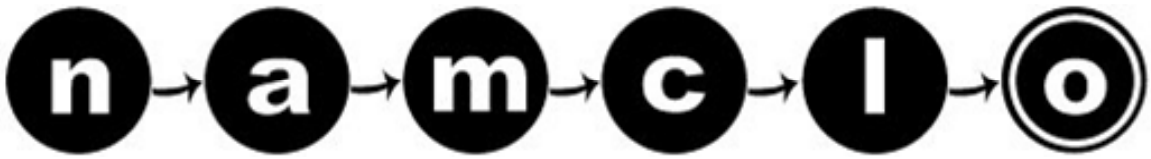


The First Annual

PITTSBURGH

Carnegie Mellon

University of Pittsburgh

INTERNET

Organized by the

University of Michigan

North American Computational Linguistics Olympiad

ITHACA

Cornell University

BOSTON

Brandeis University

March 29, 2007

National Science Foundation

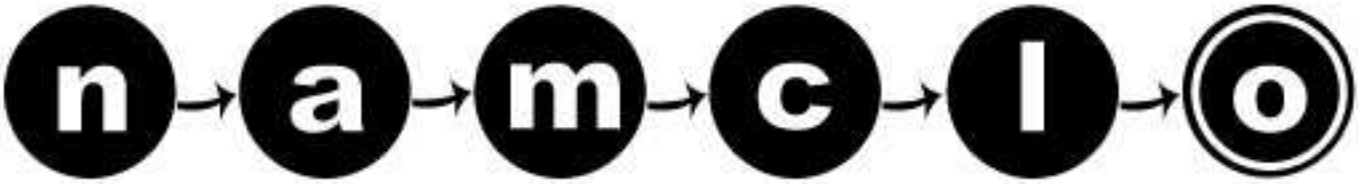
Google™

The Linguist List

The Association for Computational Linguistics

North American Chapter

Cambridge University Press



The North American Computational Linguistics Olympiad
www.namclo.org

Contest Booklet

Your Name: _____
Your School: _____
City, State, Zip: _____
Your Grade: _____
Your E-mail: _____

Start Time: _____
End Time: _____

Your Teacher's Name: _____
Your Teacher's Phone: _____
Your Teacher's E-mail: _____
Your Teacher's Signature: _____

March 29, 2007

(10 points)

(A) We are all molistic in a way

Imagine that you have 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 frumsy, 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.

A1. Then which of the following would you be likely to hear?

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

A2. What quality or qualities would you be looking for in a person?

a. blitty
b. weasy
c. sloshful
d. frumsy

A3. Explain all your answers. (Hint: The sounds of the words are not relevant to their meanings)

(10 points)

(B) Pooh's encyclopedia

Once upon a time, a very long time ago, Winnie-the-Pooh and his friends bought an electronic encyclopedia, and tried to find answers to several important questions:

Winnie-the-Pooh:
Where should a bear stock his jars of honey?
How much honey should a bear store for the winter?

Eeyore:
Where should I look for my lost tail?
Which animals sleep during the winter?

Christopher Robin:
What is the shortest way from my place to the house of Winnie-the-Pooh?
Who wrote the books about Pooh Bear?

The encyclopedia's search engine identified a number of articles related to their questions; for example, it returned the following matches:

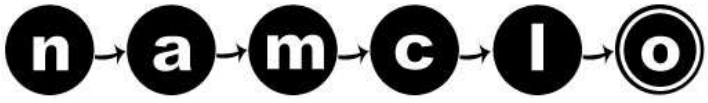
Winter food storage (for Winnie-the-Pooh)
Sleep patterns in mammals and other animals (for Eeyore)
Short stories and movies about Winnie-the-Pooh (for Christopher Robin)
Writers of children's books (for Christopher Robin)

On the other hand, the search engine missed several other relevant articles; in particular, it did *not* retrieve the following articles:

Planning of food supplies
Lost-and-found agencies
Finding shortest paths on a map
Biography of A.A. Milne, the author of Winnie-the-Pooh

Your task is to determine who received each of the following matches; two of these matches were for Winnie-the-Pooh, two for Eeyore, and two for Christopher Robin. Explain why!

Books about care and feeding of bears
Effects of honey on the sleep quality of humans and animals
Lost tales of "Bulls vs. Bears" stock trading
Ways to look for lost things
Ways to store food in the house
Winter hibernation of bears and rodents



(15 points)

(C) A donkey in every house

Consider these phrases in Ancient Greek (in a Roman-based transcription) and their unordered English translations:

(A) ***ho tōn hyiōn dulos***

(B) ***hoi tōn dulōn cyrioi***

(C) ***hoi tu emporu adelphoi***

(D) ***hoi tōn onōn emporoi***

(E) ***ho tu cyriu onos***

(F) ***ho tu oicu cyrios***

(G) ***ho tōn adelphōn oicos***

(H) ***hoi tōn cyriōn hyioi***

(1) the donkey of the master
(2) the brothers of the merchant
(3) the merchants of the donkeys
(4) the sons of the masters
(5) the slave of the sons
(6) the masters of the slaves
(7) the house of the brothers
(8) the master of the house

C1. Place the number of the correct English translation in the space following each Greek sentence. Explain your answers!

C2. Translate into Ancient Greek:

the houses of the merchants;
the donkeys of the slave

Explain your answers!

Note: The letter ***ō*** stands for a long ***o***.

(20 points)

(D) Hmong

Hmong Daw (which belongs to the Hmong Mien language family, along with several other Hmong languages) is spoken by approximately 165 thousand people in south-eastern China, Laos, Thailand, Vietnam, and some other countries.

In the 1960s Shong Lue Yang, a peasant from the Hmong Daw nation (also known as White Miao), invented an original writing system for his native language. This writing system is still in use alongside a Roman-based alphabet created by Christian missionaries.

Here are several words and phrases in the Hmong Daw language written in Shong Lue Yang's script and the missionaries' alphabet, as well as their English translations:

1.		**kev ntsuas no**	degree
2.		**hauv**	inside
3.		**raug raws cai**	legal
4.		**hloov mus**	transfer
5.		**qhua**	guest
6.		**yog los nag**	it is raining
7.		**kwv yees**	guess
8.		**ris ceg luv**	Bermuda shorts

In the missionaries' alphabet the letter **w** stands for a specific vowel. The letters **g**, **s** and **v** at the ends of the syllables aren't consonants; instead, they denote the so-called tones (specific ways of pronouncing the vowels).

D1. Write in the missionaries' alphabet (and explain):

9.		bird
10.		lobster
11.		speak
12.		dizzy

D2. Write in Shong Lue Yang's script (and explain):

13.	**hluav**	ash
14.	**li cas**	how?
15.	**neeg ntse**	smart, wise
16.	**yawg**	grandfather

(5 points)

(E) Better sorry than shunk

Here is an English sentence with a nonsense verb in it (*in italics*):

"After the monster had *shunk* its prey, it dragged it back into the cave."

E1. Fill in the other forms of this verb in the following sentences:

"She used to _____ groundhogs."
"Now she _____ possums for a living."
"When she was in Eugene she _____ thirty-three possums in one day."
"Then she took us possum-_____ in the Cascades."

E2. Are there any other possible solutions to this problem? Please give all solutions, sorted by how likely they are correct, and explain your answer.

(20 points)

(F) The lost tram

Consider these three text fragments:

(1)
The tram makes no stops; you sit clown and are served; there are no further intrusions, no late-corners, no one hurrying to get off. The businessmen leaf through their financial reports, the lady with the hatbox is alone with her novel and her sirloin. Diners reading: you never see that on a plane. When the coast approaches arid dinner is over, everyone retires to his compartment to he transferred to the boat in peace, horizontally.

(Sunrise With Seamonsters, by Paul Theroux)

(2)
Usually, Howie could legitimately claim to have no dear of any man or beast… Howie knew in his heart that it was he vulnerable positions he ended up in that scared him. He was used to operating from a position of strength, either real or projected. Now here he was, injured and alone, standing with and empty handgun in an open filed, while hid opponent or opponents fried their weapon from behind solid cover.

(Rough Justice, by Mark Johnstone)

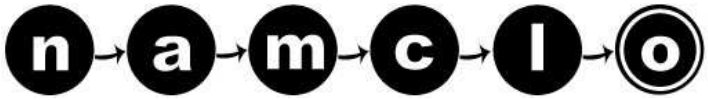
(3)
Two other factors effect the body's temperature regulation: age and acclimatization. As we grow older, we loose our ability to quickly regulate temperature… Very small children are also subject to heat disorders. There small size allows them to take on heat much faster then adults. They also cannot indicate their thirst, accept through irritability. They are completely dependent upon adults to make certain they get enough fluids.

(Doctor in the House: Your Best Guide to Effective Medical Self-Care, by John Harbert)

F1. Each text contains some deviations from what the original author wrote. Try to find all the deviations and restore the original text.

F2. For each text, explain why the deviations occurred.

F3. Could you use a computer program to fix deviations of these types? If yes, how should it work?



(5 points)

(G) Rewrite me badd

You speak a little differently than your parents do. They probably say that you're speaking "bad English". Every generation of parents says this, but this is just how language works. In fact, this is where languages come from: enough generations of young people speaking "bad Latin" and eventually you have Spanish, French, and Italian!

Huishu is a language in the Tangkhulic family and is spoken in the easternmost part of India. Over time, enough changes occurred in this one village that the villagers now speak a different language than any of their neighboring villages. So where they used to say "-lo" ("buy"), they now say "-lu", and where they used to say "-muk" ("cattle") they now say "-mu?". (That symbol at the end represents the sound in the middle of "Uh-oh!", and the dashes in front just mean that these have to occur as parts of larger words.)

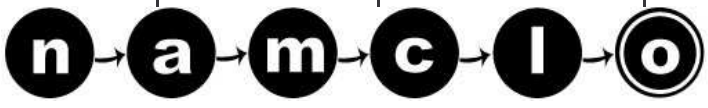
Linguists model historical sound changes as "string-rewrite rules". These are very much like a "find-and-replace" procedure in a word processor: look for one character or pattern and replace it with another one. As the old language changed into modern Huishu, the following string-rewrite rules applied:

K-Insertion: When you find an [u] at the end of the word, add a [k] after it.
Vowel-Raising: When you find an [o] at the end of the word, replace it with [u].
K-Deletion: When you find a [k] at the end of a word, replace it with [ʔ].

These changes didn't all just happen at once, though. They happened one after another – although not necessarily in the order above! – and we can see in which order they happened by comparing the old forms to the new forms. Only one order will work; if these changes had happened in any other order, we would have different modern words.

Here are a few such pairs (the old form is at the top; the new one is at the bottom). From these, can you determine the order in which the above changes must have occurred? Write the names of the rules in the blanks on the left. The blanks in between each pair are for your benefit: if you write how each word changed as each rule applied, you should be able to work out their ordering in time.

Proto-Tangkhulic form:	**-ru** ("bone")	**-khuk** ("knee")	**-ko** ("nine")
Rule 1: _____			
Intermediate form 1:	_____	_____	_____
Rule 2: _____			
Intermediate form 2:	_____	_____	_____
Rule 3: _____			
Huishu form:	**-ruk**	**-khuʔ**	**-ku**



(15 points)

(H) This problem is pretty // easy

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.

(1) Don't bother coming // early.
(2) Take the turkey out at five // to four.
(3) 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.

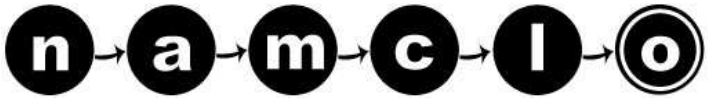
In the specific examples above, on hearing the first part, one incorrectly assumes that the sentence is over. However, when more words arrive, the original interpretation will need to be abandoned.

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

H1. 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.

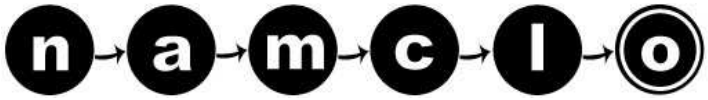
For full credit, your sentences need to be such that the interpretation of the first part should change as much as possible on hearing the second part. For example, in sentence (6) above, the interpretation of the word "fat" changes from an adjective ("fat people") to a noun ("fat [that] people eat...").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

H2. 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?



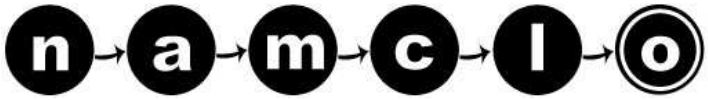
Problem A (your solution)

your name:



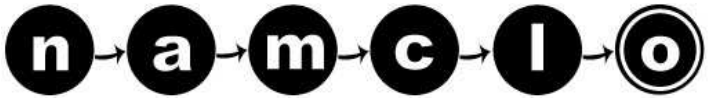
Problem B (your solution)

your name:



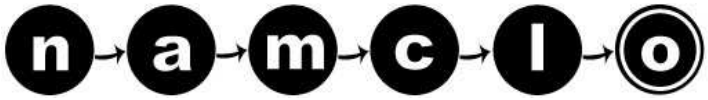
Problem C (your solution)

your name:



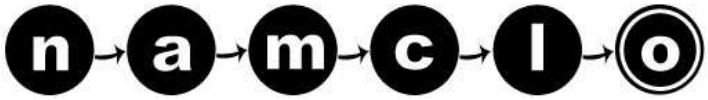
Problem D (your solution)

your name:



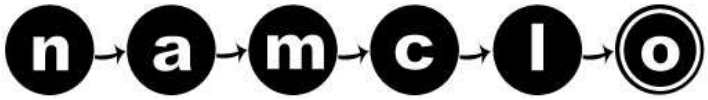
Problem E (your solution)

your name:



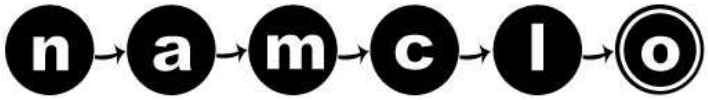
Problem F (your solution)

your name:



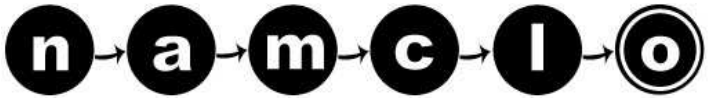
Problem G (your solution)

your name:



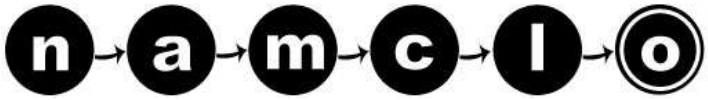
Problem H (your solution)

your name:



Problem ___ (your solution)

your name:



NAMCLO 2007 organizers

General co-chairs:
Lori Levin, Carnegie Mellon University
Thomas Payne, University of Oregon

Program chair:
Dragomir Radev, University of Michigan

Program committee:
Emily M. Bender, University of Washington
John Blatz, Johns Hopkins University
Ivan Derzhanski, Bulgarian Academy of Sciences
Jason Eisner, Johns Hopkins University
Eugene Fink, Carnegie Mellon University
Boris Iomdin, Russian Academy of Sciences
Mahesh Joshi, Carnegie Mellon University
Anagha Kulkarni, Carnegie Mellon University
William Lewis, University of Washington
Patrick Littell, University of Pittsburgh
Ruslan Mitkov, University of Wolverhampton
Thomas Payne, University of Oregon
James Pustejovsky, Brandeis University
Roy Tromble, Johns Hopkins University

Follow up chair:
Barbara Di Eugenio, University of Illinois, Chicago

Outreach chair:
William Lewis, University of Washington

Administrative assistant:
Mary Jo Bensasi, Carnegie Mellon University

School liaison:
Amy Troyani, Taylor Allderdice High School

Practice session chairs:
Rebecca Hwa, University of Pittsburgh
Noah Smith, Carnegie Mellon University

Web site:
Maria Moreno-Rollins, Eastern Michigan University
Helen Aristar-Dry, Eastern Michigan University

Problem credits:
Dragomir Radev (problems A and H)
Eugene Fink (problem B)
Todor Tchervenkov (problem C)
Ivan Derzhanski (problem D)
Jason Eisner and Roy Tromble (problem E)
Boris Iomdin (problem F)
Patrick Littell and David Mortensen (problem G)

Additional thanks to:
Steve Abney, University of Michigan
Richard Sproat, University of Illinois
Many testers

Site coordinators:
Claire Cardie, Cornell University (Ithaca site)
Lillian Lee, Cornell University (Ithaca site)
Lori Levin, Carnegie Mellon University (Pittsburgh site)
James Pustejovsky, Brandeis University (Boston site)
Dragomir Radev, University of Michigan (remote participants)