Such efforts have been made before, but our method and results are new. Our results are, we believe, not surprising in the sense that they are subject to “natural” interpretation. For example, we find that separating the letters of the alphabet into vowels and consonants, as Markov did for his analysis, is proper in a very strong statistical sense for English. We are further able, roughly speaking, to “refine” the original separation into two classes by making more classes. We have succeeded in analyzing a separation for up to twelve classes. (These classes are not disjoint as will be seen.)

The text chosen for analysis was from the Brown University Sample of Present Day English. We have included word space as a twenty-seventh letter but have eliminated all case, punctuation and hyphenations.

All of this work was done at the Communications Research Division of the Institute for Defense Analyses.

II. The Type of Model

In order to analyze his text, Markov reduced the Russian alphabet to just two symbols, vowel and consonant, and explored the chains of symbols which resulted. We are more interested in the chains of English letters themselves, so that we must provide in our model means for generating letters. We could look upon the sequence of