

The North American Computational Linguistics Olympiad (NACLO)

Dragomir R. Radev
SI, EECS, and Linguistics
University of Michigan
radev@umich.edu

Lori S. Levin
Language Technologies Institute
Carnegie-Mellon University
lsl@cs.cmu.edu

Thomas E. Payne
Department of Linguistics
University of Oregon
tpayne@uoregon.edu

Abstract

In this paper, we report on the 2007 and 2008 North American Computational Linguistics Olympiad (NACLO) competitions. NACLO is a competition whose goal is to expose high school students to the field of Linguistics, including Computational Linguistics and to serve as a talent search for future students and researchers in these areas. The 2007 edition attracted 195 high school participants. Two teams of four students represented the United States at the International Linguistics Olympiad in Russia where they tied for first place at the team event and one student obtained the highest score at the individual event. The 2008 event expanded to Canada and included 763 participants.

1 Introduction

NACLO (North American Computational Linguistics Olympiad) is an annual Olympiad-style contest for high school students, focusing on linguistics, computational linguistics, and language technologies.

The goal of NACLO is to increase participation in these fields by introducing them before students reach college. Since these subjects are not normally taught in high school, we do not expect students to have any background of these areas before the contest. The contest consists of self-contained problems that can be solved with analytical thinking, but in the course of solving each problem, the

students learn something about a language, culture, linguistic phenomenon, or computational tool.

The winners of NACLO are eligible to participate in the International Linguistics Olympiad as part of the US team.

1.1 History of the LO and ILO

The International Olympiad in Linguistics is one of twelve international Science Olympiads (the others include Mathematics, Physics, Chemistry, Biology, Informatics, Philosophy, Astronomy, Geography, and Earth Science). It has existed since 2003 and has, so far, been held exclusively in Europe (Russia, Estonia, Bulgaria, and the Netherlands). ILO 2007 took place in Zelenogorsk near St. Petersburg, Russia whereas ILO 2008 will be in Slantchev Bryag near Burgas, Bulgaria. ILO 2009 will be held in Poland.

Individual national linguistics Olympiads have been held in Russia since 1965 (based on an initiative by Andrey Zaliznyak) and in other countries more recently¹. Recently, a collection of problems from different decades appeared in Russian (Belikov et al., 2007).

1.2 Linguistics Contests in the US

Thomas Payne pioneered LO-style competitions in the USA by organizing three consecutive contests for middle and high school students in the Eugene, Oregon area in 1998-2000. In the course of publicizing NACLO, we have discovered that other local linguistics contests have taken place in Tennessee, San Jose, and New York City.

¹ The first author of this paper participated in the Bulgarian national LO in the early 1980s.

1.3 Origin of NACLO

NACLO began with a planning workshop funded by NSF in September 2006. The attendees included faculty and graduate students from about ten universities as well as representatives from NSF and ACL. Two high school teachers were present. The workshop opened with presentations from organizers of other Olympiads and contests in linguistics and computer programming. In particular we received excellent advice from Ivan Derzhanski, representing the International Linguistics Olympiad, and Boris Iomdin, representing the Moscow Olympiad. The remainder of the workshop dealt with scheduling the first contest, electing committee chairs, and making organizational decisions.

1.4 Pedagogical goals

We have two goals in organizing NACLO. We want to increase broad participation and diversity in all language-related careers. We want every student to have a fun and educational experience and have a positive attitude toward taking linguistics and language technologies courses in college. However, we also want to conduct a talent search for the most promising future researchers in our field. NACLO uses two mechanisms to be sure that we reach all levels of participation. The first mechanism is to separate an open round with easier problems from an invitation-only round with harder problems. The second mechanism is related to grading the problems. Forty percent of the score is for a correct answer and sixty percent is for explaining the answer. The students who write the most insightful explanations are the focus of our talent search.

When publicizing NACLO in high schools we have been focusing on certain aspects of linguistics and computer science. With respect to linguistics, we emphasize that languages have rules and patterns that native speakers are not aware of; that there are procedures by which these rules and patterns can be discovered in your own language; and that the same procedures can be used to discover rules and patterns in languages other than your own. With respect to computer science the term *computational thinking* has been coined (Wing 2006) to refer to those parts of the field that are not

about computers or programming: thinking algorithmically, using abstraction to model a problem, structuring and reducing a search space, etc.

1.5 Organization at the national level

NACLO has two co-chairs, currently Lori Levin, Carnegie Mellon University, and Thomas Payne, University of Oregon. Dragomir Radev is the program chair and team coach. Amy Troyani, a high school teacher with board certification, is the high school liaison and advisor on making the contest appropriate and beneficial to high school students.

NACLO has several committees. James Pustejovsky currently chairs the sponsorship committee. The other committees are currently unchaired, although we would like to thank William Lewis (outreach and publicity) and Barbara Di Eugenio (followup) for chairing them in the first year. NACLO is not yet registered as a non-profit organization and does not yet have a constitution. We would welcome assistance in these areas.

The national level organization provides materials that are used at many local sites. The materials include a comprehensive web site (<http://www.naclo.cs.cmu.edu>), practice problems, examples of flyers and press releases, PowerPoint presentations for use in high schools, as well as contest booklets from previous competitions.

The contest is held on the same day in all locations (universities and "online" sites as described below). In 2007 there was a single round with 195 participants. In 2008 there was an open round with 763 participants and an invitation-only round with 115 participants. Grading is done centrally. Each problem is graded at one location to ensure consistency.

Three national prizes are awarded for first, second, and third place. National prizes are also given for the best solution to each problem. Local hosts can also award prizes for first, second, and third place at their sites based on the national scores.

1.6 Funding

The main national expenses are prizes, planning meetings, and the trip to the International Linguistics Olympiad (ILO). The trip to the ILO is the largest expense, including airfare for eight team members (two teams of four), a coach, and two

chaperones. The national level sponsors are the National Science Foundation (2007, 2008), Google (2007, 2008), Cambridge University Press (2007, 2008), and the North American Chapter of the Association for Computational Linguistics (2007). The organizers constantly seek additional sponsors.

1.7 Publicity before the contest

At the national level, NACLO is publicized through its own web site as well as on LinguistList and Language Log. From there, word spreads through personal email and news groups. No press releases have been picked up by national papers that we know of. Local level publicity depends on the organization of local schools and the hosting university's high school outreach programs. In Pittsburgh, publicity is facilitated by a central mailing list for gifted program coordinators in the city and county. Some of the other local organizers (including James Pustejovsky at Brandeis, Alina Johnson at the University of Michigan and Barry Schiffman at Columbia University as well as several others) sent mail to literally hundreds of high schools in their areas. Word of mouth from the 2007 contest also helped reach out to more places.

1.8 Registration

NSF REU-funded Justin Brown at CMU created an online registration site for the 2008 contest which proved very helpful. Without such a site, the overhead of dealing with close to 1,000 students, teachers, and other organizers would have been impossible.

1.9 Participation of graduate and undergraduate students

Graduate and undergraduate students participate in many activities including: web site design, visiting high schools, formulating problems, testing problems, advising on policy decisions, and facilitating local competitions.

2 Problem selection

We made a difficult decision early on not to require knowledge of linguistics, programming or mathematics. Requiring these subjects would have reduced diversity in our pool of contestants as well as its overall size. Enrollment in high school pro-

gramming classes has dropped, perhaps because of a perception that programming jobs are not interesting. NACLO does not require students to know programming, but by introducing a career option, it gives them a reason to take programming classes later.

2.1 Problem types

The NACLO problem sets include two main categories of problems: "traditional" and "computational/formal". The ILO includes mostly traditional problems which include translations from unknown languages, glyph decoding, calendar systems, kinship systems, mathematical expressions and counting systems, among others. The other category deals with linguistic phenomena (often in English) as well as algorithms and formal analyses of text.

2.2 Problem committee

A problem committee was formed each year to work on the creation, pre-testing, and grading of problems. The members in 2007 included Emily Bender, John Blatz, Ivan Derzhanski, Jason Eisner, Eugene Fink, Boris Iomdin, Mahesh Joshi, Anagha Kulkarni, Will Lewis, Patrick Littell, Ruslan Mitkov, Thomas Payne, James Pustejovsky, Roy Tromble, and Dragomir Radev (chair). In 2008, the following people were members: Emily Bender, Eric Breck, Lauren Collister, Eugene Fink, Adam Hesterberg, Joshua Katz, Stacy Kurnikova, Lori Levin, Will Lewis, Patrick Littell, David Mortensen, Barbara Partee, Thomas Payne, James Pustejovsky, Richard Sproat, Todor Tchervenkov, and Dragomir Radev (chair).

2.3 Problem pool

At all times, the problem committee maintains a pool of problems which are constantly being evaluated and improved. Professional linguists and language technologists contribute problems or problem ideas that reflect cutting-edges issues in their disciplines. These are edited and tested for age appropriateness, and the data are thoroughly checked with independent experts.

2.4 Booklets

The three booklets (one for 2007 and two for 2008) were prepared using MS Publisher. Additionally, booklets with solutions were prepared in MS Word. All of these are available from the NACLO web site.

2.5 List of problems

This is the list of problems for NACLO 2007 (8 problems) and 2008 (12 problems). They can be divided into two categories: traditional (2007: C, D, G and 2008: A, C, D, E, G, J, K) and formal/computational (2007: A, B, E, F, H and 2008: B, F, H, I, L). The traditional problems addressed topics such as phonology, writing systems, calendar systems, and cognates, among others. The other category included problems on stemming, finite state automata, clustering, sentence similarity identification, and spectrograms.

2007
A. English (Molistic)
B. English (Encyclopedia)
C. Ancient Greek
D. Hmong
E. English (Verb forms)
F. English (Spelling correction)
G. Huishu (Phonology)
H. English (Sentence processing)
2008 (A-E Open; F-L Invitational)
A. Apinaye (Brazil)
B. Hindi
C. Ilocano (Philippines)
D. Swedish and Norwegian
E. Aymara (South America)
F. Japanese
G. Manam Pile (Papua New Guinea)
H. English (Stemming)
I. Rotokas (Automata; Bougainville Island)
J. Irish
K. Mayan (Calendar)
L. English (Spectrograms)

Figure 1: List of languages used in NACLO 2007 and 2008.

3 Contest administration

NACLO is run in a highly distributed fashion and involves a large number of sites across the USA and Canada.

3.1 Local administration

NACLO is held at hosting universities and also "online". The online category includes students who cannot get to one of the hosting universities, but instead are monitored by a teacher at a convenient location, usually the student's high school. There were three hosting universities (Carnegie-Mellon, Brandeis, and Cornell) in 2007 and thirteen hosting universities (the three above + U. Michigan, U. Illinois, U. Oregon, Columbia, Middle Tennessee State, San Jose State, U. Wisconsin, U. Pennsylvania, U. Ottawa, and U. Toronto) in 2008. Any university in the US or Canada may host NACLO. Local organizers are responsible for providing a room for the contest, contacting high local high schools, and facilitating the contest on the specified contest date. Local organizers may decide on the number of participants. The number of participants at the 2008 sites ranged from a handful to almost 200 (CMU-Pitt).

Local organizers may choose their level of investment of time and money. They may spend only a few hours recruiting participants from one or two local high schools and may spend a small amount of money on school visits and copying. But they may also run large scale operations including extensive fundraising and publicity. The site with the largest local participation, Carnegie Mellon/University of Pittsburgh, donated administrative staff time, invested hundreds of volunteer hours, and raised money for snacks and souvenirs from local sponsors². The CMU-Pitt site also hosts a problem club for faculty and students where problems are proposed, fleshed out, and tested. At the University of Oregon, a seminar course was taught on Language Task Creation (formulation of problems) for which university students received academic credit.

3.2 Remote ("online") sites

We had about 65 such sites in 2008. All local teachers and other facilitators did an amazing job following the instructions for administering the competition and for promptly returning the submissions by email or regular mail.

² We are grateful to the Pittsburgh sponsors: M*Modal, Vivísimo, JustSystems Evans Research, and Carnegie Mellon's Leonard Gelfand Center for Service Learning and Outreach.

3.3 Clarifications

During each of the three competitions, the jury was online (in some cases for 8 hours in a row) to provide live clarifications. Each local facilitator was asked to be online during the contest and relay to the jury any questions from the students. The jury then, typically within 10 minutes, either replied “no clarification needed” (the most frequent reply) or provided an answer which was then posted online for all facilitators to see. We received dozens of clarifications requests at each of the rounds.

3.4 Grading

Grading was done by the PC with assistance from local colleagues. To ensure grade consistency, each problem was assigned to a single grader or team of graders. Graders were asked to provide grading rubrics which assigned individual points for both “practice” (that is, getting the right answers) and “theory” (justifying the answers).

3.5 Results from 2007

195 students participated in 2007. The winners are shown here. One of the students was a high school sophomore (15 years old) while three were seniors at the time of the 2007 NACLO.

1. Rachel Zax, Ithaca, NY
2. Ryan Musa, Ithaca, NY
3. Adam Hesterberg, Seattle, WA
4. Jeffrey Lim, Arlington, MA
5. (tie) Rebecca Jacobs, Encino, CA
5. (tie) Michael Gottlieb, Tarrytown, NY
7. (tie) Mitha Nandagopalan, San Jose, CA
7. (tie) Josh Falk, Pittsburgh, PA
- Alternate. Anna Tchetchetkine, San Jose, CA

Figure 2: List of team members from 2007. Mitha was unable to travel and was replaced by Anna Tchetchetkine.

3.6 2008 Winners

The 2008 contest included 763 participants in the Open Round and 115 participants in the Invitational Round. The winners of the Invitational Round are listed below. These are the eight students who are eligible to represent the USA at the 2008 ILO. As of the writing of this paper, all eight

were available for the trip. One of the eight is a high school freshman (9th grade).

1. Guy Tabachnick, New York, NY
2. Jeffrey Lim, Arlington, MA
3. Josh Falk, Pittsburgh, PA
4. Anand Natarajan, San Jose, CA
5. Jae-Kyu Lee, Andover, MA
6. Rebecca Jacobs, Encino, CA
7. Hanzhi Zhu, Shrewsbury, MA
8. Morris Alper, San Jose, CA

Figure 3: List of team members from 2008.

3.7 Canadian Participation

Canada participated for the first time in 2008 (about 20 students from Toronto, a handful from Ottawa and one from Vancouver). Two students did really well at the 2008 Open (one ranked second and two tied for 13th) but were not in the top 20 at the Invitational.

3.8 Diversity

About half of the participants in NACLO were girls in 2007 and 2008. In 2007, 25 out of the top 50 students were female.

The two US teams that went to the ILO in 2007 included three girls, out of eight total team members (two teams of four). The 2008 teams include only one girl.

3.9 Other statistics

Some random statistics: (a) of the top 20 students in 2008, 14 are from public schools, (b) 26 states, 3 Canadian provinces, and the District of Columbia were represented in 2008.

4 Preparation for the ILO

Preparation for the ILO was a long and painful process. We had to obtain visas for Russia, fund and arrange for the trip, and do a lot of practices.

4.1 Teams

One of the students who was eligible to be on the second USA team was unable to travel. We went down the list of alternates and picked a different student to replace her.

4.2 Funding

The ILO covered room and board for the first team and the team coach. The second team was largely self-funded (including airfare and room and board). Everyone else was funded as part of the overall NACLO budget. The University of Michigan covered the coach's airfare.

4.3 Training

We ran multiple training sessions. The activities included individual problem solving, team problem solving (using Skype's chat facility), readings, as well as live lectures (both at the summer school in Estonia and on the day before the main ILO in Russia).

4.4 Travel logistics

Four students, two chaperones, and one parent left early to attend a summer school organized by the Russian team in Narva, Estonia. The third chaperone and three students traveled directly to the ILO. The eighth student traveled with her parents and did some sightseeing in Russia prior to the ILO.

5 Participation in the ILO

The ILO was organized by a local committee from St. Petersburg chaired by Stanislav Gurevych. The organization was extraordinary. Everything (problem selection, grading, hotel, activities, food) was excellent.

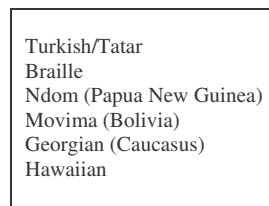
5.1 Organization of the ILO

The ILO was held at a decent hotel in Zelenogorsk, a suburb of St. Petersburg on the Baltic Sea. The first day included an orientation, the second day was the individual contest and team building activities, the third day – an excursion to St. Petersburg, the fourth day – the team contest and awards ceremony.

5.2 Problems

The problems given at the ILO were quite diverse and difficult. The hardest problems were the one in the Ndom language which involved a non-standard number system and the Hawaiian problem

given at the team contests which involved a very sophisticated kinship system.



Turkish/Tatar
Braille
Ndom (Papua New Guinea)
Movima (Bolivia)
Georgian (Caucasus)
Hawaiian

Figure 4: List of languages used in ILO 2007.

5.3 Results

Adam Hesterberg scored the highest score in the individual contest. One of the two US teams (Rebecca Jacobs, Joshua Falk, Michael Gottlieb, and Anna Tchetchetkine) tied for first place in the team event.

6 Future directions

The unexpected interest in the NACLO poses a number of challenges for the organizers. Further challenges arise from our desire to cover more computational problems.

6.1 Grading and decentralization?

Grading close to 5,000 submissions from 763 students in 2008 took a toll on our problem committee. The process took more than two weeks. We are considering different options for future years, e.g., reducing the number of problems in the first round or involving some sort of self-selection (e.g., asking each potential participant to do a practice test and obtain a minimal score on it). These options are suboptimal as they detract from some of the stated goals of the NACLO and we will not consider them seriously unless all other options (e.g., recruiting more graders). have been exhausted.

6.2 Problem diversity

We would like to include more problem types, especially on the computational end of the contest. This is somewhat of a conflict with the ILO which includes mostly "traditional" LO problems. One possibility is to have the first round be more computational whereas the invitational round would be more aimed at picking the team members for the ILO by focusing more on traditional problems.

6.3 Practice problems

We will be looking to recruit a larger pool of problem writers who can contribute problems of various levels of difficulty (including very easy problems and problems based on the state of the art in research in NLP). We are also looking for volunteers to translate problems from Russian, including the recently published collection “Zadachi Lingvisticheskoy Olimpiad”.

6.4 Other challenges

The biggest challenges for the NACLO in both years were funding and time management.

In 2007, four of the students had to pay for their own airfare and room and board. At the time of writing, the budget for 2008 is still not fully covered. The current approach with regard to sponsorship is not sustainable since NSF cannot fund recurring events and the companies that we approached either gave nothing or gave a relatively small amount compared to the overall annual budget.

The main organizers of the NACLO each spent several hundred hours (one of them claims “the equivalent to 20 ACL program committee chairmanships”), mostly above and beyond their regular appointments. For NACLO to scale up and be successful in the future, a much wider pool of organizers will be needed.

6.5 Other countries

Dominique Estival told us recently that an LO will take place in Australia in Winter 2008 (that is, Summer 2008 in the Northern Hemisphere). OzLO (as it is called) will be collaborating with NACLO on problem sets. Other countries such as the United Kingdom and the Republic of Ireland are considering contests as well. One advantage that these countries all have is that they can share (English-language) problem sets with NACLO.

6.6 Participant self-selection

Some Olympiads provide self-selection problems. Students who score poorly on these problem sets are effectively discouraged from participation in the official contest. If the number of participants keeps growing, we may need to consider this option for NACLO.

6.7 More volunteers

NACLO exerted a tremendous toll on the organizers. Thousands of hours of volunteer work went into the event each year. NACLO desperately needs more volunteers to help at all levels (problem writing, local organization, web site maintenance, outreach, grading, etc).

7 Overall assessment

While it will take a long time to properly assess the impact of NACLO 2007 and 2008, we have some preliminary observations to share.

7.1 Openness

We made a very clear effort to reach out to all high school students in the USA and Canada. Holding the contest online helped make it truly within everyone’s reach. Students and teachers overwhelmingly appreciated the opportunity to participate at no cost (other than postage to send the submissions back to the jury) and at their own schools. Students who participated at the university sites similarly expressed great satisfaction at the opportunity to meet with peers who share their interests.

7.2 Diversity and outreach

We were pleased to see that the number of male and female participants was nearly equal. A number of high schools indicated that clubs in Linguistics were being created or were in the works.

7.3 Success at the ILO

Even though the US participated for the first time at the ILO, the performance shown there (including first place individually and a tie for first place in the team contest) was outstanding.

Acknowledgments

We want to thank everyone who helped turn NACLO into a successful event. Specifically, Amy Troyani from Taylor Allderdice High School in Pittsburgh, Mary Jo Bensasi of CMU, all problem writers and graders (which include the PC listed above as well as Rahel Ringger and Julia Workman) and all local contest organizers (James Pustejovsky, Lillian Lee, Claire Cardie, Mitch Marcus,

Kathy McKeown, Barry Schiffman, Lori Levin, Catherine Arnott Smith, Richard Sproat, Roxana Girju, Steve Abney, Sally Thomason, Aleka Blackwell, Roula Svorou, Thomas Payne, Stan Szpakowicz, Diana Inkpen, Elaine Gold). James Pustejovsky was also the sponsorship chair, with help from Paula Chesley. Ankit Srivastava, Ronnie Sim and Willie Costello co-wrote some of the problems with members of the PC. Eugene Fink helped with the solutions booklets, Justin Brown worked on the web site, and Adam Hesterberg was an invaluable member of the team throughout. Other people who deserve our gratitude include Cheryl Hickey, Alina Johnson, Patti Kardia, Josh Cannon, Christina Hunt, Jennifer Wofford, and Cindy Robinson. Finally, NACLO couldn't have happened without the leadership and funding provided by NSF and Tanya Korelsky in particular as well as the generous sponsorship from Google, Cambridge University Press, and the North American Chapter of the ACL (NAACL).

The authors of this paper are also thankful to Martha Palmer for giving us feedback on an earlier draft.

NACLO was partially funded by the National Science Foundation under grant IIS 0633871 Planning Workshop for a Computational Linguistics Olympiad.

References

- Vasileios Hatzivassiloglou and Kathleen McKeown. 1997. Predicting the Semantic Orientation of Adjectives, ACL 1997.
- Jeannette Wing, Computational Thinking, CACM vol. 49, no. 3, March 2006, pp. 33-35.
- V. I. Belikov, E. V. Muravenko and M. E. Alexeev, editors. *Zadachi Lingvisticheskikh Olimpiad. MTsNMO. Moscow, 2007.*

Appendix A. Summary of freeform comments

"I think it's a great outreach tool to high schools. I was especially impressed by the teachers who came and talked to [the linguistics professors] about starting a linguistics club"

"The problems are great. One of our undergraduates expressed interest in a linguistics puzzle contest (on the model of Google's and MS's puzzle contests) at the undergrad level."

"We got a small but very high-quality group of students. To get a larger group, we'd need to start earlier."

"Things could be more streamlined. I think actually *less* communication, but at key points in the process, would be more effective."

"It also would have been nice if there were a camp, like with the other US olympiads, so that more students would get the chance to learn about linguistics"

"Just get the word out to as many schools as possible. You could also advertise on forums like AOPS, Cogito, and even CollegeConfidential ... where students are looking for intellectual challenges"

"The problems helped develop the basic code breaking."

"Having a camp would be a huge benefit, but otherwise I think the contest was done very well. Thank you for bringing it to the US."

"Maybe send a press release to school newspapers and ask them to print something about it."

"My 9 students enjoyed participating even though none of them made it to the second round. Several have indicated that they want to do it again next year now that they know what it is like."

"I used every opportunity to utter the phrase "computational linguistics" to other administrators, at meetings, with parents, students, other teachers. People inevitably want to know more!"

"As I mentioned previously, we are all set to start up a new math/WL club next year. YAY!"

"Advertise with world language professional organizations (i.e., ACTFL) and on our ListServs (i.e., FLTeach)"

"It was wonderful. KUDOS!"

"There were several practice sessions, about half run by a math teacher (who organizes many of the competitions of this nature) and half by the Spanish teacher. Also, several of the English teachers got really excited about it (especially the teacher who teaches AP English Language, who teaches often about logical reasoning) and offered extra credit to the students who took it."

"The preparation for the naclo was done entirely by the math club."

"It was a very useful competition. First, it raised awareness about linguistics among our students. They knew nothing about this area before, and now they are looking for opportunities to study linguistics and some started visiting linguistic research seminars at the University of Washington."

"The Olympiad was interesting to most students because it was very different from all the other math Olympiads we participate in. Students saw possibilities for other application of their general math skills. In addition, the students who won (reasonably succeeded in) this Olympiad were not the same students that usually win math contests at our school. This was very useful for their confidence, and showed everybody that broadening skills is important."

"I was the only one to take the contest from my school, so it didn't really increase awareness that much. I, however, learned a lot about linguistics, and the people who I told about the contest seemed to find it interesting also."

"As a result of this competition, an Independent-Study Linguistics Course was offered this spring for a few interested students."

"Three students who participated in NACLO are now doing an Independent Study course with my colleague from the World Languages dept (who had a linguistics course in college)"

“I’d like to see more linguistic indoctrination, so that math nerds are converted over to the good side.”

“next year I will teach a Computational Linguistics seminar”

Appendix B. Related URLs

```
http://www.naclo.cs.cmu.edu/  
http://www.cogito.org/ContentRedirect.aspx?  
ContentID=16832  
http://www.cogito.org/Interviews/  
InterviewsDetail.aspx?ContentID=16901  
http://www.ilolympiad.spb.ru/  
http://cty.jhu.edu/imagine/PDFs/Linguistics.pdf  
http://www.nsf.gov/news/news_summ.jsp?  
cntn_id=109891  
http://photofile.name/users/anna_stargazer/2949079/
```

Figure 5: List of additional references URLs.

Appendix C. Sample problems

We include here some sample problems as well as one solution. The rest of the solutions are available on the NACLO Web site.

C.1. Molistic

This is a problem from 2007 written by Dragomir Radev and based on [Hatzivassiloglou and McKeown 1997].

Imagine that you heard these sentences:

Jane is molistic and slatty.
Jennifer is cluvious and brastic.
Molly and Kyle are slatty but danty.
The teacher is danty and cloovy.
Mary is blitty but cloovy.
Jeremiah is not only sloshful but also weasy.
Even though frumсы, Jim is sloshful.
Strungy and struffy, Diane was a pleasure to watch.
Even though weasy, John is strungy.
Carla is blitty but struffy.
The salespeople were cluvious and not slatty.

1. Then which of the following would you be likely to hear?

- Meredith is blitty and brastic.
- The singer was not only molistic but also cluvious.
- May found a dog that was danty but sloshful.

2. What quality or qualities would you be looking for in a person?

- blitty
- weasy
- sloshful
- frumсы

3. Explain all your answers. (Hint: The sounds of the words are not relevant to their meanings.)

Figure 6: “Molistic” problem from 2007.

C.2. Garden Path

This is another problem from 2007.

True story: a major wireless company recently started an advertising campaign focusing on its claim that callers who use its phones experience fewer dropped calls.

The billboards for this company feature sentences that are split into two parts. The first one is what the recipient of the call hears, and the second one - what the caller actually said before realizing that the call got dropped. The punch line is that dropped calls can lead to serious misunderstandings. We will use the symbol // to separate the two parts of such sentences.

- Don't bother coming // early.
- Take the turkey out at five // to four.
- I got canned // peaches.

These sentences are representative of a common phenomenon in language, called "garden path sentences". Psychologically, people interpret sentences incrementally, before waiting to hear the full text. When they hear the ambiguous start of a garden path sentence, they assume the most likely interpretation that is consistent with what they have heard so far. They then later backtrack in search of a new parse, should the first one fail.

In the specific examples above, on hearing the first part, one incorrectly assumes that the sentence is over. However, when more words arrive, the original interpretation will need to be abandoned.

- All Americans need to buy a house // is a large amount of money.
- Melanie is pretty // busy.
- Fat people eat // accumulates in their bodies.

1. Come up with two examples of garden path sentences that are not just modifications of the ones above and of each other. Split each of these two sentences into two parts and indicate how hearing the second part causes the hearer to revise his or her current parse.

For full credit, your sentences need to be such that the interpretation of the first part should change as much as possible on hearing the second part. For example, in sentence (6) above, the interpretation of the word "fat" changes from an adjective ("fat people") to a noun ("fat [that] people eat..."). Note: sentences like "You did a great job... // NOT!" don't count.

2. Rank sentences (4), (5), (6) as well as the two sentences from your solution to H1 above, based on how surprised the hearer is after hearing the second part. What, in your opinion, makes a garden path sentence harder to process by the hearer?

Figure 7: “Garden Path” problem from 2007.

