CS I dr. Ana-Maria Barbu

# Theoretical linguistics – Computational and applied linguistics

## ABSTRACT

From the beginning of my career, I led my scientific concerns towards formal linguistic theories, such as GB, Categorial Grammar or unification-based grammars: HPSG, LFG and Construction Grammar. At the same time, I participated, as a linguist, in computational-linguistics research projects within the National Institute of Informatics and the Research Institute for Artificial Intelligence, where I had the opportunity to combine theoretical linguistics with natural language processing. A result of this combination was the work based on my PhD thesis *Sintaxa determinatorilor. Analiză lingvistică şi aplicaţie computaţională* [Determiner Syntax. Linguistic Analysis and computational application], which approaches an absolutely new topic in Romanian linguistics and at the same time a very recent one in the international literature. The work deals with the syntactic functions of determiners and adjectives and their function-oriented order in the noun phrase, from an X-Bar theory perspective.

The PhD thesis foreshadowed the two main directions of my subsequent scientific activity: *theoretical linguistics* and *computational and applied linguistics*. Besides that, this thesis set the beginning of a modern research style based on data in big electronic corpora, instead of introspection style. Based on these two directions, I have elaborated numerous studies in *syntax* and *syntax-semantics interface*, as well as in *conceptual methods* and in building *linguistic resources*.

## 1. Theoretical linguistics.

### 1.1 Syntax

From the results obtained in the PhD thesis, two new original analyses, very important for formalisation, have come out. The first one, *Situaţii de acord ale adjectivelor pronominale* [Agreement situations of pronominal adjectives], is intended to solve native speakers' continuous hesitation between utterances of the type *multor altor probleme* [many-GEN other-GEN problems] and *multor alte probleme* [many-GEN other-Ø problems], by means of the documented argument that only the first determiner has to be case-marked. The second

analysis, *"Numeralele" un, o: o analiză critică în comparaţie cu niciun, nicio* [The numerals *un, o*: a critic analysis in comparison with *niciun, nicio*], demonstrates that determiners *un, o* cannot be (also) cardinal numerals, due to their syntactic distribution, which is essential for formalisation, in the same vein with *niciun, nicio*. In this sub-direction, I also mention an innovative study, *Construcţiile substantiv-substantiv. Atributul substantival în nominativ* [The noun − noun constructions. The nominal attribute in nominative], which proves, with data in corpus, that Romanian, and, very likely, other Romance languages as well, has developed a borrowed compound pattern (e.g. *cuvânt-cheie* [word-key ˈkeywordˈ]) into a new type of free and productive phrase (e.g. *substantivul centru de sintagmă* [the noun - head of phrase]).

### 1.2  Syntax-semantics interface

The researches in this sub-direction are the most recent ones and concern the syntactic realization of the valence elements. I mention here, as a notable result, the identification of a new phrase type that I called *relational coordination phrase*, which can occur both independently, and as a valence element of a relational noun (e.g. *mother-daughter* interaction). This phrase type has been surprisingly ignored despite its frequency in many European languages. The paper analysing this phrase type, *A data-driven analysis of the structure type 'man-nature relationship' in Romanian* has been received with interest by the international community and published in a prestigious volume. However the biggest realisation in this sub-direction is *Dicţionarul de contexte verbale* [Verbal Contexts Dictionary], which represents, in my opinion, a solid base for further research on many and diverse topics regarding Romanian verbs. The syntax-semantics interface offers a large unexplored field, inside which I intend to develop my forthcoming scientific activity.

### 2.  Computational and applied linguistics

### 2.1 Conceptual methods

This sub-direction of research regards my computational linguistics studies, brought out in collaboration or as a unique author, especially about methods of translation equivalence extraction from parallel corpora and word-to-word alignment. The results have been published in prestigious journals (ISI-indexed).

### 2.2  Linguistic resources (data bases)

In the natural language processing domain, the linguistic resources are absolutely essential. Therefore, in numerous projects I contributed to the building of big data bases for Romanian

language. I mention, in this respect, the inflected forms dictionary RoMorphoDict and that of syllabic forms, described in *Romanian Lexical Data Bases: Inflected and Syllabic Forms Dictionaries*, as well as the data base of phonetic forms of Romanian words, from which the work *Fonetica limbii române* „The Phonetics of Romanian Language", 4 vol., came out. Another outstanding resource for natural language processing is the XML data base representing a variant of the Verbal Contexts Dictionary. The works obtained from these data bases—to which I would add my book *Conjugarea verbelor româneşti* [The Conjugation of Romanian Verbs]—are very useful auxiliary materials for foreign students learning Romanian. Building new resources constitutes another research line for the future, which continues, for instance, my study *Monitoring Neologisms in Newspapers within The Neorom Project* on new aspects of Romanian language used online.

Beside the research activity, I had activities of student education and formation, by teaching courses and/or seminars of computational linguistics at the Faculty of Letters of the University of Bucharest, where I also organised in cooperation with other colleagues the master programme in Theoretical Linguistics. In the same vein, I participated in organising some summer schools, sessions of invited talks with well-known foreign teachers or I involved young researchers and students in my projects. My scientific recognition is sustained by hundreds of citations and mentions in international scientific data bases.