it, Life Insurance being one of the most successful business activities –, we speak of stability of the Life Table.

We shall deal with the stability of each type of linguistic distribution in turn. The stability of the distribution of letters, phonemes and grammatical forms will occupy us in Chapters 3 and 6; the distribution of word length in Chapters 9 and 10, that of vocabulary in Chapters 5 to 7.

3. Explanation of Stability of Linguistic Distributions

3.1. Overlap between texts
in vocabulary and frequency of occurrence

There are two explanations possible for the observed stability of linguistic distributions.

a) A sample may be similar to another, and to the total from which it was withdrawn, because of a high percentage of the same words, provided the relative frequencies were also similar in both collectives; or, b) in spite of little overlap in words, because different words may be conceived to be random arrangements of linguistic elements, letters or phonemes. These factors may operate jointly.

We shall deal with the first alternative in this Chapter and with the second in Chapter 9.

It is obvious that the phoneme distributions will agree as far as the two samples overlap in words and their frequency of occurrence. A common batch of vocabulary items plus occurrences must yield the identical frequency distribution of phonemes. The question now arises to what extent we may credit different samples of word counts, be it from different parts of one text, or from different texts or groups of texts, to have such overlap.

Some idea – though in a restricted way only – we may get from the correlation between authors (Part II), when, as in Sect. 7.4., we consider not the overlap in vocabulary only, but the correlation between authors as regards vocabulary and occurrence.

Let us take as an illustration the three correlation tables of Sect. 7.6, exhibiting the relationship between the 'Imitatio' and other works of THOMAS À KEMPIS, between the 'Imitatio' and works of GERSON and between works by THOMAS À KEMPIS and works by GERSON. We are here interested only in the vocabulary as far as it is used equally often in both sources. Looking at the tables, we see immediately which of the words belonging to the common vocabulary of the authors were also used the same number of times by them: they are the words entered in the diagonal of the table (Yule's tables are not quite symmetrical because the
scale of the distribution is different for the horizontal and the vertical ordinate beyond the frequency of 5, yet the lack of symmetry is not considerable).

The phoneme distribution of the words in the diagonal comprises the part of the total phoneme distribution which is common to both writers. Therefore, the proportion which this part of the vocabulary represents of the total common vocabulary may be taken as an index (not a direct measure) of the proportion of phoneme occurrences common to both writers.

<table>
<thead>
<tr>
<th>Sources</th>
<th>2</th>
<th>3</th>
<th>col. 3/col. 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Imitatio – à Kempis</td>
<td>803</td>
<td>308</td>
<td>38.3%</td>
</tr>
<tr>
<td>Imitatio – Gerson</td>
<td>808</td>
<td>261</td>
<td>32.3%</td>
</tr>
<tr>
<td>à Kempis – Gerson</td>
<td>904</td>
<td>302</td>
<td>33.4%</td>
</tr>
</tbody>
</table>

This gives some idea of the common component of the two phoneme distributions. But only for a special case, since Yule’s samples refer to nouns only and to texts with rather related content; also the proportions of column 4 depend to some extent on the grouping of the distributions which in Yule’s tables is somewhat coarse.

In order to be able to arrive at a more general conclusion about the common factor of style which produces a common part of the phoneme distribution, we must try to avoid such restrictions as far as possible.

From an extensive Russian word count carried out by Josselson* in the Slavic Languages Department of Wayne University on the frequency of occurrence of over 5000 different vocabulary items extracted from a great variety of sources, a list of 490 words (List 2 of the count) was subjected to the analysis needed for answering our question.

Here all types of words are included: nouns, verbs, adverbs, pronouns etc. (except the most frequently occurring words, mostly grammar forms); the sources are not confined to certain special subjects and are, moreover, chronologically different, and the grouping of the words in our table is, on the whole, by intervals of 10 occurrences.

We now compare the two frequency distributions for word occurrences of this part of the vocabulary (490 items) for the literary period of Russian

* See Ref. 27