

Preprint No. 63

Classification: IR 1.4

COMPUTATIONAL LINGUISTIC TECHNIQUES
IN AN ON-LINE SYSTEM FOR TEXTUAL ANALYSIS

- Donald E. Walker -

INTERNATIONAL CONFERENCE

ON

COMPUTATIONAL *LINGUISTICS



C O L I N G

1969



RESEARCH GROUP FOR QUANTITATIVE LINGUISTICS

Address: Fack Stockholm 40, SWEDEN

Computational Linguistic Techniques
in an On-Line System for Textual Analysis

Donald E. Walker
The MITRE Corporation
Bedford, Massachusetts

Prepared for the
International Conference on Computational Linguistics
1-4 September 1969 Sönga-Söby, Sweden

ABSTRACT

For several years we have been involved in the development of an on-line text-processing system intended for use by information analysts in the establishment and manipulation of their own personal files. In the initial implementation of the system, a transformational syntactic analysis was applied to sentences formulated by the analyst as summaries of information content in the text he was scanning on a display scope. This analysis procedure begins with a morphological analysis and a lexical lookup that provides syntactic feature information; then, a context-free parsing takes place; finally, transformations are applied to reject inappropriate parsings, derive the base structure of the sentence, and convert the result into a canonical tree format. These canonical representations are stored and can be searched by analyzing questions in the same manner and then matching their structures against those in the data base for correspondences. On-line interaction allows the analyst to advise the program in case of ambiguity, to expand the lexicon, and to modify previous actions taken.

Recently, we have added other, less sophisticated techniques to provide a range of on-line processing capabilities from simple to complex. The analyst can identify or annotate lines, paragraphs, or whole selections, creating files and subfiles for temporary or long-term storage. Text-searching procedures allow him to match on stems, words, or phrases and on combinations of any of these elements--with synonym substitutions. Sets of synonyms can be established on-line to provide a personalized thesaurus. The various components of the system can be used flexibly, and additional features can be added easily without disrupting the ones already established.

The augmented system is being used now in studies which will provide information on the differential utility of these techniques in relation to the tasks that text-oriented information analysts undertake.