n a c l o

LSA

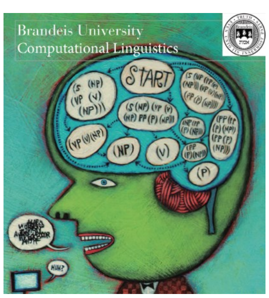
The Association for Computational Linguistics
North American Chapter

Carnegie Mellon

YAHOO!

National Science Foundation

MICHIGAN



Choosito! ©
search and learn

The Ninth Annual

North American Computational Linguistics Olympiad

2015

www.nacloweb.org

Invitational Round

March 12, 2015

Welcome to the ninth annual North American Computational Linguistics Olympiad! You are among the few, the brave, and the brilliant, to participate in this unique event. In order to be completely fair to all participants across North America, we need you to read, understand, and follow these rules completely.

Rules

1. The contest is four hours long and includes eight problems, labeled I to P; there is no break.
2. Follow the facilitators' instructions carefully.
3. If you want clarification on any of the problems, talk to a facilitator. The facilitator will consult with the jury before answering.
4. You may not discuss the problems with anyone except as described in items 3 & 12.
5. Each problem is worth a specified number of points, with a total of 100 points. Make sure to fill out all the answer boxes properly. You may be expected to include explanations for some problems in this round.
6. All your answers should be in the Answer Sheets at the end of this booklet. ONLY THE ANSWER SHEETS WILL BE GRADED.
7. Write your name and registration number on each page:
   Here is an example: Jessica Sawyer #850
8. The top 100 participants (approximately) across the continent in the open round will be invited to the second round.
9. Each problem has been thoroughly checked by linguists and computer scientists as well as students like you for clarity, accuracy, and solvability. Some problems are more difficult than others, but all can be solved using ordinary reasoning and some basic analytic skills. You don’t need to know anything about linguistics or about these languages in order to solve them.
10. If we have done our job well, very few people will solve all these problems completely in the time allotted. So, don’t be discouraged if you don’t finish everything.
11. If you have any comments, suggestions or complaints about the competition, we ask you to remember these for the web-based evaluation. We will send you an e-mail shortly after the competition is finished with instructions on how to fill it out.
12. **DO NOT DISCUSS THE PROBLEMS UNTIL THEY HAVE BEEN POSTED ONLINE! THIS MAY BE SEVERAL WEEKS AFTER THE END OF THE CONTEST.**

Oh, and have fun!

NACLO 2015 Organizers

Program Committee:

Susan Barry, Manchester Metropolitan University
Aleka Blackwell, Middle Tennessee State University
Jordan Boyd-Graber, University of Colorado
Alan Chang, University of Chicago
Dorottya Demszky, Princeton University
John Denero, Google
Jason Eisner, Johns Hopkins University
Caroline Ellison, Stanford University
Josh Falk, University of Chicago
Anatole Gershman, Carnegie Mellon University
Linus Hamilton, Massachusetts Institute of Technology
Lars Hellan, Norwegian University of Science and Technology
Adam Hesterberg, Massachusetts Institute of Technology (chair)
Jordan Ho, University of Toronto
Dick Hudson, University College London
Alex Iriza, Princeton University
Rowan Jacobs, University of Chicago
Wesley Jones, University of Chicago
Ben King, University of Michigan
Greg Kondrak, University of Alberta
Jonathan Kummerfeld, University of California, Berkeley
Mary Laughren, University of Queensland
Lori Levin, Carnegie Mellon University
Richard Littauer
Patrick Littell, University of British Columbia
Daniel Lovsted, McGill University
Jonathan May, Information Sciences Institute
Tom McCoy, Yale University
Rachel McEnroe, University of Chicago
David Mortensen
Heather Newell, Université du Québec à Montréal
Babette Newsome, Aquinas College
David Palfreyman, Zayed University
James Pustejovsky, Brandeis University
Dragomir Radev, University of Michigan
Verna Rieschild, Macquarie University
Catherine Sheard, Oxford University
Ben Sklaroff, University of California, Berkeley
Harold Somers, All Ireland Linguistics Olympiad
Chelsea Voss, Massachusetts Institute of Technology
Alex Wade, Stanford University

NACLO 2015 Organizers (cont'd)

Problem Credits:

Problem I: Graeme Trousdale
Problem J: Dorottya Demszky and Adam Hesterberg
Problem K: Daniel Lovsted
Problem L: Tom McCoy
Problem M: John de Nero
Problem N: Alexander Wade and Pat Littell
Problem O: David Mortensen
Problem P: Emily Bender

Organizing Committee:

Mary Jo Bensasi, Carnegie Mellon University
Aleka Blackwell, Middle Tennessee State University
Janis Chan, University of Western Ontario
Dorottya Demszky, Princeton University
Caroline Ellison, Stanford University
Josh Falk, University of Chicago
Matthew Gardner, Carnegie Mellon University
Harry Go, Washington University in St. Louis
Adam Hesterberg, Massachusetts Institute of Technology
Jordan Ho, University of Toronto
Simon Huang, University of Waterloo
Alex Iriza, Princeton University
Wesley Jones, University of Chicago
Ben King, University of Michigan
Aaron Klein, Harvard University
Andrew Lamont, Indiana University
Stella Lau, University of Cambridge
Lori Levin, Carnegie Mellon University (chair)
Jeffrey Ling, Harvard University
Patrick Littell, University of British Columbia
Daniel Lovsted, McGill University
Tom McCoy, Yale University
Rachel McEnroe, University of Chicago
Graham Morehead, University of Maine
David Mortensen
Helen Mukomel, Carnegie Mellon University
Heather Newell, Université du Québec à Montréal
David Penco, University of British Columbia
James Pustejovsky, Brandeis University
Dragomir Radev, University of Michigan
Alex Wade, Stanford University
Yilu Zhou, Fordham University

NACLO 2015 Organizers (cont'd)

Website and Registration:
Graham Morehead, University of Maine

US Team Coaches:
Dragomir Radev, University of Michigan (head coach)
Lori Levin, Carnegie Mellon University (coach)

Canadian Coordinator and Team Coach:
Patrick Littell, University of British Columbia

USA Contest Site Coordinators:
Bemidji State University: Porter Coggins
Brandeis University: James Pustejovsky
Brigham Young University: Dirk Elzinga, Deryle Lonsdale
Carnegie Mellon University: Mary Jo Bensasi, Lori Levin
College of William and Mary: Anya Lunden
Columbia University: Amy Cooper, Kathy McKeown
Cornell University: Abby Cohn, Sam Tilsen
Dartmouth College: Sravana Reddy
Emory University: Jinho Choi, Phillip Wolff
Georgetown University: Daniel Simonson
Goshen College: Peter Miller
Johns Hopkins University: Mark Dredze
Massachusetts Institute of Technology: Adam Hesterberg, Sophie Mori
Middle Tennessee State University: Aleka Blackwell
Minnesota State University Mankato: Rebecca Bates, Dean Kelley
Northeastern Illinois University: J. P. Boyle, R. Hallett, Judith Kaplan-Weinger, K. Konopka
Ohio State University: Micha Elsner, Michael White
Princeton University: Dorottya Demszky, Christiane Fellbaum, Alex Iriza
San Diego State University: Rob Malouf
San Jose State University: Maya Sun
Stanford University: Sarah Yamamoto
Stony Brook University: Kristen La Magna, Lori Repetti
Union College: Kristina Striegnitz, Nick Webb
University of Alabama, Birmingham: Steven Bethard
University of Colorado at Boulder: Silva Chang
University of Houston: Thamar Solorio
University of Illinois at Urbana-Champaign: Julia Hockenmaier
University of Maine: George Markowsky
University of Maryland: Julia Buffinton, Tom Roberts
University of Memphis: Vasile Rus
University of Michigan: Steven Abney, Sally Thomason
University of North Carolina, Charlotte: Wlodek Zadrozny
University of North Texas: Genene Murphy, Rodney Nielsen
University of Notre Dame: David Chiang
University of Pennsylvania: Chris Callison-Burch, Cheryl Hickey, Mitch Marcus
University of Southern California: Ashish Vaswani
University of Texas: Stephen Weschler
University of Texas at Dallas: Vincent Ng
University of Utah: Aniko Czirmaz, Mengqi Wang, Andrew Zupon
University of Washington: Jim Hoard
University of Wisconsin, Eau Claire: Lynsey Wolter
University of Wisconsin, Madison: Steve Lacy
University of Wisconsin, Milwaukee: Joyce Boyland, Hanyon Park, Gabriella Pinter
Western Michigan University: Shannon Houtrouw, John Kapenga
Western Washington University: Kristin Denham
Yale University: Aidan Kaplan, Alexa Little, Tom McCoy, Raffaella Zanuttini

NACLO 2015 Organizers (cont'd)

Canada Contest Site Coordinators:
Dalhousie University: Magdalena Jankowska, Vlado Keselj, Armin Sajadi
McGill University: Junko Shimoyama, Michael Wagner
Simon Fraser University: John Alderete, Marion Caldecott, Maite Taboada
University of Alberta: Herbert Colston, Sally Rice
University of British Columbia: David Penco, Hotze Rullman
University of Lethbridge: Yllias Chali
University of Ottawa: Diana Inkpen
University of Toronto: Jordan Ho, Pen Long
University of Western Ontario: Janis Chan

High school sites: Dragomir Radev

Booklet Editors:
Andrew Lamont, Indiana University
Dragomir Radev, University of Michigan

Sponsorship Chair:
James Pustejovsky, Brandeis University

Corporate, Academic, and Government Sponsors:
Linguistic Society of America
The North American Chapter of the Association for Computational Linguistics
Carnegie Mellon University
Yahoo!
The National Science Foundation
The University of Michigan
Brandeis University Computer Science Department
Choosito
Many generous individual donors

Special thanks to:
Tatiana Korelsky, Joan Maling, and D. Terrence Langendoen, US National Science Foundation
James Pustejovsky for his personal sponsorship
And the hosts of the 100+ High School Sites

NACLO 2015
Sites

University of Notre Dame
Goshen College
University of Colorado Boulder
University of Wisconsin Milwaukee
Middle Tennessee State University
Michigan
Cornell University
Carnegie Mellon
Brandeis
University of Washington
University of Southern California
University of Lethbridge
College of William & Mary
Emory University
San Diego State University
McGill
Massachusetts Institute of Technology
Western Michigan University
Minnesota State University Mankato
UTD
SFU Simon Fraser University
uOttawa
UNC Charlotte
The University of Utah
Union College
Penn University of Pennsylvania
SJSU San José State University
BYU Brigham Young University
Columbia University
The University of Maine
University of Maryland
Stanford University
Illinois University of Illinois at Urbana-Champaign
UAB The University of Alabama at Birmingham
University of North Texas
Bemidji State University
Georgetown University
Western Washington University
UBC
Ohio State
Western
The University of Wisconsin Madison
Stony Brook State University of New York
Yale
Johns Hopkins University
Northeastern Illinois University
University of Wisconsin Eau Claire
University of Houston
Dartmouth
University of Alberta
The University of Memphis
Dalhousie University Inspiring Minds
The University of Texas at Austin
University of Toronto
Princeton University

As well as more than 90 high schools throughout the USA and Canada

(I) Old English (1/1) [5 points]

English has changed a lot since the period in which Germanic languages were brought to the British Isles over 1500 years ago. In this puzzle, you will look at some ways in which Old English (the name we give to the varieties of English that existed from the 5th century to the 12th century) differs from Modern English.

Note: The letter þ is pronounced like the 'th' in *thin*; the letter æ is pronounced like the 'a' in *cat*.

Old English	Modern English
wit lufodon þæt mægden	we two loved the girl
þæt cild unc lufode	the child loved us two
ge lufodon þone cyning	you all loved the king
se cyning inc lufode	the king loved you two
þæt mægden we lufodon	we all loved the girl
we inc lufodon	we all loved you two
wit eow lufodon	we two loved you all
se æþeling unc lufode	the prince loved us two

Answer these questions in the Answer Sheets.

I1. Translate the following from Old English into Modern English.

a. se cyning eow lufode
b. ge lufodon þæt mægden
c. wit inc lufodon

I2. Translate the following from Modern English into Old English.

a. The prince loved the child.
b. The child loved the prince.
c. We all loved the child.
d. The child loved you two.

(J) Georgian Transitive Verbs (1/1) [20 points]

Georgian (Kartuli) is the official language of the country Georgia that is unrelated to neighboring Russian and Turkish. Below are twelve Georgian sentences with their English translations. (sg.) stands for singular; (pl.) stands for plural.

1.	კარს ვაღებ.	I am opening the door.
2.	სახლს აშენებ.	You (sg.) are building a house.
3.	ედუარდ წერილს იღებს.	Eduard is receiving a letter.
4.	ფარდულს ვაშენებთ.	We are building a barn.
5.	საჩუქარს აღებთ.	You (pl.) are opening a gift.
6.	ანდრია და მისი დედა პაკეტს იღებენ.	Andria and his mother are receiving a package.
7.	ხაჭაპური გავაკეთე.	I have made khachapuri[1].
8.	კარი შეაკეთე.	You (sg.) have repaired the door.
9.	ანდრიამ პოემა დაწერა.	Andria has written a poem.
10.	ლევან და მე აკტომობილი შევაკეთეთ.	Levan and I have repaired the car.
11.	სადილი გააკეთეთ.	You (pl.) have made dinner.
12.	ედუარდ და ეკატერინემ წერილი დაწერეს.	Eduard and Ekaterine have written a letter.

Answer these questions in the Answer Sheets.

J1. Translate the following sentences into English:

a. წერილს აღებენ.
b. კომპუტერი შეაკეთეთ.
c. ადამ მუზეუმს აშენებს.
d. ისტორია დავწერეთ.

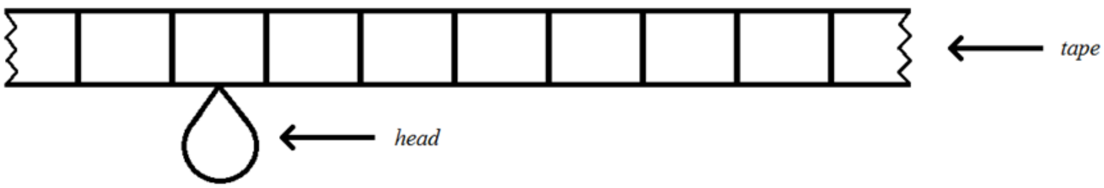
J2. Translate the following sentences into Georgian:

a. You (sg.) are opening a barn.
b. I have repaired the car.
c. We are receiving dinner.
d. They have made a gift.

[1] Khachapuri is a traditional Georgian dish of cheese-filled bread.

(K) The Dualization Game (1/5) [15 points]

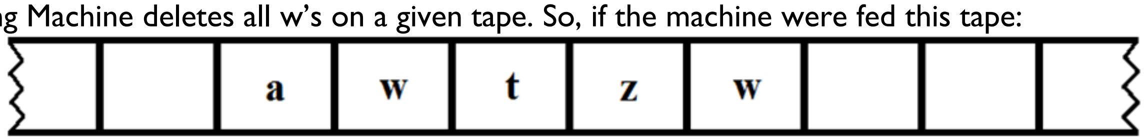
Turing Machines are a type of abstract computing machine first described by Alan Turing[1] in 1936. Although they have a very simple design, Turing Machines are very powerful – in fact, every computational task that a modern computer is capable of can also (theoretically) be done by a Turing Machine.



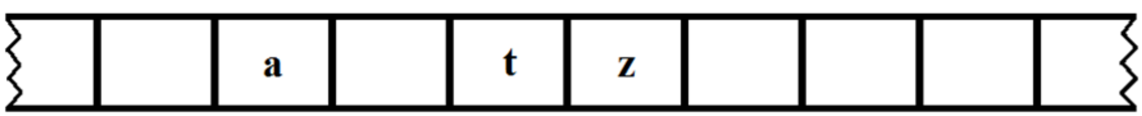
Turing Machines consist of a tape (a series of cells, infinite in both directions, each containing a blank or a symbol), and a head, which reads a particular cell on the tape and performs an operation according to the 'state' it is in: either writing something in that cell, moving left or right, both of these, or neither. A Turing Machine is defined by its instructions, which determine what operations it performs. Below are the instructions for a particular Turing Machine. Note that the symbol Ø indicates a blank on the tape; it is not the zero in state names (0):

Entry state	Read	Write	Move	Exit State
S0	w	Ø	R	S0
S0	Ø	[N/A]	[N/A]	HALT
S0	[otherwise]	[N/A]	R	S0

This Turing Machine deletes all w's on a given tape. So, if the machine were fed this tape:



…the tape would look like this when the machine was finished:



This transformation can be summarized as:

awtzw ⇒ atz (blanks within the letter sequence are not transcribed)

Some things to note:

- Turing Machines always start on the leftmost non-blank space on the tape
- Input tapes always contain a single string of symbols, unbroken by blanks
- A Turing Machine will only stop if it arrives at a HALT state or if there are insufficient instructions to proceed
- The initial state of a Turing Machine is always S0

Turing Machines can operate on strings of 0's and 1's, on arbitrary strings of letters (like above), or on words – in this last case, Turing Machines can be used to perform useful linguistic tasks.

[1]Alan Turing (1912-1954) was a British mathematician and logician who played a crucial role in the foundation of the field of computer science. He was the subject of the recent biopic 'The Imitation Game'.

(K) The Dualization Game (2/5)

Here is a simple Turing Machine designed for the English language, called Pluralizing Machine 1.0:

Entry state	Read	Write	Move	Exit State
S0	∅	s	[N/A]	HALT
S0	[otherwise]	[N/A]	R	S0

This machine makes the following successful (i.e., linguistically valid) transformations:

cat ⇒ cats   apple ⇒ apples   microscope ⇒ microscopes

However, this machine also makes the following unsuccessful (i.e., linguistically invalid) transformation:

* fox ⇒ foxs (* indicates an unsuccessful transformation)

The machines shown so far have used only a single state, S0 (not including the HALT state). Turing Machines that perform more complex tasks, however, will require multiple states, each with its own set of instructions. In multi-state machines, some lines of instructions will cause the machine to change state – in other words, exit the line in a different state than it entered.

Consider Pluralizing Machine 2.0:

Entry state	Read	Write	Move	Exit State
S0	∅	[N/A]	L	S1
S0	[otherwise]	[N/A]	R	S0
S1	x, s, z	[N/A]	R	S2
S1	[otherwise]	[N/A]	R	S3
S2	[otherwise]	e	R	S3
S3	[otherwise]	s	[N/A]	HALT

In addition to making the successful transformations made by Pluralizing Machine 1.0, this machine also makes successful transformations for many new words, including 'fox':

fox ⇒ foxes

Answer these questions in the Answer Sheets.

K1. Give three more English words for which Pluralizing Machine 2.0 makes successful transformations, but that Pluralizing Machine 1.0 transforms unsuccessfully. Try to take advantage of all the added capabilities of the new machine.

K2. Pluralizing Machine 2.0 is not without its faults: what outputs does this machine give for the inputs 'quiz' and 'child'?

Of course, Turing Machines can deal with any written language – not just English. The remaining Turing Machines in this problem perform tasks in Navajo – a language in the Na-Dené family, spoken primarily in Arizona, Utah, and New Mexico. With almost 170,000 speakers, Navajo is the most widely-spoken Indigenous language in the United States.

(K) The Dualization Game (3/5)

Consider the following verb forms from the Navajo language. Note: ł, ', and y are consonants in Navajo. An accent above a vowel, as in é, indicates high tone, pronounced with raised pitch. A hook beneath a vowel, as in ą, indicates that the vowel is nasal, pronounced through the mouth and nose. In the eyes of a Turing Machine, vowels that differ in tone or nasality are entirely different symbols.

Navajo	English
nidaahné	you (pl.) play
dajidlą́	people (pl.) drink it
biyadahodiilyéés	we (pl.) frighten him
áchą́dahídéelni'	they (pl.) are greedy
bidajil'į́	people (pl.) imitate him
nidaniiché	we (pl.) are on the run

Navajo	English
naahné	you (du.) play
jidlą́	people (du.) drink it
biyahodiilyéés	we (du.) frighten him
áchą́hídéelni'	they (du.) are greedy
bijil'į́	people (du.) imitate him
naniiché	we (du.) are on the run

Sam designs a Turing Machine to transform the plural (pl.) form of a Navajo verb into its dual (du.) form. A dual verb has exactly two people/entities as its subject, and in Navajo this form contrasts with singular verbs (one person as subject) and plural verbs (three or more people as subject). Here is Sam's Dualizer Machine 1.0:

Entry state	Read	Write	Move	Exit State
S0	d	Ø	R	S1
S0	[otherwise]	[N/A]	R	S0
S1	[otherwise]	Ø	[N/A]	HALT

K3. Sam's Dualizer Machine 1.0 makes successful transformations for only four of the six plural verbs given above. Identify the other two, for which the machine makes unsuccessful transformations, and show the machine's output in the Answer Sheets.

Here is the outline of Dualizer Machine 2.0:

Entry state	Read	Write	Move	Exit State
S0	d	Ø	R	S1
S0	[otherwise]	[N/A]	R	S0
S1	[otherwise]	(a)	(b)	S2
S2	(c)	[N/A]	(d)	S3
S3	i	[N/A]	(e)	S4
S3	(f)	(g)	[N/A]	HALT
S4	(h)	[N/A]	(i)	S5
S4	[otherwise]	[N/A]	[N/A]	HALT
S5	[otherwise]	(j)	[N/A]	HALT

(K) The Dualization Game (4/5)

K4. Fill in the blanks of Dualizer Machine 2.0. in the Answer Sheets The machine should make successful transformations for all six of the verbs given above.

Next, consider this set of Navajo verbs, with their English translations:

Navajo	English
dádi'nishkaad	I sew it shut
ná'iishgááh	I bleach it
nistséés	I extinguish it
yishchxǫǫh	I destroy it
yishdééh	I scrape it off
hadishbin	I fill it up
ííníshta'	I read
ałk'íisgis	I entwine them
yiists'ił	I break it
yishhį́į́h	I melt it
iishchííh	I dye it red

Navajo	English
dádi'níłkaad	you (sg.) sew it shut
ná'iiłgááh	you (sg.) bleach it
níłtséés	you (sg.) extinguish it
niłchxǫǫh	you (sg.) destroy it
niłdééh	you (sg.) scrape it off
hadíłbin	you (sg.) fill it up
íínílta'	you (sg.) read
ałk'íiłgis	you (sg.) entwine them
yiiłts'ił	you (sg.) break it
niłhį́į́h	you (sg.) melt it
iiłchííh	you (sg.) dye it red

Sam intends to design a Second-Personizer Machine that will transform the "I" form of each verb (the first-person singular form) to the "you (sg.)" (second-person singular) form.

Here is an outline of Sam's Second-Personizer Machine:

Entry state	Read	Write	Move	Exit State
S0	(a)	(b)	(c)	S1
S0	[otherwise]	[N/A]	R	S0
S1	(d)	Ø	(e)	S2
S1	[otherwise]	(f)	L	S2
S2	[otherwise]	(g)	(h)	S3
S3	(i)	[N/A]	[N/A]	HALT
S3	[otherwise]	(j)	L	S4
S4	(k)	[N/A]	[N/A]	HALT
S4	(l)	(m)	[N/A]	HALT
S4	[otherwise]	[N/A]	(n)	S5
S5	[otherwise]	(o)	[N/A]	HALT

(K) The Dualization Game (5/5)

K5. Fill in the blanks of Sam's Second-Personizer Machine in the Answer Sheets. This Machine should be able to successfully transform every verb above from its first-person form to its second-person form.

Here are two last Navajo verbs, with their English translations:

Navajo	English
íísínísts'ą́ą́'	I listen
bigháníshdééh	I sift it (as flour)

K6. Only one of the two verbs above is transformed successfully by the Second-Personizer Machine. Which verb do you think will be transformed unsuccessfully, and what will the machine output for this verb? Write your answer in the Answer Sheets.

(L) Easy-Peasy-Malagasy (1/2) [10 points]

Below is an example of a crossnumber puzzle. This type of puzzle is much like a crossword, except that each square in the grid is meant to hold a single digit rather than a single letter. Also note that the clues are not given in the order that the answers appear in the grid.

English Crossnumber Puzzle Example

1	2	■	3	4
5		6		
■	7			■
8				9
10		■	11	

Clues

Across

Answer #	Answer
Eight	Sixty-seven thousand eight hundred ninety-three
Eleven	Forty-one
Five	Thirty-eight thousand nine hundred one
One	Seventy-four
Seven	Two hundred forty-five
Ten	Fifty-five
Three	Twenty-six

Down

Answer #	Answer
Eight	Sixty-five
Four	Sixty-one
Nine	Thirty-one
One	Seventy-three
Six	Nine hundred forty-eight
Three	Twenty thousand five hundred ninety-four
Two	Forty-eight thousand two hundred seventy-five

The answer to this example puzzle is given here:

^{1} 7	^{2} 4	■	^{3} 2	^{4} 6
^{5} 3	8	^{6} 9	0	1
■	^{7} 2	4	5	■
^{8} 6	7	8	9	^{9} 3
^{10} 5	5	■	^{11} 4	1

(L) Easy-Peasy-Malagasy (2/2)

Here is another crossnumber puzzle; this time, all of the numbers are written in Malagasy, an Austronesian language and one of the official languages of Madagascar. Your task is to complete this puzzle and answer the questions that follow it. Remember, the clues are *not* given in the same order as the answers in the grid. Also, note that no answer has zero as its first digit. Note that you will only be graded on your answers to questions 1a through 1e; you will not be graded on the filled-in grid.

Malagasy Crossnumber Puzzle

1	2	■	3	4
5		6		
7				
8				
9		■	10	

Across

Answer #	Answer
Dimy	Efatra amby enimpolo sy telonjato sy sivy alina
Fito	Dimy ambin'ny folo sy roanjato sy arivo sy fito alina
Folo	Iraika amby fitopolo
Iray	Fito amby fitopolo
Sivy	Folo
Telo	Fito ambin'ny folo
Valo	Valo amby enimpolo sy sivinjato sy dimy arivo sy alina

Down

Answer #	Answer
Efatra	Iraika amby valopolo sy dimanjato sy efatra arivo sy fito alina
Enina	Sivy amby roapolo sy telonjato
Iray	Iraika ambin'ny folo sy fitonjato sy sivy arivo sy fito alina
Roa	Dimampolo sy zato sy fito alina
Telo	Fito amby enimpolo sy zato sy enina arivo sy alina

L1. Write the following numbers in Malagasy in the Answer Sheets.

a. 7
b. 15,968
c. 99,573
d. 80,638
e. 817

(M) Minimum Spelling Trees (1/2) [10 points]

Computing devices are limited by their available memory. One can, however, use clever methods to reduce the amount of memory needed to store a text. If one word is already represented in memory, then a similar word can be represented compactly by recording only the differences from the original word.

The German word for "house" is "Haus". German has multiple ways to say "the house" depending on the syntactic role (case) it plays in the sentence. "Das Haus" is used as the subject or direct object of a sentence (nominative or accusative case): "Das Haus ist grün" (The house is green) or "Tom liebt das Haus" (Tom loves the house). "Des Hauses" is used for possessives (genitive case) as in "Des Hauses Tür" (the house's door). "Dem Haus" (dative case) is used with indirect objects and some locations as in "auf dem Haus" (on the house). In this puzzle, we will just be concerned with the noun Haus itself, not with the articles (das, dem, des). The table below shows the singular and plural forms of Haus in each syntactic role (case).

	singular	plural
nominative	Haus	Häuser[2]
genitive	Hauses	Häuser
dative	Haus, Hause (archaic)	Häusern
accusative	Haus	Häuser

For a computer to remember all of these forms, it must efficiently encode them. To do so, we will encode each of the words in the set based on either an already-encoded word in the set or the empty word, along with a series of changes. The changes specify letters to add or remove, and are the shortest description possible given the pair of words. We write the changes and resulting words after each word already encoded. For example:

(“ “ (1+H 2+a 3+u 4+s →
“ H a u s “ (5+e 6+s →
“ H a u s e s “ (2-a 2+ä[3] 6-s 6+r →
“ H ä u s e r “)
(2-a 2+ä 6-s 6+r 7+n →
“ H ä u s e r n “))))

Above, "Häuser" and "Häusern" are encoded from "Hauses", which is encoded from "Haus", which in turn is encoded from the empty word "". The total cost of encoding the set of words is the total number of character changes overall: 4 + 2 + 4 + 5 = 15.

[1] German also marks case on the article, but that doesn't matter for this problem.

[2] Vowels with umlauts (the two dots) are pronounced differently from their unmarked counterparts. Some nouns in German, like *Haus,* have vowels which change depending on their form. Here, the singular *Haus* has no umlaut, but the plural *Häuser* does. English has a similar phenomenon in words like *foot/feet.*

[3] Note that *a* and *ä* are two different symbols and that *ä* is one symbol, not the combination of *a* and ¨.

(M) Minimum Spelling Trees (2/2)

Answer these questions in the Answer Sheets.

M1. What is the least costly way of encoding the words "Haus", "Hauses", "Häuser", and "Häusern"? What is its cost?

M2. What is the least costly way of encoding all of these words as well as the archaic form "Hause"? What is its cost?

M3. There can be several ways to encode the same set of words. Two ways to encode a set of words are considered different if any word in the set is encoded based on different words in the two ways. For example, if the set of words is {A, B, C}, one way to encode C is based on A and another way to encode C is based on B. Because C is encoded based on different words, these two ways are different. How many different ways are there to encode the five forms of *Haus* with a cost less than 12?

M4. A computer manufacturer just figured out that it will make more money if more memory is required! They ask you: what's the greatest cost encoding of the five forms of *Haus*? What is its cost?

(N) Maxakalí (1/1) [15 points]

Maxakalí is an Amazonian language in the Macro-Jê language family. It is spoken in the eastern Brazilian region of Minas Gerais by about only about 1200 indigenous inhabitants of small villages. In this problem you will have to match words and phrases to their meanings in English; write your answers in the Answer Sheets. Words and phrases in Maxakalí and English have been alphabetized in the table below.

1.	ka'õgãhã	A.	cover
2.	ka'ok	B.	eye
3.	kuxa	C.	eyelid
4.	kuxa ka'ok	D.	foot
5.	mĩkax	E.	to go
6.	mĩkaxxax	F.	hard
7.	mĩptut	G.	to harden
8.	mĩptut mõg	H.	heart
9.	mĩptut mõg kuxa	I.	house
10.	mĩptut mõg pata	J.	knife
11.	mõg	K.	knife sheath
12.	mõgãhã	L.	to lead
13.	pa	M.	motor
14.	pa ka'ok	N.	motor vehicle
15.	pata	O.	shoe
16.	pataxax	P.	to be stubborn
17.	paxax	Q.	tire
18.	xax	R.	wide awake

Pronunciation Notes: A tilde (~) above a vowel means that it is pronounced nasally. The letter <x> represents a sound like English 'sh,' and the letter <'> represents a glottal stop, or a catch in the throat like that represented by the hyphen in 'uh-oh.' Other letters are reasonably similar to their equivalents in Spanish, Japanese, or Latin.

(O) Do-This-Do-That (1/2) [20 points]

Hmong is a language of Southern China and Southeast Asia, especially the countries of Vietnam, Laos, and Thailand. One dialect of Hmong that is spoken in all of these countries is called Hmong Daw (or White Hmong). The sentences and phrases that you see below are from Hmong Daw. They are written in a writing system called RPA (Romanized Popular Alphabet). In this writing system, the initial consonant of a syllable is written first, followed by the vowel. If this vowel is doubled in writing, it is pronounced as nasalized or with a following "ng" sound. Last of all, a consonant symbol is used to represent the tone of the syllable (tone is a meaningful difference in pitch; for instance, Mandarin Chinese differentiates syllables with high, low, rising, and falling pitch). Hmong Daw has seven tones, one of which is indicated by the absence of any symbol. In this system, Hmong is written as *Hmoob*.

O1. In the following section, you are given a set of Hmong Daw sentences and phrases. English translations are presented in no particular order. Your task is to match the Hmong phrases and sentences with English translations; write your answers in the Answer Sheets. You should be aware that the order of words may differ significantly between the original phrases/sentences and their translations into the other language. Also note that the same word in English may translate to more than one word in Hmong, and vice versa.

Abbreviations: sg. = singular; du. = dual (two participants); pl. = plural.

1. Neeg them nyiaj rau koj.
2. Nej sib pom lawm.
3. phem dab phem tuag
4. Kuv lub tsev nqeeb phem heev.
5. ua ib pawg
6. Wb sib ntsib.
7. Khib kuv heev.
8. Dab tsis ntshai.
9. Nwg cog ib tsob tauj.
10. Nej tsis tau pom kuv.
11. Ntuj no lawm.
12. Neeg Nplog pe mlom.
13. tej plaub tej ntug
14. Kuv tsis tau ntsib nwg.
15. tsev hais plaub
16. neeg loj neeg siab
17. hais lus Hmoob
18. sib ntsib sib pom
19. tsis khib tsis chim
20. pe dab pe mlom
21. ib pab liab
22. Koj muaj kub.
23. ntau pab ntau pawg
24. Nej muaj tej tsev loj.
25. Wb phem heev.
26. Neeg ntshai dab.
27. Nwg chim heev.
28. ntsis tauj ntsis nqeeb
29. Cuam zoo li neeg.
30. thov ntuj thov dab
31. ib tsob ntoo siab heev

(O) Do-This-Do-That (2/2)

A. ‘a very tall tree’
B. ‘behave as one group’
C. ‘grass sprouts’
D. ‘He planted a grass plant.’
E. ‘People fear spirits.’
F. ‘some legal proceedings’
G. ‘He is very upset.’
H. ‘It offends me very much.’
I. ‘We (du.) are very ugly.’
J. ‘speak Hmong language’
K. ‘Heaven is cold now.’
L. ‘petition deities’
M. ‘a band of monkeys’
N. ‘You (pl.) did not see me.’
O. ‘You (pl.) see each other now.’
P. ‘ugly as spirits and death’
Q. ‘People pay silver to you.’
R. ‘Lao people reverence images.’
S. ‘not angry or offended’
T. ‘My grass house is very ugly.’
U. ‘big and tall people’
V. ‘Apes are like people.’
W. ‘Spirits don’t fear.’
X. ‘You (sg.) have gold.’
Y. ‘many factions’
Z. ‘I did not meet him.’
AA. ‘worship images and spirits’
BB. ‘court (house of speaking legal proceedings)’
CC. ‘encounter one another’
DD. ‘You (pl.) have some big houses.’
EE. ‘We (du.) met each other.’

O2. Translate these English phrases into Hmong; write your answers in the Answer Sheets.

a. Lao and Hmong people
b. behave as primates
c. have precious metals

(P) The Old Man the Boats (1/1) [5 points]

Ambiguity in sentences can be a source of humor. For example, "flies" can be a noun or a verb and "like" can be a verb or a preposition, resulting in the amusing juxtaposition of

Time flies (verb) like (preposition) an arrow.   Fruit flies (noun) like (verb) a banana.

For this puzzle, we will be concerned with ambiguity in parts of speech (noun, verb, adjective, adverb, preposition) as well as ambiguity in the structure of a sentence. The sentences below are bracketed to show the structure of the sentences.

[[Time] [flies [like [an arrow]]]]   [[Fruit flies] [like [a banana]]]

Local ambiguity arises when there is not enough information midway through a sentence to decide on the parts of speech or the structure. For example, when you hear or see the words "the old man" you don't know whether the sentence will continue with "man" being a verb as in "The old man the boats" or as a noun as in "The old man is wise." The structure of these two sentences is shown below. Notice that in addition to changing its part of speech, the word "man" participates differently in the structure of the sentence. In "The old man the boats", "man" is grouped with "the boats" to make a verb phrase "man the boats", but in "The old man is wise", "man" is grouped with "the old" to make a noun phrase "the old man".

[[The old] [man [the boats]]]   [[The [old man]] [is wise]]

P1. Each of the following sentences contains a local ambiguity. For each sentence, provide an alternate continuation which shows what the other interpretation is. For example, if you were given the sentence "The old man is wise" your answer could be "The old man the boats".

Your answers should always start with at least the first three words of the sentences provided. In some cases, more words should be shared. Your answers should differ from the original sentence in at least one word changing its part of speech or in at least one word being grouped with different words in the structure.

1. The old train the young.
2. The thief seized by the police turned out to be our cousin.
3. I convinced her children to do their homework.
4. The man who whistles tunes pianos.
5. The cotton clothing is drying in the sun.

The North American Computational Linguistics Olympiad
www.nacloweb.org

Contest Booklet

REGISTRATION NUMBER			

Name: ______________________________

Contest Site: ______________________________

Site ID: ______________________________

City, State: ______________________________

Grade: ______

Start Time: ______________________________
End Time: ______________________________

Please also make sure to **write your registration number and your name on each page** that you turn in.

SIGN YOUR NAME BELOW TO CONFIRM THAT YOU WILL NOT DISCUSS THESE PROBLEMS WITH ANYONE UNTIL THEY HAVE BEEN OFFICIALLY POSTED ON THE NACLO WEBSITE IN APRIL.

Signature: ______________________________

YOUR NAME: REGISTRATION #

Answer Sheet (1/6)

(I) Old English

1. a.

b.

c.

2. a.

b.

c.

d.

(J) Georgian Transitive Verbs

1. a.

b.

c.

YOUR NAME: REGISTRATION #

Answer Sheet (2/6)

(J) Georgian Transitive Verbs (cont.)

d.

2. a.

b.

c.

d.

(K) The Dualization Game

1.

2. quiz: child:

3. Verb: Output:

Verb: Output:

4. a. b. c.

d. e. f.

g. h. i.

j.

5. a. b. c.

nacl o

YOUR NAME: REGISTRATION #

Answer Sheet (3/6)

(K) The Dualization Game (cont.)

d. e. f.

g. h. i.

j. k. l.

m. n. o.

6. Verb: Output:

(L) Easy-Peasy-Malagasy

1. a.

b.

c.

d.

naclo

YOUR NAME: REGISTRATION #

Answer Sheet (4/6)

e.

(M) Minimum Spelling Trees

1. Cost:

2. Cost:

3.

YOUR NAME: REGISTRATION #

Answer Sheet (5/6)

(M) Minimum Spelling Trees (cont.)

4. Cost: [ ]

(N) Maxakalí

1. [ ] 2. [ ] 3. [ ] 4. [ ] 5. [ ] 6. [ ] 7. [ ] 8. [ ] 9. [ ] 10. [ ] 11. [ ] 12. [ ]
13. [ ] 14. [ ] 15. [ ] 16. [ ] 17. [ ] 18. [ ]

(O) Do-This-Do-That

1. 1. [ ] 2. [ ] 3. [ ] 4. [ ] 5. [ ] 6. [ ] 7. [ ] 8. [ ] 9. [ ] 10. [ ] 11. [ ]
12. [ ] 13. [ ] 14. [ ] 15. [ ] 16. [ ] 17. [ ] 18. [ ] 19. [ ] 20. [ ] 21. [ ] 22. [ ]
23. [ ] 24. [ ] 25. [ ] 26. [ ] 27. [ ] 28. [ ] 29. [ ] 30. [ ] 31. [ ]

2. a.

b.

c.

YOUR NAME: REGISTRATION #

Answer Sheet (6/6)

(P) The Old Man the Boats

1.

2.

3.

4.

5.

YOUR NAME: REGISTRATION #

Extra Page - Enter the Problem Name Here: ___________