AUTOLING : AN AUTOMATED LINGUISTIC FIELDWORKER

by Sheldon KLEIN
Computer Sciences Department
University of Wisconsin
MADISON-WISCONSIN 53706
U. S. A.

Sheldon Klein, Barbara G. Davis, William Fabens, Robert G. Herriot, Bill J.
Katke, Michael A. Kuppin and Alicia E. Towster.

ABSTRACT

A system intended to act as a linguistic fieldworker via teletype
interaction with a linguistically unsophisticated informant has been designed
and is being programmed in extendedALGOL for the Burroughs 5500 and 8500
computers.

The system consists of the three major analytic components ; a program
for performing morphological analyses and deriving a segmentation algorithm for
sentences in any language ; a syntactic learning program that formulates context
free and context sensitive phrase structure rules (monolingual learning component
to be added later) ; and a machine translation program that learns to translate
in both directions between the query language (English) and the language of the
live informant via bi-lingual transformations.

The informant may be viewed as a fourth component, and is assumed
to be able to read and write English in standard graphemics, and to be able to
read and write his non-English language in a phonemic notation.

-1-

The analytic methods used by the AUTOLING are heuristic rather tha algorithmic, and hence do not guarantee complete success. The various components make use of protocols derived from the experience of live fieldworkers . The representations of these strategies are relatively separeted from the analytic mechanisms of the program and, accordingly, may be deleted, altered or incremented at the discretion of the system designers.

As a partial illustration of the system's operation, assume that the analystic process is at an intermediate stage, i.e. some morphological, syntactic and translation rules have been posited. Under the control of a monitor program, the system would test a nexly formulated syntactic rule by :

1- Generating (via usage of the new rule) a form implied by the
   grammar but never elicited from the live informant.

2- Translating the Hypothetical form into the query language
   (English).

3- Asking the informant via a teletype query to translate the
   Enlsglish output of step 2.

If the informant's reply is equivalent to the form derived from step 2, the newly posited rule is maintened in the tenative grammar (although subject to later check). If the informant's reply deviates from the output of step 2, the new rule is treated as suspect and subjected to further verification procedures. If the deviating reply contains new morphological material, the monitor program instructs the morphological analytic component to update its analysis. The machine translation learning component would also recheck and update its rules.

The program as a whole continues its interaction with the informant indefinitely, using the just mentioned strategies as well as others. At any given time the system designers may interrupt the field work process and ask the machine to list its corpus and current rules.