Brief Outline of the School Activities in Mathematical and Computational Linguistics in Bulgaria in the 1980s and 1990s

This is an outline of the high school activities in Mathematical and Computational Linguistics in Bulgaria in the eighties and nineties, as they were when I was in charge of them. I had the privilege to initiate these activities in 1983. All started when the first Olympiad for high school students was organised in May 1983 as part of a Mathematical Linguistics seminar. In the years which followed I remember travelling to a number of schools all over the country to get teachers and students interested. I never expected that a few years later, these activities would become a well-organised national network involving a large number of high school students, high school teachers, university students, university/research institute staff and lecturers as well as professional organisations, with their regular activities including competitions, summer schools and seminars, newsletters and other supporting materials.

1. Objectives

The main objective of these activities was to get school students interested initially in linguistics but later and more importantly in computational linguistics. We had the pleasure of working with very academically motivated and talented students with interests and excellent performance in Mathematics and Languages.

2. Types of activities

It started with competitions in mathematical and computational linguistics, but as seen below, there were a variety of additional events/ways to support these.

3. Competitions/Olympiads in Mathematical and Computational Linguistics

In the 1980s and 1990s at least two (but in most years 3 and more) competitions took place. The largest-scale event was the National Competition in Mathematical and Computational Linguistics which took place in May/June every year (initially in Plovdiv, and as from 1990 in Haskovo). Each school could participate with 4 students; I remember competitions with a total of more than 220 students participating. Another annual competition would take place in January as part of the traditional competitions in Mathematics and Computer Science 'Winter Mathematical Days'. For several years (1985-1989) a competition in Mathematical and Computational Linguistics was a regular part of the annual Mathematics contests of the language high schools in Bulgaria. In addition, regular 'exhibition competitions' were organised as part of the autumn seminars for teachers (see section 7 below). There were also competitions for younger students (see section 5 below). Several international competitions took place as well, with teams from Russia, Poland, Kazakhstan and other countries (1988, 1989, 1994 are the years that I remember clearly when such international events took place). The winners from Bulgarian competitions travelled to several times to Moscow to participate in the Russian Linguistic Olympiads (starting in 1990). Moscow students also participated at one of the summer schools and at least at one of the National Competitions.

The competitions offered 3 problems (a total score of 20; usually one of the problems was easier and had a weight of 5 marks; the second and the third one usually carried marks of 7 and 8) which the students had to solve within 4 hours. The students had to find and enter not only the

expected answers (e.g. sentences to be translated in unknown languages, number names to be written, rules to be formulated etc.) but also had to explain their reasoning - how they had arrived at these answers. The papers were marked by the National 'Committee for extracurricular activities in mathematical and computational linguistics' which I led and which consisted of either other fellow researchers from the Laboratory of Mathematical Linguistics, Institute of Mathematics, Bulgarian Academy of Sciences and increasingly in the last several years of the 1990s, of motivated university students who had usually been successful at such competitions as high school students. High school teachers were not part of this committee, also termed 'jury', because of conflict of interest associated with their school participating.

The competitions were based on both individual and team participation/performance (each school participated with 4 students). The results were made known the following morning, where the winners were awarded prices. There were two types of winners: individual winners and team winners where the schools were ranked according to the aggregates of their individual participants' scores.

4. Types of problems

We designed different types of problems ranging from purely linguistic problems such as translation correspondences through mathematically orientated problems such as these related to different number names systems to problems which had a distinct computational linguistics flavour. More specifically, the problems were classed as follows:

(i) translation correspondences

A typical example of such problems would include several sentences in unknown languages with their translation in a known language (in our case Bulgarian). Based on this data, the students had to find the lexical correspondences and arrive at specific syntactic or semantic rules, which would help them translate a set of given sentences from Bulgarian into the unknown language. A usually simpler problem of this class represented a scenario where a list of words were given in unknown language and a list their translations was provided, the task being to find the exact correspondences.

(ii) number name systems

These problems offered the students a list of number names in a specific language so that they could work out the base on which the number name formation was done (e.g. vigesimal system, (base 20) as in the case of Mayan) and in general, find out rules used in the formation of number names. In addition, the students had to write down the number names for a given list of numbers.

(iii) encoding and decoding

These problems normally involved the encoding of a text (usually in Bulgarian) and the student had (based on linguistic clues or constraints) to decode the text.

(iv) calendar systems

The students were given a list of dates represented according to a specific calendar, had to work out the system and had to represent a list of given dates according to this calendar system. One particular problem that I remember discussed the system of forming the dates in an ancient Japanese calendar based on the 'circle of animals'.

(v) problems with Computational Linguistics orientation

My goal was to offer as many problems as possible with Computational Linguistics / NLP orientation as I firmly believed that these school activities could be a promising springboard for future careers in NLP. As a rule, at least one of the problems offered in the nineties was of this type with some competitions featuring even 2 out of 3 problems discussing a Computational Linguistics topic – of course in a very accessible, simplified, even naïve fashion so that they would not be too much of a challenge to the students. Wherever possible, the problems with Computational Linguistics orientation were presented in an entertaining way for the students.

It is worth mentioning that on a couple of occasions problems with a Mathematical Linguistics flavour were offered – the students had to go through a very simple formal grammar or a finite automata model for Bulgarian.

(vi) other (usually entertaining) problems of logical-linguistic nature

This class of problems represented entertaining problems where logical (some mathematical) thinking was needed, coupled most of the time with linguistic decisions. From the sample problems attached, problems 3 and 7 fall into this class.

It was essential that the problems were 'self-sufficient' in that they contained all the information necessary for their successful solution. It was not possible to assume that high school students had background in computational linguistics, or linguistics or anything related, given the different curricula of the high schools.

5. Age groups

We had usually two age groups: high school students from class 8 on (from 14 to 18 years old) and secondary school students from class 5 to class 7 (from 11 to 13 years old). The problems for the younger age group fell more or less in the same categories as the ones listed above – however, they were considerably easier and adapted for this age group.

6. Summer schools

A highlight of the activities at the end of every school year was a summer school on mathematical and computational linguistics where the most successful students were invited (in fact, our goal was to secure the participation of as many students as possible). Usually students had lectures in the morning given by leading Bulgarian and international speakers and practical activities in the afternoon (e.g. techniques for solving problems). The summer school would often feature an internal competition as part of its programme. These summer schools were usually 2 weeks long and took place in picturesque locations in the Rhodope mountains (the first summer school took place in 1984 and the last one in the early nineties). In addition to attending useful courses which enhanced their background in linguistics, mathematics, computer science in general and in computational linguistics or mathematical linguistics specifically, students were attracted by the social nature of these schools as they made friends with other students.

7. Autumn seminar

The summer schools mentioned above were mostly organised for high school students. I felt the necessity to organise an event which would benefit the teachers mainly. At this one-week seminar the teachers attended lectures given mainly by the 'committee for extracurricular activities in mathematical and computational linguistics'. However, there was special emphasis on training the teachers how to organise and lead a club for extracurricular activities in mathematical and computational linguistics.

8. Local clubs

In most regional towns interested teachers organised 'clubs for extracurricular activities in mathematical and computational linguistics' whose task was to attract students to be involved in the competitions. Such clubs would take place once or twice a week and the teacher would offer the students problems for practice.

9. Organisations

The activities were supported by a number of national and regional organisations. It is worth mentioning above all the Union of Bulgarian Mathematicians who have been active over the years in working with able students. I would also like to mention the National Commission for Olympiads, the Regional Councils for Education in a number of Bulgarian towns as well as the National Ministry for Education.

10. Support materials

It was necessary to produce supporting materials for both teachers and their students – this was a must and was emphasised at all major events. In addition to the numerous handouts and topical papers (e.g. a paper which looks at a specific type of theory/problems), in the late 1980s and early 1990s I wrote two books (unfortunately in Bulgarian) which represented collections of problems (with solutions) grouped in chapters by type with each chapter featuring an introduction explaining the methodology and main techniques to be used in solving the problems. The second book was an improved, revised and extended version of the first book and these are still used by teachers and students today. Before these books, in 1985 I wrote another book on number name problems. A student journal/newsletter called 'Mathematical and Computational Linguistics' was launched in 1997. The journal featured articles, problems to solve including new problems offered by the students themselves! Unfortunately, several issues of the journal were published only.

Ruslan Mitkov, Wolverhampton, 7 September 2006