

Establishing a Linguistic Olympiad in Spain, Year 1

Estableciendo una olimpiada lingüística en España, año 1

Antonio Toral[†], Guillermo Latour^{*}, Stanislav Gurevich[△],
Mikel Forcada[‡], Gema Ramírez-Sánchez[◇]

[†] NCLT, School of Computing, Dublin City University, Ireland

^{*} IES Gran Vía. Alacant, Spain

[△] IOL Board and Liceo Anichkov, Saint Petersburg, Russia

[‡] Dep. Llenguatges i Sistemes Informàtics, Universitat d'Alacant, Spain

[◇] Prompsit Language Engineering, S.L., Elx, Spain

oleinfo@abumatan.eu

Resumen: Presentamos la OLE, una actividad de divulgación del proyecto europeo Abu-MaTran cuyo objetivo es establecer una olimpiada lingüística en España. Damos una introducción a las olimpiadas lingüísticas, presentamos nuestra motivación, nuestros objetivos y el plan de implementación. Pensamos que nuestro trabajo puede ser útil para otros países que quieran empezar una olimpiada lingüística.
Palabras clave: divulgación, lingüística, olimpiada

Abstract: We present OLE, a dissemination activity of the EU Abu-MaTran project which aims to establish a linguistic Olympiad in Spain. We introduce the Linguistic Olympiads, our rationale and objectives for setting up OLE as well as our implementation plan. We foresee our work to be useful for other countries looking to start a Linguistic Olympiad.

Keywords: dissemination, linguistics, olympiad

1 Introduction

The Linguistics Olympiad is a competition in which second level students are challenged to solve problems in the field of linguistics. It should be noted that the students do not need to have any prior knowledge of languages or even linguistics, as the puzzles are expected to be addressed with problem solving techniques such as logic and lateral thinking.

Most problems present samples of text in a language that the student does not know and involve finding certain patterns arising in these samples (and probably identifying exceptions too). In this regard, the problems of the Linguistic Olympiad can be considered as closely related to the field of computational linguistics (Littell et al., 2013).

The Linguistic Olympiads originated in the 1960s, and were initially run in eastern European countries such as Russia and Bulgaria. Recently the Olympiads have spread to other areas such as Asia, western Europe and North America (Radev, Levin, and Payne, 2008). Moreover, since 2003 there is an annual International Linguistics Olympiad¹ (IOL), which is attended by the

winners of the national competitions. The last edition of the IOL was held in Manchester (United Kingdom) with 138 contestants in 35 teams, representing 26 countries. Until now, Spain has not taken part in the IOL (except for an ad-hoc team² in the 2007 edition).

This paper discusses the establishment of a sustainable Linguistics Olympiad in Spain (OLE).³ OLE is a current dissemination activity of Abu-MaTran,⁴ a research project in the field of machine translation funded under Marie Curie's Industry-Academia Partnerships and Pathways (IAPP) action.⁵ IAPP projects focus in transfer of knowledge between academia and industry.

The rest of the paper is organised as follows. In the following section we cover the

²Ad-hoc teams are teams from a country where no national competition exists that compete in the international Olympiad.

³OLE stands for Olimpiada Lingüística d'Espanya (Catalan), Olimpiada Lingüística de España (Spanish), Espainako Olinpiada Linguistikoa (Basque) and Olimpiada Lingüística de España (Galician).

⁴<http://www.abumatan.eu/>

⁵http://ec.europa.eu/research/mariecurieactions/about-mca/actions/iapp/index_en.htm

¹<http://www.ioling.org/>

rationale and the objectives of the OLE. Subsequently we outline the overall plan to implement the activity. This is followed by an account of the first edition of the olympiad. We then report on the exercise types used. Finally we draw conclusions and plan our future lines of work.

2 *Rationale and Objectives*

There are two main reasons that encouraged us to carry out this project:

- As mentioned before, there has not been a Linguistics Olympiad in Spain so far. Linguistic Olympiads promote problem solving skills such as logic, lateral thinking, etc. which we deem of paramount importance for students. The fact that the performance by Spanish students in the latest PISA test on problem solving was significantly lower than OCDE's average (477 points vs 500)⁶ seems to support our hypothesis that more emphasis on problem solving is needed.
- The Abu-MaTran consortium is in a good position to run this project as one of the partners, Dublin City University (DCU), has extensive experience in the area. In fact, DCU organises the All Ireland Linguistics Olympiad⁷ annually since 2009.

As previously mentioned, the Abu-MaTran project focuses on intersectoral transfer of knowledge. In the case of the OLE, this implies the transfer of organisational knowledge from DCU's experience to Prompsit, the industrial partner of the consortium, based in Spain.

The main objectives of the OLE can be summarised as follows:

- Foster the acquisition and mastering of problem solving techniques by the participant students.
- Develop the interest of students in the area of linguistics and in the knowledge of new languages.
- Get students acquainted with the area of computational linguistics and related

⁶http://www.mecd.gob.es/inee/Ultimos_informes/PISA-2012-resolucion-de-problemas.html

⁷<http://www.cngl.ie/ailo/>

disciplines such as linguistics, translation and computer science. In this regard, OLE is contributing to the area of computational linguistics in the long term by bringing this area to the next generation of researchers.

3 *Plan and Implementation*

The aim of the Abu-MaTran project with respect to OLE is to set up a sustainable Olympiad in Spain over the course of the project (January 2013 – December 2016).

In this respect, the plan is to run three annual Olympiads during the second (2014), third (2015) and final year (2016) of the project as follows:

- First edition (2014). Pilot task, targeting the area of Alacant/Alicante province (1,917,012 inhabitants, 5,816 km²).
- Second edition (2015). The area is extended to the Valencian Community (5,111,706 inhabitants, 23,255 km²).
- Third edition (2016). The area is extended to Spain (46,815,916 inhabitants, 505,992 km²). Starting with 2016 we hope to proceed with that level further on.

This iterative approach allows us to adopt initially the organisative model of AILO for the two first years, since our target area is similar (Valencian Community) or smaller (Alacant/Alicante) to that of AILO, Ireland (6,378,000 inhabitants, 84,421 km²). During these first two years we are gaining experience that will allow us to adapt the model as necessary in order to tackle successfully and efficiently our whole target area, Spain, substantially bigger than that of our initial organisative model.

4 *First Edition*

We now discuss in more detail the first edition of OLE, carried out from September 2013 to July 2014. The main phases have been the following:

- September 2013. Development of the website of OLE,⁸ as well as its corporate image (logo, font, colours, etc) and the relevant materials (e.g. brochure).

⁸<http://ole.abumatran.eu>

- October–December 2013. Registration period for interested schools. Two sets of training exercises were prepared and sent to registered schools.
- January 2014. First round, carried out in each school. The top 75 students qualify for the second round.
- March 2014. Second round, carried out at Miguel Hernandez University (Elx), where Prompsit is based. The top 4 students qualify for the IOL final.
- June 2014. Training session for students that have qualified.
- July 2014. The top 4 students from the second round represent the OLE at the final of the IOL (Beijing, China).

The reception of OLE has been already very satisfactory in its first year. 20 schools registered to take part in OLE's first edition. From these schools, over 400 students took part in the first round.

5 Exercises

We report on the exercises used in the first edition of OLE.⁹ Exercises can be classified in different types, according to the linguistic phenomena they are about. Exercises can also be classified according to their level of difficulty.

Individual tests (training, first round and second round) in OLE are made up of a number of exercises (6 in all our tests). We have designed the tests so that they have exercises of different levels of difficulty (and accordingly different punctuation) and the exercises belong to different types. It should be mentioned that it is almost impossible to solve all of the problems in the time provided (3 hours). Therefore, time management is a relevant skill to tackle the tests. It is up to the student to manage their time to maximise their chances to obtain high marks.

Apart from these tests, which are tackled by students individually, we have also designed a group test (for the second round). The group test consisted of one exercise, considerably more difficult than the exercises in the individual tests. In the group test, students are expected to apply techniques of work in group (e.g. divide the problem in simpler tasks).

Most of the exercises used in the training and the two rounds are previous years' exercises from other chapters of the IOL (Ireland and Russia). They were adapted and translated accordingly (from English and Russian into Catalan and Spanish). In this respect, we should note that the collaboration with other chapters of the IOL was very beneficial, as it allowed us to use exercises of good quality and appropriate difficulty.

For illustrative purposes, we provide details of the languages and topics of the exercises used in the first and second rounds of OLE's first edition (see Table 1). Furthermore, an exercise is shown in Appendix A.

6 Conclusions and Future Work

We have presented OLE, a dissemination activity of the Abu-MaTran project which aims to establish a sustainable linguistic Olympiad in Spain. We have introduced the Linguistic Olympiads, our rationale and objectives for setting up OLE as well as our plan for implementing it. We have then reported on the first edition of OLE, corresponding to the academic year 2013–2014. Finally, we have given a detailed account of the exercise types and the tests of the Olympiad.

The outcome of the first year of the Olympiad is very positive, with over 400 students taking part. As a side effect, we foresee our current work to be useful for other countries looking to start a Linguistic Olympiad.

Looking into the near future, we will face challenges due to the plan to extend the area covered by the Olympiad. In this regard we are looking at how to adapt our organisative model to be able to run the Olympiad while staying within our limits (budget and manpower). On-line tests (used e.g. by the North American chapter of the Olympiad) and a distributed Olympiad (as done e.g. by the Russian chapter), among others, are models that we consider exploring.

As another corresponding activity we may be looking forward to introducing special courses for high school students as well as organising vocational training camps (as it has been done for many years in Bulgaria, Russia, Estonia and some other countries) where students could get preliminary acquaintance with the basics of linguistics, a discipline which is to a very subtle extent included into traditional school educational patterns.

⁹http://ole.abumatran.eu/?page_id=8

Language	Topic
Invented	Numbers, order of number positions
La-Mi	Syllable alternation
Unua	Translation, declination and order of sentence constituents
Amharic	Translation, morphology
Japanese	Adjectives, declination
Panyabi	Translation, tones
Latvian	Conjugation of verbs
Aroma	Numbers, morphology
Mundari	Translation, morphology
Tokhari	Translation, imperative verbs
English	Numbers, numbering system and rhyme
Turkish	Translation, morphology

Table 1: Language and topic of each of the exercises of the first and second rounds of OLE’s first edition

Acknowledgements

The research leading to these results has received funding from the European Union Seventh Framework Programme FP7/2007-2013 under grant agreement PIAP-GA-2012-324414 (Abu-MaTran).

We would like to thank organisers of the All Ireland Linguistic Olympiad (Cara Green, Harold Somers and Laura Grehan) for their advice in setting up a Linguistic Olympiad and for allowing us to use their exercises.

Similarly, we would like to thank the organisation (Polina Pleshak) and authors of problems of the Russian chapter of the Linguistic Olympiad for allowing us to use their exercises, assisting us with their translation and, most of all, for attending OLE’s second round and helping us with its organisation.

We also thank institutions for having given their support for OLE’s first edition, namely the Department of Software and Computing Systems of the University of Alicante, Fundación Quorum and the Vicerrectorado de Investigación of the Miguel Hernández University.

References

Littell, Patrick, Lori Levin, Jason Eisner, and Dragomir Radev. 2013. Introducing Computational Concepts in a Linguistics Olympiad. In *Proceedings of the Fourth Workshop on Teaching NLP and CL*, pages 18–26, Sofia, Bulgaria, August. 8 pages.

Radev, Dragomir R., Lori S. Levin, and Thomas E. Payne. 2008. The

North American Computational Linguistics Olympiad (NACLO). In *Proceedings of the Third Workshop on Issues in Teaching Computational Linguistics*, TeachCL ’08, pages 87–96, Stroudsburg, PA, USA. Association for Computational Linguistics.

A Sample Exercise: Complicated Numeration System

There isn’t any language that uses such a complicated numeration system, but let’s imagine that there is a language with numbers such as:

1 nut
 2 dok
 3 tris
 4 kwat
 5 kwin
 6 ses
 7 sep
 8 ok
 9 nou
 10 des
 100 hun
 1000 mil
 11 desnut
 14 deskwat
 20 dokdes
 32 doktrisdes
 47 sepkwatdes
 185 hunkwinokdes
 237 dokhunseptrisdes
 1234 dokhunkwattrisdesmil
 4567 kwinhunsepksesdeskwatmil

Task. Write the following numbers:
 78, 874, 3210, 215