

## **CHALLENGER - The First Romanian Automate Translation System from English into Romanian**

Liviu-Virgil Ciortuz  
Fac. de Informatica, Univ. "A.I. Cuza"-Iasi  
6600 Iasi, ROMANIA

Anca Ignat  
Institutul de Matematica, Academia Romana, Filiala Iasi  
6600 Iasi, ROMANIA

**Abstract:** This paper reports our work during 1987-1990 at ITC - Romanian Academy, Branch of Iasi, namely the automate translation system from English into Romanian, CHALLENGER 1.0. It was based mainly on ATNs and it reached a demonstrative level.

The system was presented at "Zilele Academice Ie<sup>o</sup>ene", in 1989 and then to the national symposium on computer science INFO-Ia<sup>o</sup>i 1989, but no written paper reported it. Now, because we are preparing a second version of this system, based on HPSGs (Head-driven Phrase Structure Grammars introduced by Pollard and Sag in 1987), we decided to take the opportunity to (re)make this report for the "Development and Application Systems" Symposium in Suceava, 1994.

The Challenger automate translation system was in fact made out of two main sub-systems:

The first sub-system was a bilingual (in fact possibly multilingual) database managing system with

- a grammar describing the English and Romanian morphology, and
- a set of files containing lexical entries and different (derivational and interlingual) relationships between them;
- an automate procedure to generate (and eventually to edit/correct) derivational entries, starting from the root and the morphological category of a given word;
- an extension of the precedent procedure to linguistic database generation from the scratch (or text files) in a user manageable format, also modularly expandable. (About 2.000 words were first introduced in the database, and then we've extended it to about 5.000 words.)

The second sub-system of CHALLENGER did

- the English phrase syntactic analysis preceded by lexical and morphological feature decoration of words according to the database. In fact, the syntactic analysis was based on ATNs (Augmented Transition Networks, see Aho and Ullman, 1973), but unfortunately we've implemented only particular ATNs for different kind of sentences noun phrases, verb phrases, adjective phrases, prepositional phrases, etc., but no general procedure was incorporated for ATN definition and implementation;
- English to Romanian structure translation using some transformational rules (including person and gender agreement refining, relative position changing, etc.);
- Romanian text generation, starting from the syntactic structure obtained so far and using lexical entries, until the phrase linear level is reached.

The translated text complexity could be illustrated by the following text which served us as a demonstrative one:

*This is the first public demonstration of our achievements in computational linguistics. You will see, the smartest translation system in our country will be this. Who are the authors ? We are two young programmers who studied informatics to the university of our town. I am a clever boy, but she is a pretty girl. The system began to work at the end of the precedent year. Its first translation was : We wish you a very happy new year ! Surely, it is only the beginning. Thinking to the future development of the system, we invite you to see us the next year.*

*Aceasta este prima demonstratie publica a realizarilor noastre in lingvistica computationala. [Tu/Voi] vei/veti vedea, sistemul cel mai istet de traducere in tara noastra va fi acesta. Cine sint autorii? [Noi] sintem doi programatori tineri care au studiat informatica la universitatea orasului nostru. Eu sint un baiat inteligent, dar ea este o fata draguta. Prima sa traducere a fost: [Noi] iti/va dorim un an nou foarte fericit! Sigur, [el] este doar inceputul. Gindind la dezvoltarea viitoare a sistemului, [noi] te/va invitam sa ne vezi/vedeti anul viitor.*

CHALLENGER was implemented in the C programming language.

The files in the linguistic database were:

WORD records of different lexical spelling of words;  
 LEXICON lexical characterisation of words, alphabetically ordered;  
 DICTIONARY morphological characterisation of 'roots' of word families;  
 TRANSLATION 'root' words, from one language to another;  
 INDEX entries for index tables, to provide a rapid access to the other files.

The entry texts were introduced to the translation process either from keyboard or on ASCII files. The output was provided on screen and optionally on ASCII file. The grammars were given on a text file under a binary tree format.

**Further Work:** The work was stopped after our transfer in 1990 to the "A. I. Cuza" University of Iasi at the Computer Science Department, respectively the Mathematical Research Institute of the Romanian Academy, branch of Iasi, but as the first author continues his work in the Natural Language Processing domain, especially in the field of HPSGs, a second version of CHALLENGER will be elaborated in the near coming future.

**Acknowledgements:** Thanks to Dan Gâlea who morally supported us in doing this work, and to Gabriel Holban for having defining the initial file formats for the database files.

#### **Bibliography:**

1. Alfred V. Aho, Jeffrey D. Ullman, *The Theory of Parsing, Translation, and Compiling*, Prentice Hall, 1973
2. Alice L. Bădescu, *Gramatica limbii engleze*, Editura "tiințifică și Enciclopedică", București, 1984
3. Liviu Ciortuz, Anca Ignat, *A Computational System for Interactive Translation*, Personal communication, Zilele Academice leone, 19-25 oct. 1989
4. Liviu Ciortuz, *Linguistic Databases Creation and Managing*, Personal communication, INFO Iași, 1989
5. Alf Lombard, Constantin Gâdei, *Dictionaire morphologique de la langue roumaine*, Editura Academiei Române, București, 1981
6. Carl Pollard, Ivan Sag, *An Information-based Theory of Syntax and Semantics*, CLSI, 1987
7. Masaru Tomita, *An Efficient Parsing Algorithm*, Technical Report, Canegie-Mellon Univ., 1986