The North American Computational Linguistics Olympiad:
Preparing high school students for the study of human language and computation

Summary

The North American Computational Linguistics Olympiad (NACLO, http://www.naclo.cs.cmu.edu) was initiated in 2007 with funding from the National Science Foundation. Each year, it introduces more than 1000 high school students to human languages and language technologies via a pencil-and-paper contest that consists of self-contained puzzles. There is no pre-requisite knowledge, and so far, no registration fee. NACLO is run largely by faculty members and students at various universities who volunteer their time, and administrative staff whose time is donated by Carnegie Mellon's Language Technologies Institute, University of Michigan's Computer Science Department, and Brandeis University's Computer Science Department.

NACLO annual expenses include payments to undergraduates to maintain the web site and registration page, payments to undergraduates to test prospective puzzles, and a trip to the International Linguistics Olympiad (IOL, http://www.ioling.org) for selected high-school students and accompanying adults. In past years, the trip to the IOL has cost between $15,000 and $20,000. When money is available, we have had a training camp for a few days at a cost of $5000 to $10,000. In 2012, we are taking a four-student Canadian team to the IOL along with the two four-student US teams, so the trip cost has increased to $36,000. In addition to the monetary cost of the trip and training camp, the cost in volunteer time spent on fund raising is quite high. We would welcome sponsors who could commit to an annual donation.

Contact:
Lori Levin
Associate Research Professor
Language Technologies Institute
Carnegie Mellon University
lsl@cs.cmu.edu

Web sites:
North American Computational Linguistics Olympiad: http://www.naclo.cs.cmu.edu
International Linguistics Olympiad: http://www.ioling.org

Schedule of NACLO 2012

February 2, 2012	Open Round
March 13, 2012	Invitational Round
April-July 2012	Team training via Skype
July 29-August 4, 2012	International Linguistics Olympiad in Ljubljana, Slovenia

Sponsorship levels

All sponsors will receive a "Sponsor of the North American Computational Linguistics Olympiad" logo that they can display freely on web sites and printed material.

Title Sponsor: $20,000 per year. A sponsor that pays for the entire cost of the trip to IOL or a training camp can be a title sponsor. Title sponsorship includes the sponsor's name next to the contest name and the sponsor's logo isolated from other logos on the contest web page, t-shirts, banners, and program.

Gold Sponsor: $10,000. Gold sponsorship includes a large sponsor logo on the contest web site, t-shirts, banners, and program. Gold sponsors will be thanked in all ceremonies.

Silver Sponsor: $5000. Silver sponsorship includes a smaller sponsor logo on the contest web site, t-shirts, banner, and program. Silver sponsors will be thanked in all ceremonies.

Activity Sponsor: prices vary. The sponsor may fund a meal, social event, sightseeing trip, conference bag, t-shirt, trophy, or other contest memorabilia. The sponsor will be acknowledged in connection with the activity, will be thanked in the contest program, and all ceremonies.

What are linguistics Olympiads?

Linguistics Olympiads are national and international contests for high school students. Unlike math and science Olympiads, which focus on material students have already mastered, linguistics Olympiads introduce new material through logic puzzles in languages that the students are unfamiliar with. In the course of solving a puzzle, the students may discover something about the grammar or sounds of a language, its conceptual system, its culture, or its historical relationship to other languages.

Linguistics Olympiads foster meta-linguistic reasoning that is useful for any career involving human language, as well as analytical problem-solving skills that are relevant for many technical and non-technical careers. Also, because it has been the tradition of linguistics Olympiads to assign half of the score for the written explanation of the answer, the contests also encourage the ability to clearly present a hypothesis and explain how it is supported by the data.

What is the International Linguistics Olympiad (IOL)?

The International Linguistics Olympiad (http://www.ioling.org) is one of twelve international high school Olympiads. It has been held every year since 2003. At least 25 countries have participated, and the number is growing slowly each year. The countries that participated in 2011 are: Australia, Brazil, Bulgaria, Canada, Estonia, India, Ireland, South Korea, Latvia, Netherlands, Poland, Russia, Singapore, Slovenia, Sweden, United Arab Emirates United Kingdom, United States, and Vietnam. More than 30 countries are expected to participate in 2012. The contest consists of a six hour individual contest and a four hour team contest. The contest is administered by pencil and paper in the languages of each participating country. An international jury formulates the problems and scores the contest results.

How much money do we need?

We need around $36,000 for the trip to the International Linguistics Olympiad, and around $10,000 to pay undergraduate assistants to maintain the web site and registration page and to test problems.

Why sponsor a linguistics Olympiad?

Our ability to use language distinguishes us from other living things and offers a window for studying human cognition and social interaction. However, linguistics, the study of all aspects of human language (cognitive, sociological, historical, etc.), is not taught in high schools. As a result, many students make plans to major in other subjects in college and do not find out about linguistics until later in their studies, when it may be too late to change majors.

Linguistics is a fascinating field on its own for those who want to devote their lives to research on human language. However, it is also an ideal subject for practicing skills that can be applied in many different careers. Insights from linguistics can give students meta-linguistic intuitions that can help them be better language learners, language teachers, or translators. Linguistics also uses logical and analytical skills for discovering structure in language, which are important in any field of study. Awareness of linguistics can lead students to be better informed about language policy issues such as bilingual education, heritage language maintenance for immigrants, and revitalization of endangered languages. It also can attract students to computer science by showing that computer scientists work on fascinating problems like machine translation, speech recognition and synthesis, and question answering systems.

Why focus on computational linguistics?

We use human language technologies on a daily basis every time we use a search engine, spam filter, telephone dialogue system, or natural language interface to a phone or car. People produce language constantly, in quantities too vast for human attention. Human language technologies help us use what we produce the most of by searching it, summarizing it, and making it available to us when and how we need it. Computational linguistics is one of the basic sciences that underlie human language technologies.

The United States does not meet its own work force needs in computer science, and NACLO is attempting to combat some of the possible causes of the shortage. NACLO shows students that computer science is "not just numbers and code." It does this by presenting fascinating subject matter related to human languages from a computational perspective. The message is that the real life of a computer scientist is to apply computational science to interesting problems in all fields of study.

NACLO also attempts to bring a more diverse group of students into computer science. According the National Science Foundation, the need for a computationally trained work force is too great to omit major segments of the population such as girls and under-represented socio-economic groups. NACLO tries to reach out to everyone by presenting a contest without pre-requisite knowledge and by making it available without a registration fee.

Who participates in linguistics Olympiads?

Students who enjoy linguistics Olympiads generally have high aptitude in both human languages and mathematics. About half of the participants are female. Because success is based on aptitude rather than amount of prior knowledge, younger teenagers can participate and win. The youngest participants are about twelve years old. The winners are usually fourteen to eighteen years old. Because there is no registration fee and no pre-requisite knowledge, the contest is accessible to every student.

Who wins linguistics Olympiads?

The winners are among the most talented teenagers in the world. Winners from the North American contest have enrolled at Princeton, MIT, Harvard, Stanford, Berkeley, Caltech, the University of Chicago, and the University of Washington. Two winners of the Indian contest are also now studying at MIT. Several of the winners of the North American contest are also math and physics Olympians. Many winners are immigrants or heritage language speaking children of immigrant parents.

Does the contest have an impact?

Although we do not procedures for following up on all NACLO participants, we have anecdotal evidence that many participants take linguistics courses or major in linguistics in college. We have more contact with the NACLO winners, who we work with closely in training for the IOL. Most of the NACLO winners major in math or computer science in college, but also take linguistics courses. Some of the contest winners are majoring in linguistics, and one or two are majoring in other areas of humanities, social sciences, or natural sciences. Some of the 2007 winners are now entering graduate school. Some have already started their research careers by publishing papers and presenting at conferences.

NACLO: The North American Computational Linguistics Olympiad

NACLO (http://www.naclo.cs.cmu.edu) began in 2007 with funding from the US National Science Foundation. We have now conducted six annual contests. Each contest consists of an open round with about 1200 participants and an invitational round with about 120 participants. Two teams of four are chosen from the winners of the invitational contest. The teams are coached by Skype conference for about three months and are then taken to the IOL.

The US is one of the top prize-winning countries in the IOL, winning the team competition four out of five years, and winning individual gold medals four out of five years, as well as many silver and bronze medals, honorable mentions, and best solution prizes for individual problems. The US team wins more prizes per person than any other team, thanks to head coach Dragomir Radev (University of Michigan) and assistant coaches Patrick Littell (Ph.D. student in linguistics at University of British Columbia) and Adam Hesterberg (Ph.D. student in math at MIT and 2007 contest winner).

NACLO emphasizes aptitude for computation more than linguistics Olympiads in other countries. In addition to using logical and analytical skills, it explicitly focuses on concepts and tools from computer science such as finite state machines and graph search, and also introduces applications of computational linguistics such as machine translation and automatic summarization.

Importance of NACLO in the US

The US does not meet its workforce needs in computer science or in language-related skills for intelligence, national security, humanitarian relief, or human services. Enrollment in computer science programs has declined in the past few years. NACLO shows students that computer science is about more than programming and machines. Computer scientists deal with fascinating subject matter on a daily basis. NACLO problems focus on computational thinking – organizing a problem into abstract data structures, algorithms, and search spaces – but do not require students to know math, computer science, or computer programming. In this way, NACLO remains accessible to students who have aptitude but have not yet had training in computer science.

It is too easy for Americans to feel that they do not need to speak other languages, and immigrant families may feel pressure to abandon their heritage languages. However, foreign language skills are critical for business, government, intelligence, national security, education, human services, health care, and humanitarian relief. For students who are inclined to go into language-related careers, NACLO strengthens meta-linguistic skills and fosters deep respect for the beauty and intricacy of each language and what it can tell us about being human.

History of Linguistics Olympiads

The International Linguistics Olympiad (IOL, http://www.ioling.org) was initiated in 2003. The 2012 contest will be the 10th in the series. Before 2003, several countries held national linguistics Olympiad contests. The oldest contest was held in Russia in the 1960's. The Russian contest is credited with the idea of turning linguistic data sets into logic puzzles, thus testing aptitude for pattern matching, and logical and analytical reasoning along with insight about the structure of human languages. Books of linguistics Olympiad problems have been published in Russia and Bulgaria. A book of US linguistics Olympiad problems is in press.

In the US, a Linguistics Challenge Contest was organized by Thomas Payne at the University of Oregon from 1998 to 2001, based on the Russian contest. Smaller contests that pre-date NACLO have been held in Tennessee, San Jose, California, and New York City.