STATISTICAL METHODS IN LEXICOLOGICAL RESEARCH IN THE BALTIC STATES

HILDA RADZIN

ABSTRACT

In this study "lexicology" refers to the dictionary form of a text, a vocabulary, without consideration of the lexical structure of an individual word and the system of its meaning.

The first attempts to apply statistical methods in lexicological research are connected with the computation of "frequency dictionaries", particularly various specialized lexica aimed at mechanical exploitation. A "frequency dictionary" is a list of words in which every word carries an indicator or its occurence frequenc in a text of a certain length. Frequency dictionaries permit one to compare words from the standpoint of their usage frequency. The data of frequency dictionaries are of great theoretical interest for studies of certain properties of the text with regard to its relation to the vocabulary.

In the Baltic states we find research work pertaining to compilation of frequency dictionaries. In Tallin, Estonia, work was started pertaining to compilation of Russian frequency dictionary for teaching purposes (1960). The project worked on a number of texts, amounting to 400,000 tokens. The completed dictionary reveals frequency distributions according to word-classes, government of case by verbs, prepositional phrases and also the substantival cases.

In other Baltic States we find linguists engaged in similar forms of study on the word level. On the basis of results gained from the statistical word count of Latvian technical-industrial texts, the first volume (pertaining to technical industrial field) of a frequency dictionary of Latvian language was compiled and published in 1966.

The statistical approach is used in the studies of style. The concept of "style" presupposes the presence of several properties inherent in a given text or texts or author, as opposed to others. One can propose that style is a sum of statistical characteristics describing the content properties of a particular text as distinct from others.

In the Baltic states we find linguists using statisical methods in the analysis of style. In scanning some prose texts of Latvian writer Blaumanis, K. Karulis found that words that Blaumanis uses in the direct speech are shorter than those in the indirect speech.

St. John's University
Jamaica, New York