CHAPTER IV

Results of the Phrase Structure Rule Count

Table 1 gives the ranks of the high frequency rules as found in the samples from the three novels. The most striking result is that the ranks for FELIX KRULL and DOKTOR FAUSTUS are exactly the same revealing 100% positive correlation. The significance of this assumes weight if we remember that the samples were randomized. It shows convincingly that beneath the thousand minute variations which an author uses in developing a long novel there is hidden order, an "invariant characteristic"[180] over which a writer has as little control as over the autonomous functions of the body. The rho values for the sequential samples I and II show the least correlation, which supports the hypothesis that the same novel can show heterogenity in parts.

Table 18 shows that among the grammar forms the prepositions occur with the highest frequency. As P——P(N), the prepositional use, is the only rule in our list which deals with a functional word, one may not be much surprised that rule 26 tops the list each time, nor that N——N(A) (the expansion of a noun by a modifier) figures as the

---

[180] The term "invariant characteristic" has been used by Sol Saporta (153).

second most frequent. But there is no particular reason why the other rules should occur in a particular order. Even assuming that the rules are likely to vary by not more than one rank either way, the number of permutations possible with the remaining 9 rules is 55 so that the probability of a particular arrangement is 1/55 only. This is the calculation under the most stringent restriction that there are unknown stabilizing factors in language by reason of which the ranks of each does not vary beyond a very narrow limit. The results can therefore be taken as significant. In the first chapter we posed the question : how far has the style of Thomas Mann varied over the years? What is the significance of the statistical results for this question? The high correlation between the <u>random</u> samples from BUDDENBROOKS and FELIX KRULL and between those from BUDDENBROOKS and FAUSTUS shows that the generative structure of the novels are same, i.e., the highly complicated sentence structure of DOKTOR FAUSTUS, where the periodic construction is longer than in BUDDENBROOKS has fundamentally the same syntactic structure. Wherein lies then the difference? Let us take a look at the mean number of syllables per sentence and the mean number of rules applied per sentence in the three novels :

	year	I	J	
Buddenbrooks	(1901)	35'3	9'78	181
Felix Krull	(1954)	38'7	12'01	
Doktor Faustus	(1947)	50'8	18'08	

These figures show that the sentence length has increased from Buddenbrooks to Faustus and also that the number of applications of the rules has consequently registered an increase. This means that the author has expanded the width of the sentence without altering his preference for and among the dominant rules. Variation in style has therefore to take into consideration both quantity as well as quality of change. The method provides a quantitative index for defining the texture of a text passage. The sequential samples I and II and random sample III from Buddenbrooks have, for the 11 dominant rules, the following mean and standard deviations :

	mean	s.d.
Sample I	6'33	7'22
Sample II	5'98	5'32
Sample III	7'71	6'70

Taking the random sample as norm, we find that the mean of the first sample is less the norm mean, but the dispersion is greater. Sample II also has a smaller mean, but the dispersion is also less. The smaller means denote that the eleven dominant rules do not cover a large percentage of

---

[181] Data regarding syllable means are from WEISS (286, page 95)

$\bar{I}$ = mean number of syllables per sentence

$\bar{J}$ = mean number of rules per sentence

the rules and that a larger number of rules have been employed in the text. In that sense we may say, the higher the mean, the denser the text. The dispersion indicates how the values of the various rules are distributed around the mean. The higher the dispersion, the greater are the differences among the various values. The lower the dispersion, the lesser the marked preference which the author shows for any particular rule or rules. The density of a text thus moves along two dimensions.

One of the interesting minor results of the investigation concerns the application of rule 38, P——PP. Although 26 has the highest frequency 38 comes at the bottom of the table. This means that Thomas Mann prefers to develop the phrase from one node. An adverbial phrase can be developed from a preposition and if at the start two or three prepositions are taken and their nouns developed a complicated structure will ensue, wherein the top rule would have been P——PP. But Mann avoids this. On the other hand Rule 36, N——NN (noun duplication) occurs with high frequency. We may illustrate this graphically :

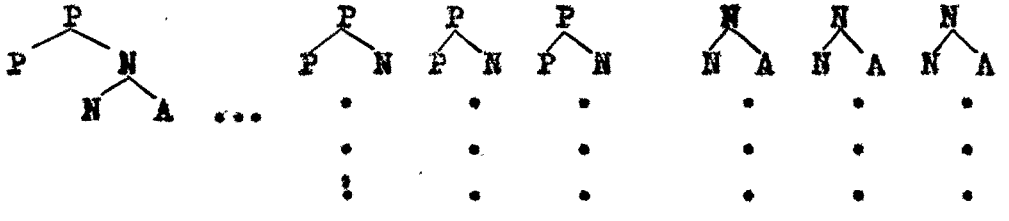

Table 3 shows that the prepositional rule 26 is applied roughly 1/5th of the times when a prepositional construction is started. Tables 10 and 11 show that the most prolific among the prepositions are in, mit, als, auf, von, aus, im, an. In Kaeding's count (249 , pages 145-150) these prepositions have the per mille values :

in	-	17'24
mit	-	8'39
als	-	5'34
auf	-	7'42
von	-	10'38
aus	-	3'72
im	-	4'65
an	-	5'10

but only in, and mit appear with maximum depth of 7 or 8 (vide Table 10). What semantic feature gives them this status among all the prepositions remains an intriguing question.

Distribution of nouns

Table 13 gives the frequency distribution of nouns grouped semantically. Here too interesting features can be observed; the parts of the body which occur with high frequency have the order : eyes, hand, head, hands, face. One may be justified in saying, this is the order of

importance which Thomas Mann assigns to the items.

Table 14 gives the 27 most frequent nouns in the samples I and II from BUDDENBROOKS and in KAEDING's count. Ranking the first eleven amongst themselves from 1 to 11 and calculating the rank difference coefficient of correlation, rho, given by

$$P = 1 - \frac{6x\Sigma D^2}{N\,(N^2-1)}$$

we obtain

$rho_{I,II}$	$rho_{I,Kae}$	$rho_{II,Kae}$
0'68	-0'255	0'05

In other words the two samples I and II from BUDDENBROOKS exhibit correlation as far as noun ranks are considered, but either of them shows little correlation with the sample of Kaeding. It is nevertheless interesting to note that out of the thousands of nouns which exist in a language some are conceptually so important that they occur in any type of population with high frequency. Such words are : Augen, Hand, Mann, Frau. Tables 16, 17, 19, 20 show some frequency distributions and also what happens to the Hapax Legomena when a portion of the text is being added each time. The sequence of losses in Hapax Legomena in one sample is (Table 19)

16, 16, 19, 39, 31, 35, 56, 35, 64, 38, 52, 28, 36, 32, 56, 42, 38, 32, 22, 43

After some time the loss reaches peak values, 56, 64, 52, 56, and then there is again a period where the loss is less (22). There seems to be unknown psychological factors operating here. The investigations of this aspect were however discontinued since it was found that even for a computer the repeated sorting involved in this process assumed gigantic proportions and called for too much computer time.