

Say What? High School Linguists Break the Code

Winners of the fourth annual North American Computational Linguistics Olympiad just announced.

More than a thousand high school students from across the USA and Canada recently competed in the fourth annual North American Computational Linguistics Olympiad. The top students are eligible to represent their country at the Eighth International Linguistics Olympiad to be held in Sweden in late July.

The competition included two rounds -- the Open round on February 4th and the Invitational round on March 10th. 1118 students participated in the Open competition at more than 100 sites, including universities such as Carnegie Mellon, Princeton, Stanford, and University of Michigan, as well as many high schools. The students with the top 100 scores in the open round advanced to the Invitational round, which featured significantly harder questions.

Top winners include:

1. Ben Sklaroff, Palo Alto, CA, Palo Alto High School
2. Brian Kong, Milton, MA, Milton Academy
3. Allen Yuan, Farmington Hills, MI, Detroit Country Day School
4. Daniel Li, Fairfax, VA, Thomas Jefferson High School for Science and Technology
5. Alan Chang, San Jose, CA, Piedmont Hills High School
6. Alexander Iriza, Astoria, NY, The Dalton School
7. In-Sung Na, Old Tappan, NJ, Northern Valley Regional High School at Old Tappan
8. Tian-Yi Damien Jiang, Raleigh, NC, North Carolina School of Science & Mathematics (Durham)

Students compete in the Computational Linguistics Olympiad by solving challenging problems using data from a variety of languages and formal systems. There is no pre-requisite knowledge. Students discover facts about languages and formal systems in the course of solving the puzzles. According to first place winner Ben Sklaroff, "Translating a language you've never heard of before just by using logic is extremely gratifying. It's like breaking a code, except languages generally make sense so all the rules are less arbitrary. Trying to figure out how another person would express common concepts in their own language, with just a few examples to work with is a fun challenge."

This year students solved sixteen problems, including deciphering the rules for a Pig-Latin-like play language in Minangkabau, the writing systems of Plains Cree, and the Vietnamese classic Tale of Kieu written in Chinese characters. Computational problems dealt with text compression and automatic expansion of abbreviated words. Alan Chang (5th place) says, "I really didn't know what I was getting myself into until the open round actually started. Those three hours didn't feel like a test at all. They were three hours of very creative and challenging puzzles. As much as I like math and physics competitions, I found this linguistics competition to be the most fun. I was very surprised at how many problems I could solve without any prior experience or knowledge of the subject, and this made me feel more accomplished every time I solved a problem."

Dragomir Radev of the University of Michigan is the chair of the program committee. Among his many responsibilities, Radev gathers ideas from industry and academic researchers around the world. Radev aims to create challenging and stimulating problems that address cutting edge issues in the field of computational linguistics. Though not yet widely known to the general public, computational linguistics is a rapidly emerging field with applications in such areas as search engine technologies, machine translation, and artificial intelligence.

While the linguistics competition is fun, it also requires dedication and hard work by many people, all of whom are volunteers. Dragomir Radev and Lori Levin (Carnegie Mellon University) co-chair the organizing committee, which also includes School Liaison Amy Troyani (Pittsburgh Allderdice High

School), Administrative Chair Mary Jo Bensasi (Carnegie Mellon University) and Sponsorship Chair James Pustejovsky (Brandeis University), as well as problem authors and jury members Eugene Fink (Carnegie Mellon University), David Mortensen (University of Pittsburgh), Patrick Littell (University of British Columbia), and 2007 international gold medalist Adam Hesterberg, now studying at Princeton University. Many other college professors, high school teachers, and college students also volunteer their time.

NACLO is sponsored by the National Science Foundation and the North American Chapter of the Association for Computational Linguistics (NAACL), Carnegie Mellon University Language Technologies Institute and Gelfand Center for Community Outreach, University of Michigan, and Brandeis University, as well as donations from academic departments and individual donors.

Programs similar to NACLO have taken place for over forty years in Eastern Europe, and the International Linguistics Olympiad is in its eighth year. More information as well as the problem sets and solutions can be found on the NACLO website www.naclo.cs.cmu.edu.

"Usually, college students don't even hear about computational linguistics until they are well along in their undergraduate studies," says Lori Levin of Carnegie Mellon University, co-chair of the North American program. "Our hope is that competitions such as the Computational Linguistics Olympiad will identify students who have an affinity for linguistics and computational linguistics before they graduate high school and encourage them to pursue further studies at the university level." The organization also hopes to see the scientific study of language incorporated into high school curricula. Charles Forster, a computer science teacher, has created a new computational linguistics course at the Dalton School in NY. "We are making an effort to cater to the students who are in the department but are less interested in following our singular "algorithms" track, as well as to students who are afraid to take computer science because it has the word "science" in it but are interested in taking language. Ling is a great crossover field in HS. For kids who are less inclined to science, they learn to see a subject that they are interested in through a science lens. Inversely, it is a great entre for sci geeks to english and language."

Universities and corporations view the program as a way of helping high school students discover their talents and interests in the areas of language, linguistics and natural language processing. "High school students are always enthusiastic about logic puzzles, and the Linguistics Olympiad provides lots of them," says Adam Hesterberg, vice-chair of the jury and winner of the 2007 International Linguistics Olympiad. "It's like a math contest without the requirement of knowing any math, although without the rigor of a math contest. Indeed, mathematicians normally do quite well in the contests." Chang adds, "Despite all being based on linguistics, the problems in NACLO are very diverse. Every time I began a new problem, I had to think carefully about what I could use to solve it. The techniques I ended up using ranged from applying basic English grammar to searching for patterns to solving systems of equations."

Dragomir Radev certainly feels that his hard work pays off. "Many of the participants are extremely bright and have broad interests. In addition to linguistics, they also excel in physics, mathematics, computing, and many other subjects. A number of linguistics clubs have been created at high schools thanks to NACLO."

And, as Eugene Fink puts it, "most importantly, it is fun for all participants, both students and organizers." Allen Yuan (third place) concurs, "NACLO has been one of the most enlightening experiences in my life, combining my love for solving puzzles with a newly sparked interest in languages. The contest was very well organized this year and I hope that this event can continue to expand in the future. It brings a great opportunity to change the way all the students think."

Contact information:

Dr. Lori Levin, Carnegie Mellon University (ls1@cs.cmu.edu)

Dr. Dragomir Radev, University of Michigan (radev@umich.edu)