On the other hand, the distribution of words according to the number of phonemes must be conceived to be the result of the operation of all those constraints of phoneme groups which are characteristic for the language. In taking the word distribution according to numbers of phonemes as our basis for deriving the entropy and the redundancy, we work from a distribution which is to be regarded as the final result of the actual Markoff process, which was introduced in a rather simplified form when trying to correct $H$ for clustering of letters. We then obtain $H$ as 2.274. As the entropy of the phoneme groups which we call words, it is close to the entropy per phoneme if all the constraints on phoneme association and the spaces between words are taken into account, i.e. 2.3 bits.

The difference between the entropy calculated for independently associating phonemes and that calculated taking into account the relative frequencies of digrams, trigrams etc. of phonemes is a measure of the structural element in the language. It enables us to compare languages objectively in this respect, instead of having to rely on subjective impressions which will, of course, vary from observer to observer.

The distribution of Latin hexameters according to the number and arrangement of $s$ and $d$, which was described in Chap. 11 is a similar case.

For the basic distribution (Table 75, last col.) we obtain $H_p = 3.754$, whereas for that of combinations (Table 76, col.2) we have $H_r = 1.88[24]$. The same value of $H$ results if it is calculated from the theoretical values, col. 3. This means that the value of 1.88 for the entropy is compatible with the assumption that the numbers of $s$ and $d$ are subject to chance only.

However, if the fact is taken into account that there is a preference for $d$ in anterior or $s$ in posterior position, the entropy of the transition probabilities results as $H_b = 3.549$. The difference $H_b - H_r = 1.67$ is a measure of the influence of word structure upon the distribution, at least in first approximation, since only pairs of $s$ and $d$, not triplets etc. were taken into account. This difference is of the order of $H_p - H_r = 1.87$ and thus explains the observed discrepancy between the two entropies, the one from the combinations and the other from the permutations, as being due to word structure. In other words: whereas the numbers of $s$ and $d$ are subject to chance only, their arrangement is governed by choice.

17. Language Translations as Bi-Variate Distributions of Coding Symbols

In this part we consider the bi-variate distribution of language symbols for coding a message. Translating from one language into another may be conceived as having that structure. Every such translation can be
regarded as an instance of double coding. The identical message is expressed in two different language codes. It is first encoded in the ‘original’ language, from which the translator has to decode it previous to encoding it into that of the ‘translation’.

The bi-variate universe which we shall consider is that of word occurrences according to the number of syllables in both languages. This concerns directly only the quantitative aspect of translation, viz., the translation in terms of the number of syllables used in the encoding operation as against that used in the code of the original. Thus, the ‘goodness’ of a translation in the sense of close representation of the meaning of the original is not the immediate object of the statistical enquiry, and could not form such an object until the manifold of meanings is conceived as a statistical collective. But even so, the statistical inquiry could be of great interest to the linguist since the quantitative aspect of translation which is amenable to statistical treatment is an element of what is sometimes vaguely described as ‘flavour’, ‘atmosphere’, ‘recapture of the spirit of the original’ or as the absence of these characteristics. We are all familiar with the fact that of two translations, both of which may be equally correct renderings of the original thoughts, one may appeal much more to us than the other, because it somehow succeeds better in bringing home to us the peculiarity of the original. It has, in spite of the other being an equally good rendering of the original ideas in the language of the translation, retained more of the original than the other. The similarity between the two languages in the articulation of thought by means of syllables, and in the quantity of phonemes or syllables used for conveying distinction between thoughts, must undoubtedly have a great deal to do with this. We shall see that the theory of information enables us to express this aspect of translation quantitatively, and thus to arrive at objective estimates of the shares of freedom and constraint when translating from one language into another, and of the ‘goodness’ of a given translation, in respect of the quantitative characteristic under consideration.

17.1. Bi-variate information theory [2,3]

The bi-variate problem arises in communication theory as that of the disturbance of a signal through noise during the transmission. This means that the received signal is not necessarily the same as that sent by the transmitter. The change of the signal may be a systematic one. In this case, the received signal will be a definite function of the transmitted signal; that is, the latter will suffer a systematic distortion. The case, however, which interests us here is that in which the signal does not always undergo the same change in transmission, so that in the individual instance it is unpredictable whether, and what, a change will take place.