up.

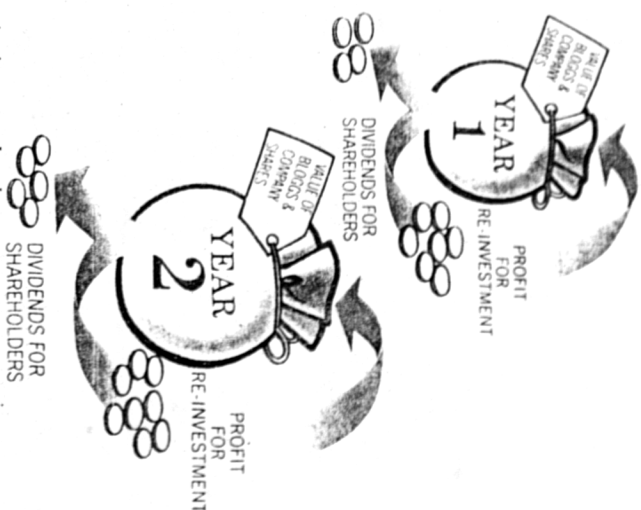
For the most part, and over the years, the general trend has been for share prices to go up. Hiccups may occur along the way, but, in the majority of cases, well-chosen shares are shares which will increase in value.



What's more, as a successful company increases its profits, it will plough some of them back into the business, by buying new machinery or building new premises. And this will add to the company's share value, too.

This is what is meant by capital growth. And alongside this growth – as an

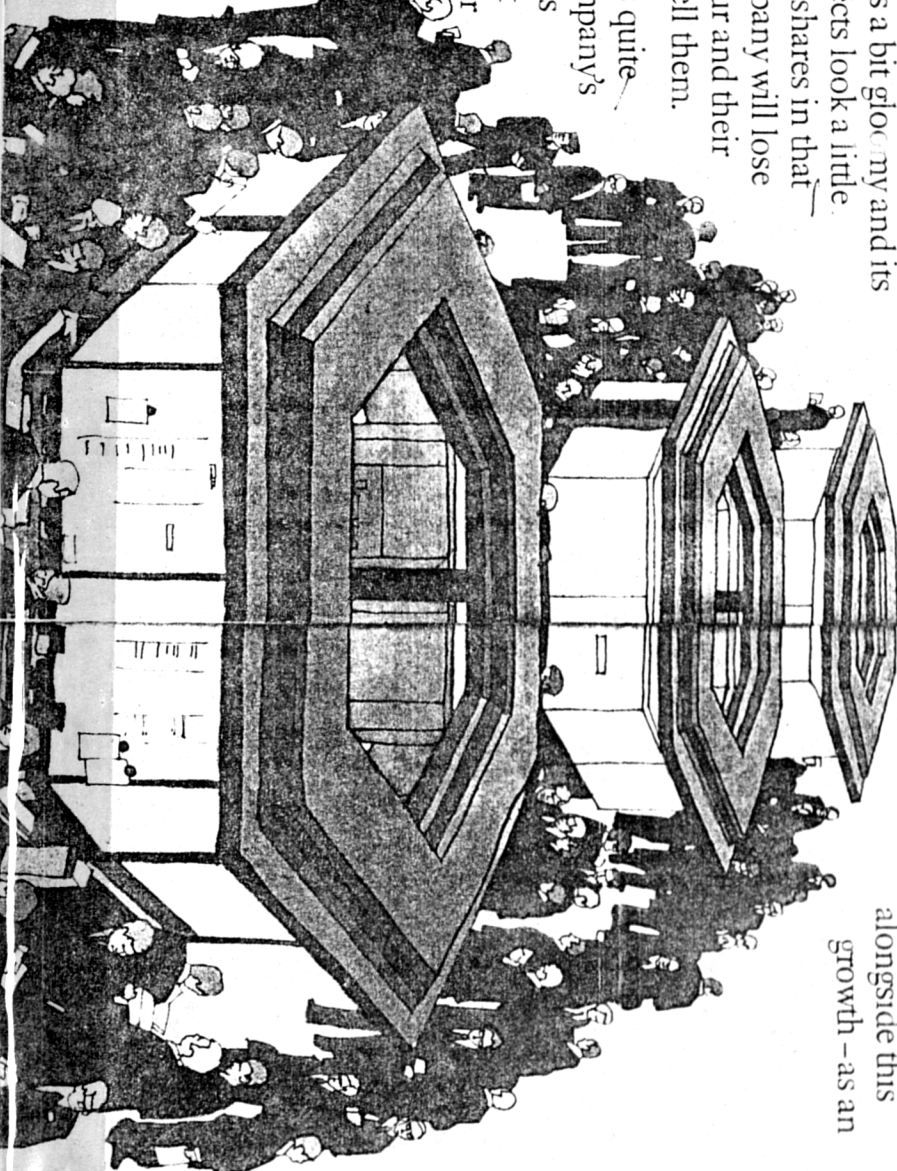
added benefit to investors – the companies will be paying out part of their profit to shareholders as dividends.



*As time goes by, the money you invest in a company's shares not only grows in value, but pays you an income as well.*

So, were you to invest in a variety of well-chosen shares, your money would have a real chance of becoming more valuable all the time. And you would be getting a rising income from it as well.

Which brings us on to unit trusts and how they fit i



## BIBLIOGRAPHY

- Bailey, Richard. (1979) 'Authorship in a forensic setting'. in Ager D, (ed) Advances in Computer aided Literary Research, AMLC, Birmingham
- Bailey, R.W. & Dolezel, L (1968) An Annotated Bibliography of Statistical Stylistics, University of Michigan Press
- Baker, C. (1985) Aspects of Bilingualism in Wales, Multilingual Matters, Clevedon
- Beardsley, Monroe, C. & Winstatt Jnr, W.K. (1959) 'The Concept of Meter: An exercise in Abstractions' in ed Sebeok, Thomas, A. (1960) Style in Language, MIT Press/Wiley, New York
- Biber, D. (1985) 'Investigating Macroscopic Textual Variation through Multifeature/Multidimensional Analysis', Linguistics, vol. 23, pp. 337-360
- Biber, D. (1986) 'Spoken and Written Textual Dimensions in English: resolving the contradictory findings' Language, vol.62
- Bolinger, D (1980) Language the Loaded Weapon, Longman, London
- Carroll, J.B. (1969) 'Vectors in Prose Style', in L. Dolezel and R. Bailey (eds.) Statistics and Style, Elsevier, New York
- Campbell, Leo, J. & Holland, V. Melissa. (1982) 'Understanding the Language of Public Documents because Readability Formulas don't', in (ed) Di Pietro, R.J.(1982) Linguistics and the Professions vol III: Advances in Discourse Processes, Ablex, Norwood New Jersey
- Catford, J.C. (1965) A Linguistic Theory of Translation, Oxford University Press, London
- Chan, R.K. (1970) Illustration of a Methodology for Stylistic Investigation with a special Reference to Formality in English, unpublished PhD Thesis, University of Reading
- Charrow, V.R. (1982) 'Language in the Bureaucracy', in (ed) di Pietro R.J. Linguistics and the Professions Vol VIII: Advances in Discourse Processes, Ablex, Norwood New Jersey
- Chase, S. (1955) The Power of Words, Phoenix House, London
- Chiu, R.K. (1973) 'Measuring Register Characteristics', International Review of Applied Linguistics, vol. 11, pp. 51-68
- Cmd. 8504 (see H.M.S.O.)
- Coleman, R. (1963) 'The Dialect Geography of Ancient Greece' in: Transactions of the Philological Society, 1963 pp.105-126

- Crystal, D. (1981) 'Objective and Subjective in Stylistic Analysis', in D. Crystal Directions in Applied Linguistics, Academic Press, London
- Crystal, D. & Davy, D. (1969) Investigating English Style, Longman, London
- Cutts, M. & Maher, C. (1980) Writing Plain English, Plain English Campaign, Salford
- Dolezel, L. & Bailey, R. (1969) Statistics and Style, Elsevier, New York,
- Dudley-Evans, A. (1986) 'Genre Analysis: An Investigation of the Introduction and Discussion Sections of Dissertations', unpublished, Aston
- Ehrenberg, A.S.C. (1975) Data Reduction, John Wiley and Sons, London
- Engels, L.K. (1968) 'The Fallacy of Word Counts', International Review of Applied Linguistics, vol 6
- Enkvist, Nils Erik, Spencer, J. & Gregory, Michael, J. (1964) Linguistics and Style, Oxford University Press, Oxford
- Forms and Notices Unit (1983) Writing Notices. A Guide, H.M. Customs and Excise, London
- Gibson, W. (1970) 'Styles and Statistics', in (ed) Freeman, D.C. Linguistics and Literary Style, Holt Reinhart and Winston, New York
- Gough, P.B. (1965) 'Grammatical transformations and speed of understanding', Journal of Verbal Learning and Verbal Behaviour, vol 4
- Gowers, E. (1962) The Complete Plain Words, Penguin, Harmondsworth
- Gregory, M.J. (1966) 'Aspects of Varieties Differentiation', Journal of Linguistics, vol. 3, pp. 177-274
- Halliday, M.A.K. & Hasan, R. (1976) Cohesion in English, Longman, London
- Halliday, M.A.K., McIntosh, A. & Strevens, P. (1964) The Linguistic Sciences and Language Teaching, Longman, London
- Harrison, Colin (1980) Readability in the Classroom, Cambridge University Press, Cambridge and London
- Hasan, R. (1973) 'Code Register and Social Dialect', in (ed) Bernstein, Basil, Class, Codes and Control vol 2: Applied Studies towards a Sociology of Language, Routledge Kegan Paul, London and Boston
- Hill, A. (1961) 'Grammaticality', Word vol 17

- H.M.S.O. (1982) Administrative Forms in Government: Govt White Paper (Cmd 8504), H.M.S.O. London
- Hogben, George, L. M.D. (1978) 'Linguistic Style and Personality', Language and Style, vol 10
- Hymes, D. (1972) 'On Communicative Competence', in J.B. Pride and J. Holmes (eds.) Sociolinguistics, Penguin, Harmondsworth
- Joos, M. (1959) 'The Isolation of Styles', in R.S. Harrel (ed.) Report of the 10th Annual Round Table Meeting on Linguistics and Language Studies No. 12, Georgetown University Press, Washington D.C.
- Joos, M. (1961) The Five Clocks, Indiana University Reserach Center in Anthropology, Folklore and Linguistics, Indiana
- Kendall, Maurice. (1975). Multivariate Analysis, Charles Griffin & Co., London
- Kenny, A. (1982) The Computation of Style, Pergamon, Oxford
- Kim, J. & Mueller, C.W. (1978a) Introduction to Factor Analysis, Sage Publications, Beverley Hills
- Kim, J. & Mueller, C.W. (1978b) Factor Analysis: Statistical Methods and Practical Issues, Sage Publications, Beverley Hills
- Kucera, H. & Francis, W. (1967) Computational Analysis of Present-Day American English
- Labov, W. (1970) 'The Study of Language in its Social Context' in (eds) Pride, J.B. & Holmes, Janet, Sociolinguistics, Penguin, Harmondsworth 1972
- Langley, R. (1979) Practical Statistics Simply Explained, Pan Books, London
- Leach, C. (1979) Introduction to Statistics, John Wiley and Sons, Chichester
- Leech, G.N. (1966) English in Advertising, Longman, London
- Leech, G.N. (1974) Semantics, Penguin, Harmondsworth
- Lyne, A.A. (1983) 'Word-Frequency Counts: their Particular Relevance to the Description of Languages for Special Purposes and a Technique for Enhancing their Usefulness', University of Nottingham Linguistics Circle (unpublished paper)
- Markworth, M.L. & Baker, W.J. (1980) 'A Discriminant Function Analysis of Co-variation of a Number of Syntactic Devices in five Prose Genres', in (eds) Prideaux, G.D, Derwing, B.L. & Baker, M.J. Experimental Linguistics, Story, Ghent



- McEntegart, D. & Le Page, R.B. (1982) 'An Appraisal of the Statistical techniques used in the Sociolinguistic Survey of Multilingual Communities' in (ed) Romaine, Suzanne (1982) Sociolinguistic Variation in Speech Communities, Arnold, London
- Moerk, E. (1973) 'An Objective, Statistical Description of Style', Linguistics, vol. 108, pp. 50-58
- Murry, John Middleton, (1922) The Problem of Style, Oxford University Press, Oxford
- Osgood, C.E. (1964) 'Semantic Differential Technique in the Comparative Study of Cultures,' American Anthropologist, vol 66,3
- Palmer, F.R. (1974) The English Verb, Longmans, London
- Plutchik, R. (1983) Foundations of Experimental Research, Harper and Row, New York
- Posner, Rebecca, (1963) 'The Use and abuse of Stylistic Statistics' Archivum Linguisticum vol 15
- Pride, J.B. & Holnes, Janet (eds) (1972) Sociolinguistics, Penguin Harmondsworth
- Prideaux, Gary D, Derwing, Bruce, L, & Baker, William, J, (1980) Experimental Linguistics, Story, Ghent
- Pytelka, J. (1972) 'Studies in the Language of Commerce', in V. Fried (ed.) The Prague School of Linguistics and Language Teaching, Oxford University Press, Oxford
- Quirk, R. (1982) Style and Communication in the English Language, E.J. Arnold, London
- Quirk, R. & Greenbaum, S. (1973) A University Grammar of English, Longman, London
- Redish, Janice C, (1985) 'The Plain English Movement' in: (ed) Greenbaum, S. The English Language Today, Pergamon, London
- Riffaterre, M. (1959) 'Criteria for Stylistic Analysis', Word vol 15
- Robson, C. (1973) Experiment, Design and Statistics in Psychology, Penguin, Harmondsworth
- Sandell, R. (1977) Linguistic Style and Persuasion, Academic Press, London
- Scott, F.S. and others (1970) English Grammar: a Linguistic Study of its Classes and its Structures, Heineman, London

- Spencer, J.W. & Gregory M.J. (1964) 'An Approach to the Study of Style', in Enkvist N.E. & Spencer & Gregory, M.J. (1964) Linguistics and Style, Oxford University Press, London
- SPSS inc. (1983) SPSS-X User's Guide, McGraw Hill, New York
- Strang, B.M.H. (1962) Modern English Structure, Arnold, London
- Strevens, P.D. (1964) 'Varieties of English', English Studies, vol. 45, pp. 1-10
- Thorndike, Edward, L. & Lorge, Irving. (1944) The Teacher's word book of 30,000 words, Columbia University
- Turner, G.W. (1973) Stylistics, Penguin, Harmondsworth
- Vestergard, T. & Schroeder, K. (1985) The Language of Advertising, Blackwell, Oxford
- Swales, J. (1981) 'Aspects of Article Introductions', Aston Research Reports, No. 1, University of Aston
- Ullmann, S. (1964) Semantics, Blackwell, Oxford
- Vernon, T. (1980) Gobbledegook, National Consumer Council, London
- Wallace, W.D. (1981) 'How Registers Register: Toward the Analysis of Language Use', International Journal of Applied Linguistics, vol. 19, pp. 267-286
- Wells, R. (1960) 'Nominal and Verbal Style' in T.A. Seboek (ed.) Style in Language, Massachusetts Institute of Technology, Massachusetts
- West, M. (1953) A General Service List of English Words, Longman, London
- Williams, C.B. (1970) Style and vocabulary numerical studies, Griffin, London
- Winter, W. (1969) 'Styles as Dialects', in L. Dolezel and R. Bailey (eds.) Statistics and Style, Elsevier, New York
- Woods, A, Fletcher, P. & Hughes, A. (1986) Statistics in Language Studies, Cambridge University Press, Cambridge
- Yule, U. (1950) An Introduction to the theory of Statistics, Griffin, London
- Wason, Peter C. (1984) 'Conformity and Commitment in Writing', Visible Language, vol 14, pp 351-363