

A DISCRIMINANT FUNCTION ANALYSIS OF
CO-VARIATION OF A NUMBER OF SYNTACTIC DEVICES
IN FIVE PROSE GENRES

MARY LOIS MARCKWORTH

WM. J. BAKER

Department of Linguistics
University of Alberta
Edmonton

Copyright 1975

Association for Computational Linguistics

ABSTRACT

This paper approaches the question of whether the styles of different prose genres (here referring to classifications such as fiction, newspaper reportage, learned journals, etc.) are partially characterized by differences in the co-variation of a number of common syntactic structures.

Thirty-six syntactic variables are tabulated in sample sentences drawn from five genres. The variables are of several different sorts; for example, sentence types, some focus phenomena, elements of the verb structure, conjoined structures, and various modification and complementation structures.

The co-variation of these syntactic variables is analysed by means of a discriminant function analysis. The analysis shows that the verbal styles of the specific genres considered are characterized by differential patterns of occurrence of the syntactic structures. The patterns appear to be primarily interpretable as artefacts of the common semantic denominator (or communicative purpose) of their genres.

TABLE OF CONTENTS

Abstract	2
Introduction	4
Procedure	7
Results	12
Conclusion	29
References	31

TABLES AND FIGURES

Table 1. Syntactic variables selected for analysis	8
Table 2. Sentence-length blocks for sampling	11
Table 3. Frequency of occurrence of each variable	13
Table 4. Variables discriminating fiction from the four non-fiction genres	17
Table 5. Variables discriminating among non-fiction genres	25
Figure 1. Plot of group centroids for the five genres on the first two discriminant functions	16
Figure 2. Plots of the group centroids for the four non-fiction genres for discriminant functions I, II and I, III	24

INTRODUCTION

It is clear that linguists should attempt to use all of the empirical evidence obtainable to describe usefully and completely the significant patterns which occur in language data, but it is not always clear just what patterns or relationships in the data should be included in such descriptions. One aspect of language data--generally referred to as style--has been known for some time to be subjectively rather obvious but extremely difficult to characterize in objective terms, especially if one ignores the obvious role of lexical selection.

It would be possible to talk about style in non-subjective terms if one were able to identify those surface elements of language that appear to participate in stylistic variations and then to relate these to subsets of language data defined by external criteria of known genre, provenance, author, or situation. However, simple counts of elements with respect to such subsets have not proven to be overly instructive, primarily because it is presumably not the simple occurrence of the elements which leads one to perceive a certain style but, rather, the far less obvious co-occurrence of sets of elements (Marckworth, 1973). These co-occurrences are ones which are not required by the

grammar of a language but which have a high probability of being linked in the genre in which they occur.

A useful definition of style might be the overall pattern or gestalt perceivable in language data which transmits information about an utterance with respect to the personal, social, and cultural relationships assumed between speaker and hearer, the historical or geographical provenance of the source, and the identity or mental state of the speaker. Recognition of such a gestalt allows the hearer or reader to assign the language data to a specific subpopulation about which he already has prototype information.

Assuming this provides a workable view of style, then the basic task of an investigation of the phenomenon is one of generating hypotheses about possible language-element participants in style-marking constellations which can be investigated in themselves and in patterns of co-occurrence. The results of such investigations, if successful in differentiating externally defined subpopulations of language, will provide additional data about language to which any grammatical description must be responsible.

Genre, the focal point of this study, is quite difficult to define in specific terms. As Marckworth (1973, pp. 24-25) has noted:

Genre tends to be a Humpty-Dumpty term, a useful concept which is redefined within the limits of each discussion of it with little requirement that such definitions have more than a superficial agreement with each other. Genre may distinguish prose from poetry from drama; it may distinguish historical plays from French farces, sonnets from epics, mystery stories from all other novels. As a construct it has, in fact, much in common with classification as defined by pattern recognition studies: a group of items along some characteristic (attribute) continuum whose members are closer to each other than to their neighbors and which are not separated by a gap across a critical perceptual boundary or distance. The characteristic of the continuum and the critical distance are continually respecifiable in terms of the stimuli perceived and the fineness of the classificatory system required.

When viewed in these terms, it is immediately clear why genre is at once such a vague and such a useful term. Continua based on a number of different characteristics may be used as the basis for defining clusters of similar texts and the genre classifications formed on different continua need not be parallel nor form mutually exclusive groups. Some of the continua used in genre classification are content, as in the case of the mystery story or the Cavalier love poems; intent, as in the case of humorous writing or fantasy; form, as in the case of drama or the novel; and context, as in the case of technical or belletristic prose. This is not to claim that such characteristics are independent of each other or that any text may not be defined in terms of all four; simply that groupings along a continuum defined by any one of them alone may be cited to establish contrasting genres.

This paper reports an investigation of a large variety of surface syntactic features, ranging from sentence type through elements of the verb, and the covariant patterns of occurrence of such structures in five genres of written modern English. It is recognized, of course, that powerful

determiners of the recognizability of various genre styles exist at both the infra-sentential (e.g., specific lexical selection) and supra-sentential (e.g., general content) level, but it is the aim of this study only to assess, at the level of the sentence, the contribution of syntactic devices to a measurable difference in genre-defined styles.

PROCEDURE

Thirty-six syntactic structures were selected for consideration on the basis of being possible contributors to the constellations determining genre style. They were chosen from four main categories: sentence-type, including the range of possible interrogative patterns; focus phenomena of various sorts; elements of the main verb phrase; and a group of conjoined or embedded structures. This latter group included structures of noun-phrase modification, verbal complementation, sentence modification, and parallel element conjoining.

Table 1 shows the syntactic structures whose frequency of occurrence was counted, along with the identification numbers assigned to each of these variables for use in subsequent tables and discussions. Note that the various structures are clearly not completely independent of each other.

TABLE 1

SYNTACTIC VARIABLES SELECTED FOR ANALYSIS

Category	Variable Number	Variable Name
Sentence Type	1.	Declarative
	2.	Interrogative: word-order inversion type
	3.	tag type
	4.	Wh- type
	5.	intonation type
	6.	Imperative
Focus Phenomena	7.	Non-standard word order
	8.	Passive construction
	9.	Cleft construction
	10.	Extraposition construction
Verb Structure	11.	Auxiliaries: modals and catenatives
	12.	progressive aspect
	13.	perfective aspect
	14.	emphatic <u>do</u>
	15.	past tense marking
	16.	Main Verb: transitive
	17.	intransitive
	18.	<u>to be</u>
	19.	other copulas
	20.	Contracted verbal forms
Conjoining & Embedding	21.	Simple: of full sentences or clauses
	22.	of phrases or words
	23.	Inclusion of direct discourse
	24.	Nominalization: in the NP (cleftable)
	25.	in the VF (non-cleftable)
	26.	adverbial clauses
	27.	indirect questions
	28.	NP modification: adjectives
	29.	locatives
	30.	appositives
	31.	full relative clauses
	32.	partially reduced relative clauses
	33.	noun-adjuncts
	34.	noun-heads, including pronouns
	35.	Adverbials: prepositional phrases
	36.	other adverbials

Extraposition, e.g., implies an embedded clause; a passive construction implies a transitive verb; conditional clauses may imply the past tense marker; contracted verbal forms imply auxiliaries; and emphatic do and other auxiliaries are mutually exclusive. However, none of these relationships (referred to here as grammatical co-occurrence restrictions), with the exception of the last, is reciprocal. Extraposition implies an embedded clause, but an embedded clause does not necessarily imply extraposition; a passive construction implies a transitive verb, but a transitive verb does not imply a passive construction, etc. Since these various elements are not completely redundant, they are able to operate at least semi-independently as possible syntactic indicators of style.

Five genres were chosen for investigation on the basis of context of utterance which permits identification by place of publication. The genres selected were: Learned Journals, Newspaper Reportage, Popular Journals, Government Documents, and Fiction.

The actual data were drawn from the Brown University million-word English corpus, A Standard Sample of Present-day Edited American English for Use with Digital Computers. This corpus consists of 500 samples of English-language texts published in the United States in 1961, each sample

approximately 2,000 words long. This large number of relatively short samples minimizes the effect of any single author or topic, and the restrictions on date and place of publication control variables associated with provenance. A complete description of this corpus and its content may be found in Francis (1964) or in Kucera & Francis (1967).

A total sample of 500 sentences was drawn, 100 from each of the five genres. Each genre subset of 100 consisted of ten sentences from each of ten sentence-length blocks. Sentence length was measured in words, and the blocks are shown in Table 2.

These block lengths were chosen to mirror roughly the distribution of sentence lengths in the entire corpus. A structured sample of this kind was drawn to prevent sentence length as such from acting as a variable since sentence-length distribution was already known to differentiate among genres (Marckworth & Bell, 1967), and also to guarantee that those syntactic devices which tend to be associated with greater sentence length would have equal opportunities to appear within each genre sample. Again, the emphasis in this study was on the sentence as the basic unit of analysis, and on the co-occurrence of syntactic structures within sentences. The reader should keep in mind, however, that a random sample of sentences would permit certain sentence

TABLE 2
 SENTENCE-LENGTH BLOCKS FOR SAMPLING

Block Number	Block Length in Words
1	1 - 5
2	6 - 9
3	10 - 12
4	13 - 16
5	17 - 20
6	21 - 25
7	26 - 30
8	31 - 40
9	41 - 49
10	50 - 240

lengths to dominate in specific genres, possibly obscuring the co-occurrence patterns of interest here but, nevertheless, reflecting another property which can clearly be said to characterize zeigenre style.

Each of the sample sentences was analyzed for the occurrence of the syntactic variables indicated in Table 1, and the number of occurrences of each structure was recorded. A discussion of the basis on which the syntactic elements were identified may be found in Marckworth (1973, pp. 44-48). The basic data for analysis thus consisted of

500 observations (sentences) with each observation scored on 36 variables and classified by genre and length. The subsequent analysis of these data was based primarily on discriminant functions which were used to determine how the variables served to distinguish one genre from another.

Basically, discriminant function analysis is the multivariate extension of the univariate F ratio which is used to distinguish among previously established groups. It represents, however, a considerable increase in both complexity and analytical power since it focuses not only on the simple differences between groups on each variable, but also on the interrelationships among differences on the several variables considered simultaneously. It serves to maximize group differences by developing maximally efficient weights which, when applied to the original data, will yield the clearest distinctions among the groups being analyzed. The method of discriminant function analysis is discussed fully in Rulon, et al. (1967, pp. 299-319)

RESULTS

The frequency of occurrence of each of the variables within each of the genre categories is shown in Table 3. Eleven of the variables, indicated by an asterisk following

TABLE 3

FREQUENCY OF OCCURRENCE OF EACH VARIABLE

Genres						
Variables	Learned Journals	Newspaper Reportage	Popular Journals	Government Documents	Fiction	Total
1	97	98	87	95	92	465
2	0	0	1	0	0	1*
3	0	0	0	0	0	0*
4	0	3	1	1	1	6*
5	0	0	0	0	0	0*
6	3	0	12	4	9	28
7	0	4	2	4	2	12*
8	43	28	33	33	12	149
9	1	0	0	1	0	2*
10	3	1	2	1	3	10*
11	30	30	26	46	50	182
12	3	9	5	5	12	34
13	11	19	13	13	41	97
14	2	0	0	3	0	5*
15	63	84	66	44	209	466
16	107	113	142	98	141	601
17	29	33	36	25	81	204
18	50	49	42	41	63	245
19	0	11	4	2	4	21*
20	0	7	6	0	32	45
21	32	37	51	27	62	209
22	141	129	108	112	108	598
23	0	9	4	0	13	26
24	39	19	24	23	51	156
25	22	24	28	25	24	123
26	29	13	19	29	59	149
27	0	0	2	0	3	5*
28	329	335	300	278	230	1472
29	0	7	2	1	0	10*
30	17	27	24	9	11	88
31	24	38	31	23	33	149
32	24	14	25	37	17	117
33	13	10	7	23	8	61
34	705	711	714	690	694	3514
35	321	285	296	345	224	1471
36	156	132	143	142	217	790

*Omitted from main analysis because of low frequency of occurrence in the total sample (≤ 25).

the total column, appeared in less than five percent of the sentences examined. Because of this low incidence, interpretation of these variables would be difficult and tenuous so they were omitted from further analyses.

Each variable was first examined individually in an analysis of variance for the five genre groups and four tests were run on each to determine if the variable would differentiate:

- a. fiction from the non-fiction genres,
- b. formal (Learned Journals and Government Documents) from informal (Newspaper Reportage and Popular Journals) non-fiction,
- c. Learned Journals from Government Documents,
- d. Newspaper Reportage from Popular Journals.

Fourteen variables (8, 11, 12, 13, 15, 16, 17, 18, 21, 24, 26, 28, 35, and 36) were found significantly to differentiate fiction from non-fiction ($p \leq .05$); three variables (16, 26, and 33) distinguished formal from informal non-fiction; none distinguished between Learned Journals and Government Documents; and only one (1) differentiated Newspaper Reportage from Popular Journals.

Two of the variables (6 and 23) could not be examined in this way because of zero incidence in some genres, but the data pattern for 23 (inclusion of direct discourse)

suggests an obvious distinction between formal and informal non-fiction. Variable 6 (imperatives) would clearly distinguish between Newspaper Reportage and Popular Journals, which is not surprising in view of the number of how-to-do-it articles in the latter genre. These two variables (6 and 23) could be and were retained for the discriminant function analyses. The remaining variables (10, 22, 25, 30, 31, 32, and 34) showed no distinction as univariate indices on the four tests but they were, nevertheless, also retained for the multivariate analysis since they could, when analyzed in conjunction with other variables, still provide information for genre distinction. This is because the simple univariate analyses discussed above do not take into account the possible intercorrelations (constellation effects) among the variables.

The first multivariate analysis was a five-group discriminant function analysis, performed on the five genres. It indicated a clear differentiation of fiction from the four non-fiction genres (see Figure 1), on the basis of the eight syntactic variables listed in Table 4. These results demonstrate that sentences from all of the non-

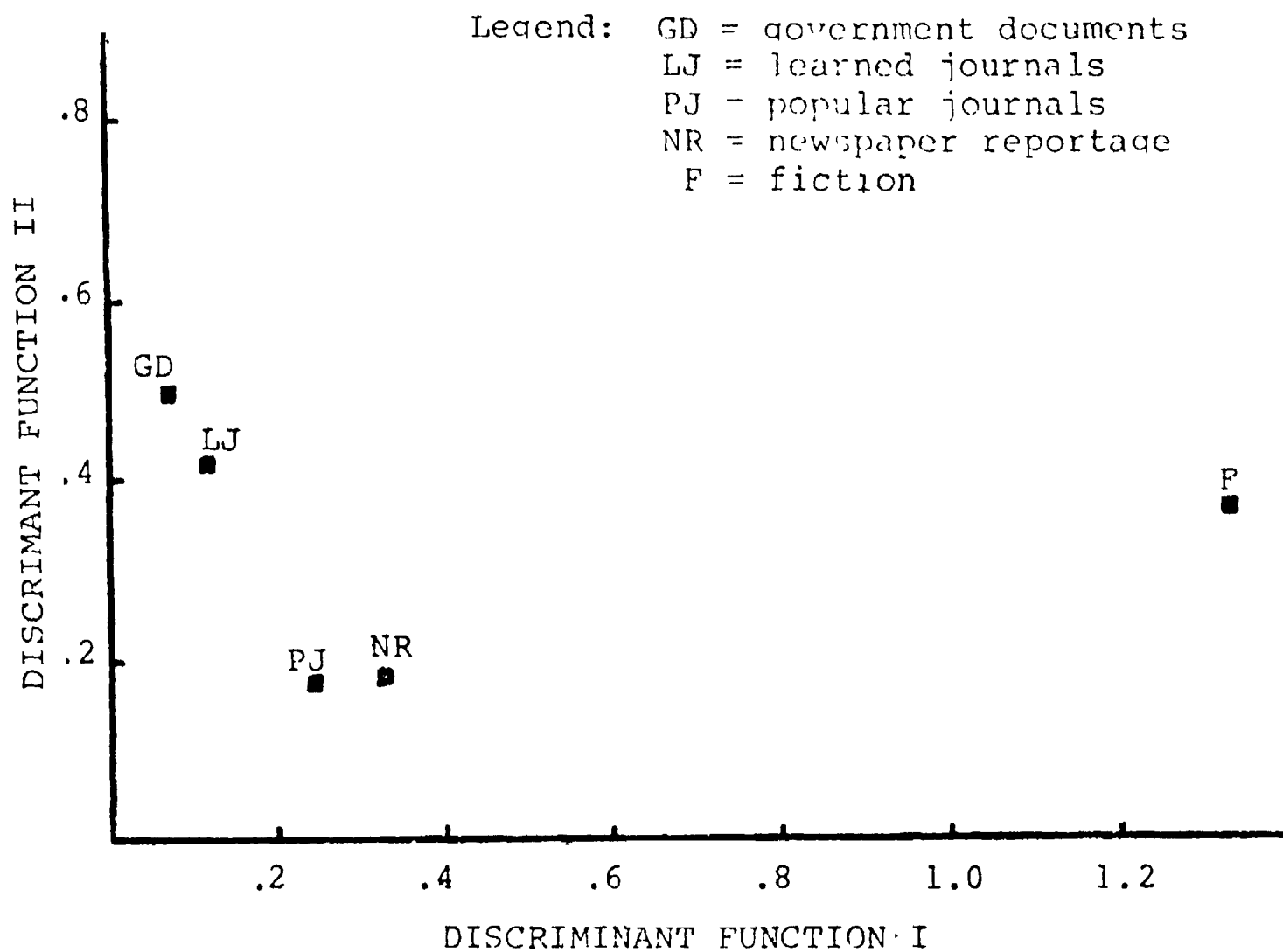


Figure 1. Plot of the group centroids for the five genres on the first two discriminant functions.

TABLE 4
 VARIABLES DISCRIMINATING FICTION
 FROM THE FOUR NON-FICTION GENRES

Variable Number	Variable	Discriminant Weight*
15	past tense marking	.73
20	contracted verbal forms	.47
17	intransitive verbs	.43
13	perfective aspect	.39
26	adverbial clauses	.36
23	inclusion of direct discourse	.32
36	other adverbials	.27
8	passive construction	-.26

* Those variables having a positive value are characteristic of fiction sentences, but not of non-fiction; those values with a negative value, vice versa.

fiction genres are more alike in syntactic structure than any of them are like sentences from fiction. This may, at first view, be surprising in light of the range of non-fiction genres included in the study, but it bears out the findings of at least one other investigation of quantitative characteristics of the language of different genres (Marckworth and Bell, 1967, on sentence-length distributions): that the major measurable stylistic distinction is between fiction and non-fiction genres.

Several interesting observations may be made about the syntactic variables that participate in the discrimination (see Table 4). The most obvious point is the heavy involvement of syntactic features of the verbal unit in differentiating fiction from non-fiction styles. With the exception of inclusion of direct discourse, which seems transparently attributable to the dialogue characteristic of fiction, and the lack of passive constructions, since voice has been shown to be a whole-sentence focus phenomenon (Andrew, 1974), all of the variables in Table 4 are associated with verbal rather than nominal elements of the sentence. Marking for past tense and perfective aspect, intransitive verbs, and contracted verbal forms are all specific to the verb phrase, and adverbial clauses and other adverbials are either specifically verb-modifying or are whole-sentence-modifying. Apparently elements of the noun

phrase, or at least those considered in this study, do not participate in the distinctively style-associated constellations of syntactic structures that distinguish fiction from non-fiction.

A second notable feature distinguishing the fiction sentence set is the amount of indication of past time action. This is conveyed not only by the formal past tense variable, which in the great majority of cases does indicate a past time action, but also by the perfective aspect, which always indicates a past time event whether marked for past or present tense. This feature is perhaps understandable in view of the usual function of fiction as a narrative of past events, and it should also be noted that this same function may utilize another role of perfective aspect--that of interrelating sequential events through time.

A question may be raised about the relationship of two of the fiction-distinguishing variables: the presence of intransitive verbs and the absence of passive constructions. Since these two exhibit a non-reciprocal grammatical co-occurrence restriction between voice and verb type--passive voice implies a transitive verb but not vice versa--it is possible that the paucity of passive constructions in fiction is simply an artefact of the frequency of intransitive verbs. In other words, we must ask whether

passive sentences occur less in fiction than in non-fiction simply because they have less opportunity to do so, or whether non-passiveness is an independent syntactic feature of fiction style. A comparison of the ratios of occurrence of passive construction to transitive verb in the fiction and non-fiction genres shows the latter case to be the true one. (Such a ratio expresses the actual occurrence of passive sentences in relation to the possible occurrences.)

The ratios for the five genres are

Fiction	.085
Learned Journals	.402
Newspaper Reportage	.248
Popular Journals	.232
Government Documents	.337

It is tempting to speculate about just why the fiction genre should be characterized by significantly more intransitive verbs to cause this variable to be discriminatory. One possible explanation, which suggests a simple characterization of fiction style, could be based on analyzing a large subset (if not the whole class) of verbs usually called intransitive as items which can occur both with no object (traditionally called intransitive verbs) and with one or two objects (traditionally called transitives).

In such an analysis it is presumed that when the grammatical object of a verb of this class is either redundant or not completely specified it is suppressed and

the result is a one-place predicate sentence, e.g., John sang. When the grammatical object carries new or required information it is present and the result is a multiple-place predicate, e.g., John sang a Greek folksong. This sort of analysis of the "intransitive" verb opens the door to a very general characterization of genre differences. Thus viewed, the intransitive verb variable characterizes sentences which are not heavily information oriented--sentences in which a major component, the grammatical object, is either so predictable or so unimportant that it is not even specified.

Such sentences are significantly more characteristic of fiction than of non-fiction writing, and this analysis of them suggests a measurable basis for the old rule of thumb that successful (although not necessarily good) fiction writing is strongly action oriented. It also suggests the validity of the common-sense intuition that a primary source of the differences between fiction and non-fiction is that the latter is designed foremost as an information-conveying instrument; that in the dichotomy of literary purpose it is more likely to teach than to delight.

Since one of the major ways in which information can be packed into a sentence is through heavy use of nominal elements, we looked at a simple measure of this characterization of fiction style as action-oriented as

opposed to non-fiction style as information-oriented: the verb/noun ratio for each genre.

These ratios are

Fiction	.416
Learned Journals	.264
Newspaper Reportage	.290
Popular Journals	.314
Government Documents	.241

bearing out the supposition that sentences in the non-fiction genres have more nouns in proportion to verbs than do those from fiction. It may also be noted that, in addition to being high in nouns in proportion to verbs, non-fiction sentences also exhibit somewhat more noun modification than fiction, as shown by the following ratios of all noun-modifier types (variables 28, 29, 30, 31, 32, and 33) to nouns:

Fiction	.431
Learned Journals	.577
Newspaper Reportage	.606
Popular Journals	.545
Government Documents	.564

Thus, from the discriminatory variables identified by the five-group discriminant function analysis and the further observations suggested by them, a picture emerges of distinctive syntactic structure constellations in at least two major genre categories: fiction, with the syntactic structures determined by the function of past-time, action-oriented, narrative communication; and non-fiction, with structures determined by an information-carrying function.

The five-group discriminant function analysis showed such a major distinction between fiction and the non-fiction genres that it seemed possible that differences in the non-fiction genres might have been obscured. In consequence, a four-group discriminant function analysis was done on data from these genres only. The result indicated a distinction in syntactic structure between the formal genres (Learned Journals and Government Documents) and the informal (Newspaper Reportage and Popular Journals) along the first axis, and along the second axis a distinction between Popular Journals and Newspaper Reportage. The third dimension distinguished Learned Journals from Government Documents (see Figure 2).

Table 5 shows the syntactic variables that participate in these three discriminations. These discriminating structures present a less distinct picture of different types of writing than do those differentiating fiction and non-fiction, but nevertheless illustrate some interesting points about genre and style.

Of the four items that characterize informal non-fiction sentences, only transitive verb is not susceptible to immediate explanation, although we can note that its presence as an informal marker must be due almost entirely to sentences from Popular Journals since, in the

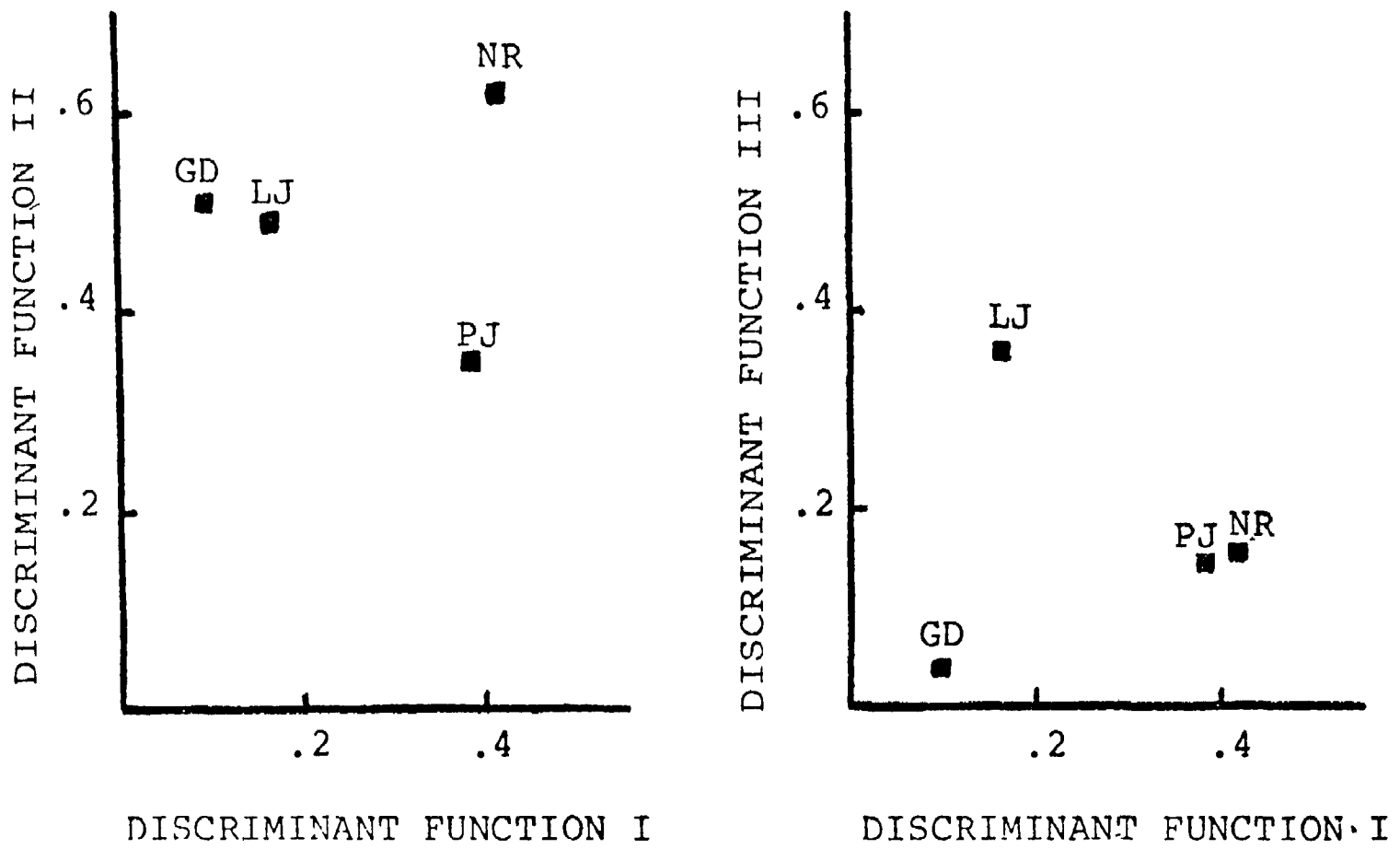


Figure 2. Plots of the group centroids for the four non-fiction genres for discriminant functions I, II and I, III.

TABLE 5

VARIABLES DISCRIMINATING AMONG NON-FICTION GENRES

Discriminant Function	Variable Number	Variable	Discriminant Weight
I. INFORMAL FROM FORMAL NON-FICTION GENRES			
	23	inclusion of direct discourse	.47
	20	contracted verbal forms	.33
	15	past tense	.30
	16	transitive verbs	.29
	32	partially reduced relative clauses	-.26
	33	noun adjuncts	-.32
	26	adverbial clauses	-.32
II. NEWSPAPER REPORTAGE FROM POPULAR JOURNALS			
	1	declarative sentences	.55
	23	inclusion of direct discourse	.27
	16	transitive verbs	-.35
	6	imperative sentences	-.65
III. LEARNED JOURNALS FROM GOVERNMENT DOCUMENTS			
	24	nominalization in noun phrase	.41
	8	passive constructions	.28
	22	conjoined words and phrases	.25
	33	noun adjuncts	-.26
	11	modals and catenatives	-.31

discrimination between Newspaper Reportage and Popular Journals, it has a negative weighting for the former; that is, in the comparison of this pair of genres, Newspaper Reportage is distinguished by the absence of transitive verbs (see Variable 16, Discriminant Function II, in Table 5).

The other three distinguishing features of the informal genres are inclusion of direct discourse, contracted verbal forms, and past tense. Inclusion of direct discourse is probably present as a result of the fact that the parsing procedure did not differentiate between true direct discourse of the sort found in fiction dialogue and the inclusion of quoted material of the sort found in Newspaper Reportage in which one or two words may be quoted. The supposition that Newspaper Reportage contributed this discriminant variable to the informal category is borne out by its appearance as a characteristic distinguishing that genre from Popular Journals, as shown by the second function in Table 5.

Contracted verbal forms are a typical and, frequently, a deliberate indicator of informal style, which probably explains the presence of this variable as a discriminator. In addition, editorial policy (or a writer's perception of it) usually discourages the use of these forms in any sort

of formal written language, which explains their absence in Learned Journals and Government Documents. Why past tense should differentiate informal from formal non-fiction is not really clear; one possible contributory cause may be that Learned Journals frequently discuss things as they are (or appear to be, Government Documents (which in this sample are largely proclamations of future legal interpretations or holidays) discuss things as they will be, but Newspaper Reportage discusses things as they have been.

Several other transparent discriminatory variables are to be seen in Table 5. Variable 8, passive constructions, which distinguishes Learned Journals from Government Documents, is almost certainly a result of conscious editorial policy. The presence of declarative sentences and the absence of imperative sentences which distinguish Newspaper Reportage from Popular Journals are a joint result of the presence in Popular Journals of a number of how-to-do-it articles: "Hold the brick in your left hand"

A less transparent, but perhaps more analytically interesting set of variables is shown in Table 5 (32, 33, 26, and 24, 22, 33). The first three, partially reduced relative clauses, noun adjuncts, and adverbial clauses, are atypical of informal non-fiction in the informal/formal discrimination; the other three are involved in

distinguishing Learned Journals from Government Documents--
rominalizations in the noun phrase and conjoined words and phrases by their presence in Learned Journals and noun adjuncts by its absence. (The activity of noun adjuncts as discriminatory by its absence in both of the informal genres and in Learned Journals indicates its heavy use in Government Documents: "I, John Chaffee, Governor of Rhode Island....") All five of these structures are significantly characteristic of one or both of the formal non-fiction genres, and all five are from the category of conjoined or embedded syntactic elements--that is, syntactic structures whose primary purpose is to compress and relate information within the sentence. Seemingly, those genres in which the author's intent is to convey maximum information with maximum explicitness are just those that make maximum use of such syntactic techniques. (We may note in passing that, except for adverbial clauses and for conjoined words and phrases, which may involve either nouns or verbs, the significant elements are members of the noun phrase; it appears that only at this level of genre discrimination is anything but a verbal or whole-sentence element a significant stylistic indicator.)

CONCLUSION

What seems to be evident from the above results is that, while there are indeed significantly different patterns syntactic occurrence between genres, these patterns (with the exception of editorially determined use of passive constructions in Learned Journal style and avoidance of contracted verbal forms in any formal style) result primarily from general semantic constraints operating within the genres and based in the communicative purposes of the genres. To wit, fiction, no matter what its topic, is typically a narration of past but interconnected actions, and the syntactic structures that differentiate fiction from non-fiction are ones which convey this semantic content; non-fiction, again no matter what it is about, is in general a data-conveying instrument, even though there are detectable differences in the manner in which the data are conveyed, e.g., degree of specificity of data (Learned Journals), degree of didacticism (Popular Journals), and degree of included narration (Newspaper Reportage). Again, these broad semantic similarities and differences are reflected in the syntactic structures that differentiate the genre styles. In summary, quantitative differences in syntactic structure can indeed be found between independently-defined sub-populations of language (genres), but they appear to correspond to--and are presumably the

result of--generic communicative purposes of the genres, and should consequently be viewed as internally-constrained artefacts of this semantic component rather than externally-defined elements of style.

REFERENCES

- Andrew, C.M. (1974). An Experimental Approach to Grammatical Focus. Ph.D. Dissertation. University of Alberta, Department of Linguistics, Edmonton, Alberta.
- Francis, W.N. (1964a). A Standard Sample of Present-day Edited American English, for Use with Digital Computers. (Magnetic Tape). Brown University, Department of Linguistics, Providence, R.I.
- Francis, W.N. (1964b). Manual of Informator of Accompany a Standard Sample of Present-day Edited American English, for Use with Digital Computers. Brown University, Department of Linguistics, Providence, R.I.
- Kucera, H. and W.N. Francis (1967). Computational Analysis of Present-day American English. Brown University Press, Providence, R.I.
- Marckworth, M.L. and I.M. Bell (1967). Sentence-length distribution in the Corpus. In Kucera and Francis, pp. 368-405.
- Marckworth, M.L. (1973). Statistical Determination of Some Elements of Genre Style. Ph.D. Dissertation. Brown University, Department of Linguistics, Providence, R.I.
- Rulon, P., D.V. Tiedeman, M.M. Tatsuoka and C.R. Langmuir (1967). Multivariate Statistics for Personnel Classification. Wiley, New York.

END

