Fig. 3

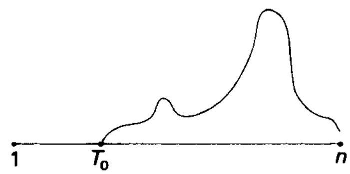

Fig. 4

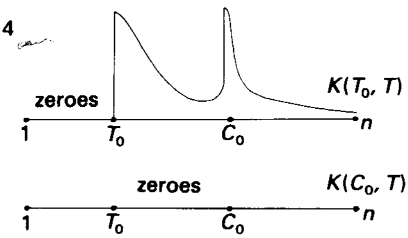

Figure 3. *Experimental name frequency graph.*

Figure 4. *Frequency graph of $K(T_0, T)$ and $K(C_0, T)$.*

5 The method of duplicate recognition

We now account for a new method of duplicate recognition in a text, and describe *the frequency duplicating principle.* Let an historical epoch (A, B) be described in some text X which is divided into individual chapters $X(T)$. Assuming that they have been numbered generally, in a chronologically correct manner, we suppose that there are two duplicates among the chapters, i.e. two chapters, or fragments of the text, speaking about the events of one and the same generation. In other words, these chapters repeat each other but are placed by the chronicler differently in the text X. Consider the simplest situation where the same chapter is repeated twice in X, numbered T_0 as a chapter $X(T_0)$ and C_0 as a chapter $X(C_0)$. We take $T_0 < C_0$. It is evident that the frequency graphs $K(T_0, T)$ and $K(C_0, T)$ have the form represented in Fig. 4. The first graph $K(T_0, T)$ clearly does not satisfy the frequency damping principle (two maxima). Hence, we have to permute the chapters of the text somehow to achieve a better agreement with the theoretical damping graph in Fig. 1. Furthermore, we can see that the second graph vanishes, that is $K(C_0, T) = 0$, which is explained by the fact that there are no new names appearing for the first time in the chapter $X(C_0)$: they all have already appeared in the earlier chapter $X(T_0)$. It then becomes evident that the best coincidence of the experimental graph with the theoretical is achieved when we juxtapose these two duplicates, i.e. chapters $X(T_0)$ and $X(C_0)$, or simply identify them with each other.

Thus, if the chapters of the text, generally speaking numbered chronologically correctly, contain two chapters whose frequency graphs give the form approximately represented in Fig. 4, then they are probably duplicates, and should be identified with each other. This is just what we call the frequency duplicating principle. A similar reasoning is also valid for the case of several duplicates in a text.

In the computer experiment performed by the author, the discovery of such double peaks (which correspond to the duplicates) of the frequency graph occurred as follows (in the first step of analysis). Let a_{ij} be the element of the matrix $K\{T\}$, placed in the ith row and jth column. Consider the matrix $\{a_{\alpha\beta}\}$, consisting of the elements $a_{\alpha\beta}$, where $\alpha \geq i$ and $\beta \leq j$, that is part of the large matrix $K\{T\}$ bounded by the ith row and jth column. We construct for it the averaged frequency graph $K^{ij}_{\text{av}}(t)$ by averaging the values positioned in the matrix $\{a_{\alpha\beta}\}$ on the diagonals parallel to the principal diagonal. We now assume that the ith and jth columns of the frequency matrix $K\{T\}$ correspond to the duplicates $X(i)$ and $X(j)$; that is $T = i$ or $T = j$. Then the averaged frequency graph $K^{ij}_{\text{av}}(t)$ has the form represented in Fig. 4; that is it possesses two maxima.

Then, marking all those elements a_{ij} (where $i < j$) in the large matrix $K\{T\}$, for which the averaged graph $K^{ij}_{\text{av}}(t)$ has such an anomalous form, we discover those chapters which may be duplicates. It was required in our computations that the averaged graph $K^{pq}_{\text{av}}(t)$,