

ments. The situation is therefore by far more complex than what Documentarians imagined it to be, particularly since there may well be some sort of interaction between SDS and DIV and possibly also between these two and the hypothetical DOC. We actually calculated the interaction effects, but cannot afford to enter into such detail for lack of space. Instead, we had better proceed to stage (B).

In stage (B), we did not acknowledge any hypothesis, but let the 96 text samples group by themselves, so to say. A few words about this strategy must precede the results. While so far each of the three dimensions (DOC, SDS, DIV) was treated separately, a sample is now being defined at once by the Document, the Sort-of-Discourse and the Division it belongs to. It goes without saying that, in addition, samples are numbered consecutively from 1 through 96, that they are of equal length (200 words with fluctuations of +1 or 2%) and that a list is available of the limits of each sample in the text, for purposes of identification. These limits may in some cases be quite narrow and fall within one single chapter, and in others be spread over half the book. For example: sample no. 17 (NE) means that it belongs to the subcategory 'The Elohist as a Narrator' and comprises ch. 30<sup>14-41</sup>; sample no. 57 (HP) means that it belongs to the subcategory 'Human direct speech as reported by the Priestly Writer' and since such 'Priestly' direct speech is extremely scarce, this sample extends from ch. 27<sup>46</sup> right until the very end of Genesis; both samples are thus constituents of Division II. As tridimensional presentation, though feasible, is confusing on paper, we are going to neglect, for the time being,

DIV, and display the distribution of samples with regard to DOC and SDS alone. The following matrix obtained.

The columns represent categories of Documents, the rows, categories of Sorts-of-Discourse. Capital letters within the cells mark the nine subcategories. The number of words per cell is also indicated as well as that of pertinent samples (number of words divided by 200).

The first task of the statistician was to find out whether samples of the same subcategory behave linguistically in a consistent manner. The majority, it was soon established, indeed do, with one notable exception, though: NP. This case is far from surprising when we recall that of the 3379 words apportioned by Documentarians to the source P, a full 73% are of the N-category (see of Fig. 4) and that P comprises genres as disparate as ch. 5, a roster of names and numbers, and ch. 23, the account of Abraham's purchase of the burial cave.

To that matrix were now administered a whole range of statistical techniques such as Cluster, Factor, Reliability and other analyses. To explain here their theoretical background would again exceed the limits of an abstract, and without such explanations the results are incomprehensible. We shall consequently concentrate on one only, namely Cluster Analysis, and that because it is the easiest to understand and to display visually.

The basic idea of Cluster Analysis is the following. Recall first that each sample is characterized by the realizations of 54 variables. The analysis is capable to work out which pair of samples is of the highest mutual affinity, which of a slightly lesser and so on, an assignment which, considering the large number of characteristics, is plainly beyond the capacity of human diligence and the human brain. The program produces a triangular chart as shown in Fig. 5.

In Col. 1 on the left-hand margin, the nos. of samples are given, their sequence being repeated on the bottom line so that the result of the comparison between any two samples can be read at the point where the two coordinates meet. The darker this point, the greater is the similarity in language behaviour between the two comparands. Col. 2 indicates the subcategory and Col. 3 the Division which a sample originated in.

The mass of information contained in this chart is so immense that one could write a new kind of commentary on Genesis founded on it. Attention can be drawn to a few features only.

We observe, first of all, a small very dark triangle in the uppermost corner which, when the proximity signs below it are inspected, turns out to have almost nothing in common with the rest of the book. Its constituents are three NP-samples. When their origin in the book is identified, they are found to be no other than the three long genealogies in chs. 5, 6 and 11. That the program was able to reveal their uniqueness is highly gratifying and proves how well the set of formal criteria describes not only the linguistic comport, but also the literary genre of a given text.

	DOC J words samples	DOC E words samples	DOC P words samples	Total words samples
SDS N words samples	5539 NJ 27	3294 NE 16	2408 NP 12	11241 N 55
SDS H words samples	2849 HJ 14	3161 HE 16	360 HP 2	6370 H 32
SDS D words samples	1021 DJ 5	161 DE 1	611 DP 3	1793 D 9
Total words samples	9409 J 46	6616 E 33	3379 P 17	19404 96

Fig. 4