CHAPTER 1

OVERVIEW

Over the past two decades, the Linguistic String Project of New York University -- in collaboration, for the past decade, with the Computer Science Department -- has developed a system for the processing of natural language text. This system comprises several grammars (including a broad-coverage English grammar, smaller grammars for teaching purposes, and specialized grammars for analyzing specific sublanguages), dictionaries, and a programming environment for developing and using these grammars and dictionaries. The programming environment includes

* The Restriction Language, a programming language for writing computer grammars of natural language.

* A Compiler to compile grammars written in Restriction Language.

* A Parser to parse and transformationally analyze sentences using a grammar written in Restriction Language.

* A Lisp interpreter which can control the operation of the Parser and process the output of the Parser.

* A set of utilities which operate on grammars and dictionaries, producing listings, inverted dictionaries, and cross-reference tables.

* A set of programs for word dictionary maintenance, including programs for updating the dictionary and for extracting subset dictionaries.

* A set of programs for manipulating texts prior to analysis: serializing sentences by paragraph and sentence number, creating concordances, and building random-access text files.

The current programming environment is coded almost entirely in standard FORTRAN-77 and operates on a Digital Equipment Corp. VAX computer. This code is an adaptation of earlier programs written in FORTRAN-66 for the Control Data 6600 [1]. The FORTRAN implementation is the third in a series of implementations of the Linguistic String Parser; the first two were in IPL-V [2] and in FAP [assembly language] for the IBM 7094 [3].

This manual is primarily a guide to the Restriction Language and the use of the Compiler, Parser, Lisp Interpreter, and word dictionary updating program. Please address any corrections or suggestions regarding this manual to R. Grishman, Computer Science Dept., New York University.

---

[1] R. Grishman, N. Sager, C. Raze, and B. Bookchin, The Linguistic String Parser. Proc. 1973 National Computer Conf., 427-434.

[2] N. Sager, J. Morris, M. Salkoff, and C. Raze, Report on the String Analysis Program. String Program Reports No. 1, New York University Linguistic String Project, 1966.

[3] C. Raze, The FAP Program for String Decomposition of Scientific Texts. String Program Reports. No. 2, New York University Linguistic String Project, 1967.