

Report on NACLO (North American Computational Linguistics Olympiad)

Lori Levin, co-chair, lsl@cs.cmu.edu
Thomas Payne, co-chair, tpayne@cs.uoregon.edu
Dragomir Radev, program chair, radev@umich.edu

Introduction

NACLO (formerly known as NAMCLO), <http://www.namclo.org>, is an Olympiad style contest for high school students. The first contest was held on March 29, 2007.

195 high school students participated in Pittsburgh, Boston, Ithaca, and "the internet". The problems and solutions and names of the winners are available on the web site. We are now planning to take the top four winners to the International Linguistics Olympiad (ILO) in St. Petersburg, Russia this summer.

Goals of NACLO

1. Increase the size and diversity of the pool of future scientists in Linguistics, Computational Linguistics, and Human Language Technologies.
2. Identify talented high school students and help them get the background that they need for higher education in Linguistics, Computational Linguistics, and Human Language Technologies
3. Get the scientific study of language into high school curricula (in cooperation with the LSA's Language in the School Curriculum committee)
4. Identify foundational skills that can be taught before college that can prepare students for coursework in linguistics, computer science, and language technologies.

NAACL Sponsorship of NACLO

NAACL contributed \$2000 to the 2007 competition. The money was used for prizes (along with additional sponsorship money from Google and books from Cambridge University Press). Prizes were awarded as follows:

National First Place: \$500
National Second Place: \$350
National Third Place: \$200

Local First Place (four locations):
\$150 and the Cambridge Encyclopedia of Language
Local Second Place (four locations plus one tie):
\$100 and the Cambridge Encyclopedia of Language
Local Third Place (three locations):
\$50 and the Cambridge Encyclopedia of Language

Best solution (eight problems with four ties = 12 recipients):
\$50 each

Total: \$2900 (\$2000 from NAACL, the rest from Google)

There was an error in not including the NAACL logo on the web site, which has now been corrected. The NAACL logo was, however, on the exam booklet and on large posters in Pittsburgh, Boston, and Ithaca.

In addition, NAACL was verbally acknowledged in Pittsburgh in front of 94 students and several teachers.

History of Linguistics Olympiads

Linguistics Olympiads have been held since the 1960's in Russia and since the 1980's in Bulgaria. They are now also held in other countries such as the Netherlands and the UK. A Linguistics Challenge was also held several times in Eugene, Oregon. There is now an International Linguistics Olympiad, whose 5th annual contest is July 31 to August 4 in St. Petersburg, Russia (www.ilolympiad.spb.ru).

This year's ILO is preceded by a one week summer school in Narva-Joesuu, Estonia.

History of NACLO

About one year ago, Tanya Korelsky (US National Science Foundation) suggested that North America should participate in the ILO and also have its own Olympiad contest. Lori Levin agreed to apply to NSF for a grant in order to hold a planning workshop. The workshop was held in September, 2006 at the Interspeech conference in Pittsburgh, PA. About forty people attended the workshop including representatives from NAACL and ACL, high school teachers, representatives from the ILO and Moscow Linguistics Olympiad, organizers of other contests, and faculty and graduate students from about 5 other universities.

At the workshop, it was determined that Spring would be the best time for the competition. It was decided to ambitiously try for a pilot competition in Spring 2007, with only six months to plan and organize it.

Officers were elected at the workshop:

Co-chairs: Lori Levin and Thomas Payne
Program chair: Dragomir Radev
Follow-up chair: Barbara Di Eugenio
Outreach chair: William Lewis
Sponsorship chair: James Pustejovsky

Other jobs were created later:

School liaison: Amy Troyani
School practice program: Rebecca Hwa and Noah Smith
Administrative assistant: Mary Jo Bensasi

Locations of NACLO

NACLO was held in Pittsburgh, Boston, and Ithaca. In addition, students could participate in remote locations provided that they were monitored by a teacher. Philadelphia planned to participate, but

couldn't due to the tight time line.

Program

The program chair (Dragomir Radev) is responsible for soliciting and reviewing problems, choosing problems for the competition, and supervising the grading of problems. A call for problems was issued in October 2006. Problems were submitted and reviewed in November and December 2006. The submitted problems were split into practice problems, which were posted on the web, and potential competition problems, which were held secretly in reserve. After NACLO was publicized on the ACL mailing list, additional problems were submitted. Eight problems were chosen for the competition, some from US faculty and some from foreign faculty with experience in the ILO. The length of the exam was five hours.

Types of problems: We attempted to augment the ILO-style linguistic puzzles with some problems that focus more on computational concepts and applications such as web search and OCR. Pedagogical goals of the problems and effective problem types are ongoing areas of research.

Outreach

The outreach chair, William Lewis, was responsible for the web site, practice session curriculum that would be presented in schools, and other kinds of publicity. Linguist List joined to help with the web site, with additional funding from NSF. A school practice session was designed by Rebecca Hwa and Noah Smith, including information about linguistics and computer science and practice problem solving. Several practice sessions were held in Pittsburgh, Boston, and Ithaca. There were also some press releases and some postings in newspaper calendars.

Sponsorship

The sponsorship chair was James Pustejovsky. Since this was our first pilot year, we did not contact very many potential sponsors. Google contributed \$6000, NAACL contributed \$2000, and Cambridge University Press contributed 12 copies of the Encyclopedia of Language.

Follow up

The follow up chair (Barbara Di Eugenio) is responsible for getting feedback on the competition and for setting up clubs and mentoring programs so that students know how to find appropriate universities and courses in order to continue studying linguistics and language technologies. The main follow up activities for this year were an evaluation form and the trip for four students to the ILO summer school and competition.

The evaluation form includes a variety of questions including how did you hear about NACLO, would you participate again, how can we make it better, etc. Students were also asked to rate each problem. Results have been tabulated and are available on request.

Diversity

About half of the top 50 scorers are girls. The national first place winner is a girl.

One of the goals of NACLO is to introduce linguistics and language technologies to populations that are under-represented in the field. However, we did not make any special effort in this direction for the 2007 contest. For example, in Pittsburgh, the contest was publicized through the network of program coordinators for gifted students at high schools. There is some economic diversity in gifted programs, and a bit of racial diversity, but not an overwhelming amount.

The future of NACLO and future financial needs

The 2007 NACLO recruited 195 participants with less than eight weeks of minimal publicity. We plan to start school outreach sessions in the Fall of 2007 for the 2008 contest. In addition, there will be more cities participating. We are estimating that the 2008 NACLO will have over 1000 participants. There will probably be an internet-based qualifying round.

Our activities for the next year will focus on: establishing a non-profit organization, writing a constitution, refining the lesson plan for high school practice sessions, and refining the types of exam problems to meet pedagogical goals.

The expenses for NACLO include:

1. Operating expenses such as copying, mailing, and room rentals and lunch for the participants.
2. Prizes: We expect the 2008 prizes to be similar to the 2007 prizes.
3. Trip to the ILO: This year's trip to the ILO for four students and two chaperones will cost at least \$17,000. The money is coming from Google and NSF's office of international programs. The costs includes a day of training in New York, 250 euros per student for the summer school in Estonia, plane tickets, and some hotel and food costs for days that are not covered by the ILO and summer school. The ILO pays for food and lodging during the competition, but since the ILO is growing, the 2008 ILO may not be able to provide food and lodging for all participants. For example, they may be able to cover the costs for one team per country, but several countries send more than one team. NACLO certainly plans to send more than one team in 2008.
4. Personnel: This year there was minimal coverage of faculty time and administrative assistant time through the NSF planning grant. It would be good to work out a way to get more coverage for faculty and administrative salaries. For comparison: chairing NACLO is takes at least as much time as being general chair of a conference; sponsorship and administration also take about the same amount of time as for a conference; program committee takes at least as much time and could take more. For example, the Moscow program committee meets weekly year round and may refine each problem for several months; local arrangements takes more time than for a conference because of outreach to highschools and practice sessions. This is ok for one year at a time, but a few devoted people will probably be involved on an ongoing

basis and could use some salary coverage.