Who's the best high school linguist of them all?

Winners from the third annual North American Computational Linguistics Olympiad just announced.

More than 1,000 high school students from across the USA and Canada recently competed in the third annual North American Computational Linguistics Olympiad. The top students are eligible to represent their country at the Seventh International Linguistics Olympiad to be held in Poland in late July.

The competition took place in two rounds -- an Open round on February 4th and an Invitational round on March 11th. 1080 students participated in the Open competition at more than 100 sites, including universities as well-known as Carnegie Mellon, Princeton, Stanford, and NYU, but also many others such as the University of Texas at El Paso and Montana State, as well as many high schools. The top 135 scoring students on the open round participated in the Invitational round which featured significantly harder questions.

Top winners include:
1st- Anand Natarajan from The Harker School in San Jose, California
2nd- Alan Huang from Detroit Country Day School in Beverly Hills, Michigan
3rd- Vivaek Shivakumar from Thomas Jefferson High School for Science and Technology in Alexandria, Virginia
4th- Daryl Hansen from Skyline High School in Sammamish, Washington
5th- Sergei Bernstein from Belmont High School in Belmont, Massachusetts
6th- John Berman from John T. Hoggard High School in Wilmington, North Carolina
7th- Morris Alper from Henry M. Gunn High School in Palo Alto, California
8th- Rebecca Jacobs from Harvard-Westlake School in North Hollywood, California

Students compete in the Computational Linguistics Olympiad by solving challenging problems using data from a variety of languages and formal systems the students have never learned. This year students solved a total of 13 problems. In most of these, students were given some data from a language they weren't expected to know anything about and asked to figure out something of the structure of the language. Some of the problems also dealt with how computational thinking may be applied to some thorny language processing problems, such as how a computer can interpret human language.

This year, Dr. Dragomir Radev, of the University of Michigan, chaired the program committee. Among his many responsibilities, Dr. Radev gathers ideas from industry and academic researchers around the world. Dr. 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.

The US program is sponsored by the National Science Foundation, Google, Cambridge University Press, Microsoft, Everyzing, M*Modal, JUST. Systems, The North American Chapter of the Association for Computational Linguistics (NAACL), Oxford University Press, Carnegie Mellon Universitys Language Technologies Institute, the University of Michigan, Brandeis University, and the University of Pittsburgh Linguistics Department. Similar programs have taken place for over forty years in Eastern Europe, and the International competition is in its seventh year. More information as well as the problem sets and solutions can be found on the organization's 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 Dr. 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. 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."

While the linguistics competition is fun, it also requires dedication and hard work by many people, all of whom are volunteers. The organizing committee is headed by Professor Dragomir Radev (University of Michigan) and Professor Lori Levin (Carnegie Mellon University), and it also includes Mary Jo Bensasi, Eugene Fink, Adam Hesterberg, Patrick Littell, Ida Mayer, James Pustejovsky, and Amy Troyani. The program committee includes twenty more people, who create new competition problems and judge the performance of contestants. The other volunteers are high-school teachers and college students who help to organize and proctor the event.

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."

Contact information:

Dr. Lori Levin, Carnegie Mellon University (lsl@cs.cmu.edu)
Dr. Dragomir Radev, University of Michigan (radev@umich.edu)