MATHEMATICAL LINGUISTICS AND COMPUTATIONAL SEMANTICS IN ROMANIA, IN THE PERIOD ’60-‘80

EMIL IONESCU

Besides reference works of Grigore Moisil and Solomon Marcus, which deserve a special analysis, mathematical and computational linguistics in Romania in the sixties are also represented by the pioneering works of two researchers, Erika Nistor and Eliza Roman. The environment of their contributions has two characteristics. The papers were published in reviews accessible only to a small scientific community; they mainly concentrated upon a specific problem, the one of the organization of information in a library (and, more generally, in a documentary center).

The interest of the two researchers for the library science is hardly surprising. A library is the ancestor of the present-day web, and problems of information retrieval and information processing, which are central today in the field of Information Technology, emerged first as challenges in the field of library documents. As for the fact that the contributions of the two researchers have been published in rather marginal publications (except in the eighties), this explains their limited audience. This article aims to do justice to these works and to emphasize – for those who are rather inclined to ignore the recent past – that the papers referred to here are part of the living history of mathematical and computational linguistics in Romania.

To put things in a chronological order, we ought to point out first to the interest of the two researchers for the so-called “languages of documentation” (Nistor, Roman 1968a). As is well-known, these languages are tools used in the process of information retrieval applied to a specific field, the one of the library organization. The book to which the authors make reference in this respect is the classic work of Maurice Coyaud (Coyaud 1966).

A very clear introduction to what was to be the main topics in computational linguistics may be found in Domonkos-Nistor and Roman (1967) (see also Nistor-Domonkos, Roman 1967). The reader discovers in these articles definitions of subfields, such as writing and speech recognition, text summarization, speech synthesis and machine translation, all along with an overview of the achievements made at that time in these subfields.

RRL, **LI**, *3–4*, p. 511–515, Bucureşti, 2006

BDD-A249 © 2006 Editura Academiei
Provided by Diacronia.ro for IP 216.73.216.75 (2026-10-30 02:01:02 UTC)

In the same line, Nistor and Roman (1968b) offer an outstanding survey of the major problems in computational linguistics in the sixties: information retrieval, Chomsky's hierarchy of formal languages and their relevance in machine translation, text summarization and problems of programming. This time, the work is a more comprehensive and systematic presentation, and resumes issues already outlined in Nistor and Roman (1968a).

Another synthesis (Nistor, Roman 1978) focuses on the main issues of Information Retrieval (with special emphasis on the field of library organization). What is new in this work is the use of the concept of fuzzy set, as a mean of investigating classes of documents.

The eighties are characterized by more attention paid to the problems of computational semantics and natural language theory. For instance, in Nistor and Roman (1980) it is proposed a computational device of detecting semantic deviations at the level of clauses. To this purpose, the authors use a numerical encoding of the properties of a concept specific to a word or phrase. The whole concept thus becomes the product (in the arithmetical sense) of these numerical encodings. A semantic deviation takes place with respect to a certain subject noun or NP, if and only if the product which encodes the properties of the concept expressed by the subject and the numerical code which represents the concept of the predicate divide with a remainder. For example, if one states that **the river Mureş flows into the Danube** (which is false and represents a form of semantic deviation), the numerical encoding of the concept of the river **Mureş** (the number 329 in the system of the authors) has to be divided with 2 (that is, with the number which represents the numerical encoding of the predicate …**flows into the Danube**). Since this division is without exact quotient, the statement is declared semantically anomalous, and marked as such. Unfortunately, this device does not work any more if it is coupled with negation.

Another interesting contribution of computational semantics is proposed in Nistor and Roman (1982). It is a research, dealing this time with aspects of the logical form of sentences, more precisely with entailments and presuppositions tied to the tense of certain verbs. The verbs subject to the analysis are the ones which, if used in the present tense (noted t) allow for inferences concerning both a time interval prior to t (noted t-1), and a time interval subsequent to t (noted t+1). Such a verb is for instance **to entry**. If used in the present tense t (for instance in the sentence **John is entering the room**) the verb presupposes that, priory to t (= t-1), John isn't in the room. Also, the verb allows for the temporal entailment that, subsequently to the same moment t, (= t+1), John is (will be) in the room. The import of this type of inferences can be hardly neglected, in spite of their apparent triviality. Speakers of a language use them all the time and obtain important information about facts which concern them.

BDD-A249 © 2006 Editura Academiei
Provided by Diacronia.ro for IP 216.73.216.75 (2026-10-30 02:01:02 UTC)

The novelty of the approach consists in the algorithm which simulates the natural computation of such inferences. The input of the program is represented by sentences containing verbs that are susceptible of supplying inferential information about moments which are prior and subsequent, respectively, to the moment described by the present tense of the verb. The output is the expected inferences. The algorithm, therefore, is able to compute automatically temporal presuppositions and entailments.

The structure of the algorithm consists in a number of ordered steps, as follows:

- The algorithm is given in the sentence for which it has to compute the expected inferences. For instance the French sentence **Jean s'assoit** ("John is sitting").
- The verb is picked up out of the sentence and analyzed, that is, it receives a category according to the categories which form a part of the data base of the program. In the situation of the verb **s'assoir**, this category is labeled 1.
- The category of the verb is associated with the corresponding list of symbolic arguments, which represents another part of the data base. The list of symbolic arguments of the verb **s'assoir** is the following symbolic structure: *S+est+argument* (where S = subject, *est* = is and *argument* = a variable for argument).
- The category of the verb is also associated with a network in which the verb lies. The network is a finite state automaton, in which every verb is assigned a number of states corresponding to the two moments t+1 and t-1. The verb **s'assoir** for instance is assigned three states for each time interval. The meaning of this assignment is that in this way one prepares the generation of the inferences appropriate to this verb. We already know therefore that temporal inferences associated with **s'assoir** are three sentences for t-1 and three for t+1. We also know that the form of each of these inferences is *S+est+argument*.
- One replaces the variable S with the appropriate NP subject, in the present case **Jean**. One thus obtains three sentence schemes for either of the two temporal moments. The pattern is **Jean**+*est*+*argument*
- The states reached by the verb also contain information about (verb) negation. This means that if the verb **s'assoir** is associated in the network with the state in which we have three sentential schemes for t-1 and another three for t+1, we will also get the information that some, all or none of these schema shall have a negative counterpart (obviously, this does not hold for all the verbs). Thus for **s'assoir**, the following association with negation is obtained. For t-1: two schemes are negated. For t+1: two schemes are negated (where negation takes the form **Jean**+*n'est*+*argument* ("Jean is not argument")).

BDD-A249 © 2006 Editura Academiei
Provided by Diacronia.ro for IP 216.73.216.75 (2026-10-30 02:01:02 UTC)

- The last operation is the replacement of the variable argument with the appropriate constant. The constant is a word or a phrase. For **s'assoir** the constants are **couché assis** and **à pied**:

t-1
Jean n'est pas *couché.*
Jean n'est pas *assis.*
Jean est *à pied.*

t+1
Jean n'est pas *à pied.*
Jean n'est pas *couché.*
Jean est *assis.*

What is thus obtained is the set of temporal inferences allowed by the verb **s'assoir.**

Finally, one has to mention the research in Nistor and Roman (1983). This research continues and extend the one in Nistor and Roman (1982). It approaches the same problem of inferences, but deals this time with inferences of location. The input of the program are sentences with two distinct syntactic structures:
NP, V, NP (for example, **John met Mary in the school**)
NP, V, PP[locative] (for instance, **John sleeps in the school**)

The PP **in the school** is privileged in this account, for reasons of simplicity in the analysis. This means that the authors prefer to work with a single locative complement. Clauses of the type described above allow for certain entailments which regard the place of the action. For instance, from **John met Mary in the school**, one can draw the entailment that both John and Mary were in the school. This does not hold with respect to a sentence like **John imagined Mary in the school**, where in one of its readings no participant in the imagining situation is bound to be in the school. So, it is obvious that different verbs may have different entailments with respect to the location of the situation they describe.

It is just with these entailments that deals the algorithm built by the authors. It is, therefore, an algorithm designed to automatically compute the location inferences for verbs incorporated in its data base.

In order to perform this task, the program is fed with four types of information:
Information concerning the subject NP.
Information concerning the object NP.
Information concerning the verb.
Information concerning the locative PP.

These are the ingredients crucially involved in computing the inferences. The program also contains a number of 28 inference patterns. When a certain verb is given, in a sentence, the program assigns it the appropriate inference and truth value. This is the output of the program.

BDD-A249 © 2006 Editura Academiei
Provided by Diacronia.ro for IP 216.73.216.75 (2026-10-30 02:01:02 UTC)

A word about external conditions of work in that time. Contributions like the one presented above have been produced in difficult – sometimes even hostile – conditions. It was almost a hero effort to get update bibliography. As for scientific contacts with foreign researchers, this was almost impossible, in any case, very risky. Although there is no heroical history of science but just history of science, these conjunctures have to be recalled, in order to evaluate the scientific results in their correct perspective.

REFERENCES

Coyaud, M, 1966, *Introduction à l'étude des langages documentaires*, Paris, Klincksieck.

Domonkos-Nistor, E., E. Roman, 1967, "Încercări de automatizare în domeniul recunoaşterii, analizei, completării, rezumării şi traducerii textelor", in *Studii şi cercetări de documentare şi bibliologie,* 49–57.

Nistor-Domonkos E, E. Roman, 1967, "Despre programarea la maşina de calcul electronică a lucrărilor de bibliotecă"*, in Studii şi cercetări de documentare şi bibliologie*, 1, 445–452.

Nistor, E., E. Roman, 1968a, "Limbajele documentare", in *Probleme de Informare şi Documentare*, vol 2, Bucureşti, 574–586.

Nistor, E., E. Roman, 1968b, *Modele în documetare şi biblioteconomie*, Bucureşti*,* Institutul Central de Documentare Tehnică.

Nistor, E., E. Roman, 1978, *Progrese în automatizarea clasificării, prelucrării şi difuzării documentelor,* Bucureşti, Institutul Național de Informare și Documentare.

Nistor, E., E. Roman, 1980, "Attempts at Automatic Detection of Some Semantic Deviations", *Revue Roumaine de Linguistique, Cahiers de linguistique théorique et appliquée*, XVII, 2, 165–170.

Nistor, E., E. Roman, 1982, "Réseau sémantique à trois temps", *Revue Roumaine de Linguistique, Cahiers de linguistique théorique et appliquée,* XXVII, 2, 145–152.

Nistor, E., E. Roman, 1983, "Analyse du rôle du sujet et de l'objet dans un système automatique d'inférences", *Revue Roumaine de Linguistique, Cahiers de linguistique théorique et appliquée*, XX, 1, 59–66.

BDD-A249 © 2006 Editura Academiei
Provided by Diacronia.ro for IP 216.73.216.75 (2026-10-30 02:01:02 UTC)