Serious language puzzles that are surprisingly fun!

-Will Shortz, Crossword editor of The New York Times and Puzzlemaster for NPR
Welcome to the tenth 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 ten problems, labeled I to R.
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 & 11.
5. Each problem is worth a specified number of points, with a total of 100 points. In the Invitational Round, some questions require explanations. Please read the wording of the questions carefully.
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 of the Answer Sheets’ Here is an example: Jessica Sawyer #850
8. The top students from each country (USA and Canada) will be invited to the next round, which involves online practices before the international competition in India.
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. **DO NOT DISCUSS THE PROBLEMS UNTIL THEY HAVE BEEN POSTED ONLINE! THIS MAY BE A COUPLE OF MONTHS AFTER THE END OF THE CONTEST.**

Oh, and have fun!
NACLO 2016 Organizers

Program Committee:
Adam Hesterberg, Massachusetts Institute of Technology (co-chair)
Alan Chang, University of Chicago
Aleka Blackwell, Middle Tennessee State University
Alex Iriza, Princeton University
Alex Wade, Stanford University
Andrew Lamont, Indiana University
Babette Newsome, Aquinas College
Ben Sklaroff, University of California, Berkeley
Caroline Ellison, Stanford University
Daniel Lovsteds, McGill University
David Mortensen, Carnegie Mellon University
David Palfreyman, Zayed University
David McClosky, IBM
Dick Hudson, University College London
Dorottya Demszky, Princeton University
Dragomir Radev, University of Michigan (co-chair)
Elycia Warner, University of Cambridge
Greg Kondrak, University of Alberta
Harold Somers, AILO
Harry Go, WUSTL
James Pustejovsky, Brandeis University
Jason Eisner, Johns Hopkins University
Jonathan Kummerfeld, University of California, Berkeley
Jonathan May, ISI
Jonathan Graehl, SDL International
Jordan Boyd-Graber, University of Colorado
Jordan Ho, University of Toronto
Josh Falk, University of Chicago
Julia Buffinton, University of Maryland
Lars Hellan, Norwegian University of Science and Technology
Lori Levin, Carnegie Mellon University
Lynn Clark, University of Canterbury
Mary Laughren, University of Queensland
Michael Erlewine, National University of Singapore
Oliver Sayeed, University of Cambridge
Patrick Littell, Carnegie Mellon University
Rachel McEnroe, University of Chicago
Richard Littauer
Susan Barry, Manchester Metropolitan University
Tom McCoy, Yale University
Tom Roberts, University of California, Santa Cruz
Verna Rieschild, Macquarie University
Wesley Jones, University of Chicago
Yejin Choi, University of Washington
NACLO 2016 Organizers (cont’d)

Problem Credits:
Problem I: Harold Somers and Simona Klemenčič
Problem J: Patrick Littell
Problem K: Tom McCoy
Problem L: Dragomir Radev and Patrick Littell
Problem M: Ollie Sayeed
Problem N: Vica Papp
Problem O: Alex Wade
Problem P: Josh Falk
Problem Q: Tae Hun Lee
Problem R: Harold Somers

Organizing Committee:
Adam Hesterberg, Massachusetts Institute of Technology
Aleka Blackwell, Middle Tennessee State University
Alex Iriza, Princeton University
Alex Wade, Stanford University
Andrew Lamont, Indiana University
Bill Huang, Princeton University
Caroline Ellison, Stanford University
Daniel Lovsted, McGill University
David Mortensen, Carnegie Mellon University
Deven Lahoti, Massachusetts Institute of Technology
Dorottya Demszky, Princeton University
Dragomir Radev, University of Michigan
Harry Go, Washington University in Saint Louis
James Pustejovsky, Brandeis University
James Bloxham, Massachusetts Institute of Technology
Janis Chang, University of Western Ontario
Jordan Ho, University of Toronto
Josh Falk, University of Chicago
Julia Buffinton, University of Maryland
Laura Radev, Harvard University
Lori Levin, Carnegie Mellon University
Mary Jo Bensasi, Carnegie Mellon University
Matthew Gardner, Carnegie Mellon University
Patrick Littell, Carnegie Mellon University
Rachel McEnroe, University of Chicago
Simon Huang, University of Waterloo
Stella Lau, University of Cambridge
Tom McCoy, Yale University
Tom Roberts, University of California, Santa Cruz
Wesley Jones, University of Chicago
Yilu Zhou, Fordham University
NACLO 2016 Organizers (cont’d)

US Team Coaches:
Dragomir Radev, University of Michigan
Lori Levin, Carnegie Mellon University

Canadian Coordinator and Team Coach:
Patrick Littell, Carnegie Mellon University

USA Contest Site Coordinators:
Brandeis University: James Pustejovsky
Brigham Young University: Deryle Lonsdale
Carnegie Mellon University: Lori Levin, Pat Littell, David Mortensen
College of William and Mary: Dan Parker
Columbia University: Brianne Cortese, Kathy McKeown
Cornell University: Abby Cohn, Sam Tilsen
Dartmouth College: Michael Lefkowitz
Emory University: Jinho Choi, Phillip Wolff
Georgetown University: Daniel Simonson
Goshen College: Peter Miller
Indiana University: Markus Dickinson, Andrew Lamont
Johns Hopkins University: Rebecca Knowles
Massachusetts Institute of Technology: Adam Hesterberg, Sophie Mori
Middle Tennessee State University: Aleka Blackwell
Minnesota State University Mankato: Rebecca Bates, Dean Kelley, Richard Roiger
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, Misha Khodak, Caleb South, Yuyan Zhao
San Diego State University: Rob Malouf
Sega Math Academy: Lucian Sega
Southern Illinois University: Vicki Carstens, Jeffrey Punske
SpringLight Education Institute: Sherry Wang
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, Ryan Musa
University of Maryland: Julia Buffinton, Kasia Hitzkenko
University of Memphis: Vasile Rus
University of Michigan: Steven Abney, Marcus Berger, Sally Thomason
University of Nebraska, Omaha: Ashwathy Ashokan
University of North Carolina, Charlotte: Hossein Hemitialam, Wlodek Zadrozny
University of North Texas: Genene Murphy, Rodney Nielsen
University of Pennsylvania: Chris Callison-Burch, Cheryl Hickey, Mitch Marcus
University of Southern California: Aliya Deri, Ashish Vaswani
University of Texas at Dallas: Luis Gerardo Mojica de la Vega, Jing Lu, Vincent Ng
University of Texas, Austin: Rainer Mueller
University of Utah: Aniko Czirmaz, Mengqi Wang, Andrew Zupon
University of Washington: Jim Hoard, Joyce Parvi
University of Wisconsin, Eau Claire: Lynsey Wolter
University of Wisconsin, Madison: Steve Lacy
University of Wisconsin, Milwaukee: Joyce Boyland, Hanyon Park, Gabriella Pinter
Washington University in Saint Louis: Harry Go, Brett Hyde, Jackie Nelligan
Western Michigan University: Shannon Houtrouw, John Kapenga
Western Washington University: Kristin Denham
Yale University: Bob Frank, Raffaella Zanuttini, and the Yale Undergraduate Linguistics Society
NACLO 2016 Organizers (cont’d)

Canada Contest Site Coordinators:
Dalhousie University: Magdalena Jankowska, Vlado Keselj, Dijana Kosmajac, 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, Jozina Vander Klok
University of Calgary: Dennis Storoshenko
University of Lethbridge: Yllias Chali
University of Ottawa: Diana Inkpen
University of Toronto: Jordan Ho, James Hyett, Pen Long
University of Western Ontario: Janis Chang

High school sites: Dragomir Radev

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

Sponsorship Chair:
James Pustejovsky, Brandeis University

Sustaining Donors
Linguistic Society of America
NAACL
NSF
Yahoo!

Major Donors
ACM SIGIR
ARL
Choosito!
DARPA
Linguistic Data Consortium (LDC)

University Donors
Brandeis University
Carnegie Mellon University
University of Maryland
University of Michigan
University of Washington

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 120+ High School Sites

All material in this booklet © 2016, North American Computational Linguistics Olympiad and the authors of the individual problems. Please do not copy or distribute without permission.
As well as more than 120 high schools throughout the USA and Canada
(I) Deriving Enjoyment (1/2) [10 points]

Slovenian is a South Slavic language spoken by approximately 2.5 million speakers worldwide, the majority of whom live in Slovenia.

Approximate pronunciation guide: this is for your information only and does not contribute to the solution. č, š, ž are pronounced like ch in ‘check’, sh in ‘sheet’, and the s in ‘measure’, j is pronounced like y in ‘yes’, c is pronounced like zz in ‘pizza’, h is pronounced like ch in ‘loch’, v is pronounced somewhat like a w.

Answer the following questions in the Answer Sheets.

**I1.** Study the following data which shows some word derivations. Fill in the gaps.

<table>
<thead>
<tr>
<th>Adam</th>
<th>Adamič</th>
<th>Adams</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>baba</td>
<td>babica</td>
<td>babica</td>
<td></td>
</tr>
<tr>
<td>boben</td>
<td>bobnič</td>
<td>small drum, eardrum</td>
<td></td>
</tr>
<tr>
<td>bog</td>
<td>god</td>
<td>small drum, eardrum</td>
<td>(b)</td>
</tr>
<tr>
<td>čokolada</td>
<td>čokoladica</td>
<td>small chocolate</td>
<td></td>
</tr>
<tr>
<td>boben</td>
<td>bobnič</td>
<td>small drum, eardrum</td>
<td></td>
</tr>
<tr>
<td>bog</td>
<td>god</td>
<td>small god</td>
<td>(b)</td>
</tr>
<tr>
<td>jamica</td>
<td>hole</td>
<td>hole</td>
<td></td>
</tr>
<tr>
<td>knjiga</td>
<td>booklet</td>
<td>booklet</td>
<td>(d)</td>
</tr>
<tr>
<td>koklja</td>
<td>kokljica</td>
<td>chicken</td>
<td></td>
</tr>
<tr>
<td>menih</td>
<td>menišč</td>
<td>young monk</td>
<td></td>
</tr>
<tr>
<td>muha</td>
<td>fly</td>
<td>fly</td>
<td>(e) midge (small fly)</td>
</tr>
<tr>
<td>nožica</td>
<td>small leg</td>
<td>small leg</td>
<td></td>
</tr>
<tr>
<td>ogenj</td>
<td>ognjič</td>
<td>small fire</td>
<td></td>
</tr>
<tr>
<td>orel</td>
<td>orlica</td>
<td>female eagle</td>
<td>(f)</td>
</tr>
<tr>
<td>osel</td>
<td>oslič</td>
<td>donkey foal</td>
<td>(g)</td>
</tr>
<tr>
<td>otrok</td>
<td>baby</td>
<td>baby</td>
<td>(h)</td>
</tr>
<tr>
<td>oven</td>
<td>sheep</td>
<td>sheep</td>
<td>(i)</td>
</tr>
</tbody>
</table>

---

**n a c l o**
(I) Deriving Enjoyment (2/2)

<table>
<thead>
<tr>
<th>Pavel</th>
<th>Paul</th>
<th>(j)</th>
<th>Paulson</th>
</tr>
</thead>
<tbody>
<tr>
<td>Peter</td>
<td>Peter</td>
<td>Petrič</td>
<td>Peterson</td>
</tr>
<tr>
<td>pob</td>
<td>boy</td>
<td>pobič</td>
<td>small boy</td>
</tr>
<tr>
<td>Primož</td>
<td>Primus</td>
<td>Primožič</td>
<td>Primusson</td>
</tr>
<tr>
<td>(k) crab</td>
<td>račič</td>
<td>baby crab</td>
<td></td>
</tr>
<tr>
<td>roka</td>
<td>arm</td>
<td>ročica</td>
<td>small arm</td>
</tr>
<tr>
<td>(l) Stephen</td>
<td>Štefanič</td>
<td>Stephenson</td>
<td></td>
</tr>
<tr>
<td>Šapa</td>
<td>paw</td>
<td>šapica</td>
<td>small paw</td>
</tr>
<tr>
<td>Tomaž</td>
<td>Thomas</td>
<td></td>
<td>Thomson</td>
</tr>
<tr>
<td>(n) thorn</td>
<td>trnič</td>
<td></td>
<td>small thorn</td>
</tr>
<tr>
<td>Urh</td>
<td>Ulrik</td>
<td>Uršič</td>
<td>Ulrikson</td>
</tr>
<tr>
<td>veter</td>
<td>wind</td>
<td></td>
<td>draft (current of air)</td>
</tr>
<tr>
<td>volk</td>
<td>wolf</td>
<td>volčič</td>
<td>wolf cub</td>
</tr>
<tr>
<td>vrh</td>
<td>peak</td>
<td></td>
<td>small peak</td>
</tr>
<tr>
<td>zid</td>
<td>wall</td>
<td></td>
<td>small wall</td>
</tr>
<tr>
<td>žep</td>
<td>pocket</td>
<td></td>
<td>small pocket</td>
</tr>
</tbody>
</table>

12. If rožič means ‘small horn’, give the two possible words for ‘horn’ from which it might be derived.

13. If čolnič means ‘small boat’, give the two possible words for ‘boat’ from which it might be derived.
(J) Get Edumacated! (1/2) [15 points]

“Homeric infixation is a morphological construction that has recently gained currency in Vernacular American English. People who are familiar with this construction invariably credit the TV animation series, *The Simpsons*, particularly the speech of the main character Homer Simpson, for popularizing this construction.”

(Yu, A.C.L. 2004. Reduplication in English Homeric infixation. NELS 34)

Many speakers of American English, particularly younger generations, can insert the syllable “ma” into a word (like “edumacation” or “saxomaphone”) to produce a humorous variant. For many words, everyone agrees on how the “edumacated” variant should be formed, but there’s a lot of disagreement, too.

Below, three people give what they feel are the correct “edumacated” versions of twelve words. We’ve capitalized the stressed syllables of the respondent’s answers. Stressed syllables are spoken with more emphasis than unstressed syllables; for example, the second syllable in poTAto is stressed.

**J1.** We’ve left out some of their responses. Fill in the blanks with the appropriate words from the list below. You should likewise indicate stress with capitalization in your answers. Write your answers in the Answer Sheets.

<table>
<thead>
<tr>
<th>Alan</th>
<th>Barbara</th>
<th>Chris</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alabama</td>
<td>AlamaBAma</td>
<td>AlamaBAma</td>
</tr>
<tr>
<td>capital</td>
<td>CApimaTAL</td>
<td>CApimaTAL</td>
</tr>
<tr>
<td>captain</td>
<td>CApamaTAIN</td>
<td>CAPtamaTAIN</td>
</tr>
<tr>
<td>congratulations</td>
<td>conGRAtumaLAtions</td>
<td>conGRAtumaLAtions</td>
</tr>
<tr>
<td>hypothermia</td>
<td>HYpomaTHERmia</td>
<td>HYpomaTHERmia</td>
</tr>
<tr>
<td>oboe</td>
<td>ObamaBOE</td>
<td>OboemaBOE</td>
</tr>
<tr>
<td>octagon</td>
<td>OCTamaGON</td>
<td>OCTamaGON</td>
</tr>
<tr>
<td>octet</td>
<td>(a)</td>
<td>(b)</td>
</tr>
<tr>
<td>purple</td>
<td>(c)</td>
<td>(d)</td>
</tr>
<tr>
<td>tuba</td>
<td>(f)</td>
<td>TUBamaBA</td>
</tr>
<tr>
<td>wonder</td>
<td>WONdamaDER</td>
<td>WONdermaDER</td>
</tr>
<tr>
<td>wonderful</td>
<td>WONdermaFUL</td>
<td>WONdermaFUL</td>
</tr>
</tbody>
</table>

(A) PURpamaPLE
(B) OCtempaTET
(C) TUBamaBA
(D) TUUUmamaBA
(E) PURplemaPLE
(F) OcamaTET
(G) PURRRmaPLE
(J) Get Edumacated! (2/2)

J2. How would each respondent say the following words? We’ve given you a few to get started.

<table>
<thead>
<tr>
<th></th>
<th>Alan</th>
<th>Barbara</th>
<th>Chris</th>
</tr>
</thead>
<tbody>
<tr>
<td>antiseptic</td>
<td>(a)</td>
<td>(b)</td>
<td>(c)</td>
</tr>
<tr>
<td>Canada</td>
<td>(d)</td>
<td>(e)</td>
<td>(f)</td>
</tr>
<tr>
<td>feudalism</td>
<td>(g)</td>
<td>FEUdamalISM</td>
<td>(h)</td>
</tr>
<tr>
<td>optics</td>
<td>(i)</td>
<td>(j)</td>
<td>(k)</td>
</tr>
<tr>
<td>party</td>
<td>PARtamaty</td>
<td>(l)</td>
<td>(m)</td>
</tr>
<tr>
<td>table</td>
<td>(n)</td>
<td>(o)</td>
<td>(p)</td>
</tr>
<tr>
<td>water</td>
<td>(q)</td>
<td>(r)</td>
<td>WAAAmater</td>
</tr>
</tbody>
</table>

J3. Respondents usually hesitate before two-syllable words, and are less sure that their answers feel “correct”. Why, and what motivates Alan’s, Barbara’s, and Chris’s eventual answers?
(K) Kings, Queens, and Counts (1/5) [15 points]

Suppose your computer asks you, “What is the meaning of life?” Not being much of a philosopher, you decide to interpret the question as “What is the definition of the word life?” because that question is much more straightforward than “What is the purpose of existing?”

Even though you’ve simplified the problem, it’s still not very easy. Language users have so much background knowledge wrapped up in their mental definitions of words that it is tough to teach a computer all the shades of meaning encompassed by each word.

One of the most effective ways of defining words for computers is also one of the simplest: Get a sample of text and define each word by counting how often it occurs near each other word. Using this method, life might be defined as “the word that occurs 657 times near the, 423 times near a, 11 times near bug’s, 0 times near gumption, 0 times near ellipsis, 8 times near preserver, …” and the list would continue for every word in the sample of text. The following questions deal with this method of representing word meaning.

K1. For question K1, the following poem will be the sample text used to obtain word counts:

Whether the weather is good, or whether the weather is not,
Whether the weather is cold, or whether the weather is hot,
We’ll weather the weather—whatever the weather—
Whether we like it or not.

The representations of some words from this poem are shown below as obtained by counting how often each other word occurs in a certain window around the word in question. Your task: in the Answer Sheets, shade in the provided graph to give the representation of the word is.
Below are 33 word representation graphs. These were obtained from a different sample text and show the counts of 15 words (word A through word O), but the identities of these 15 words are not given. Your task: Study these 33 word graphs and then answer the questions that follow.
Antananarivo (Madagascar’s capital)

mystery word #1

mystery word #2

mystery word #3

mystery word #4

mystery word #5

mystery word #6

mystery word #7

mystery word #8

mystery word #9

mystery word #10

mystery word #11

n a c l o
(K) Kings, Queens, and Counts (5/5)

K2. The 11 mystery words have the following definitions (but not in this order):

a. antismartnessesquely
b. aunt
c. big
d. can
e. cats
f. Kenya
g. Kenyan
h. meow
i. strange
j. strangest
k. the

On your Answer Sheet, write the number of the mystery word corresponding to each definition.

K3. You might expect the graph for mystery word #4 to look something like the graph below, but it does not. Explain why.
(L) The Short Hand of the Law (1/2) [10 points]

Shorthand machines, commonly known as stenotype machines, are used in many courts of law to record court proceedings. They are a special kind of typewriter with an unusual keyboard, in which many keys can be punched simultaneously ("chorded"), and the results are output onto a thin strip of paper. Court stenographers using stenotype machines can transcribe court proceedings very quickly; the world record is 375 words per minute!

Below, we have taken an example of court stenography, 25 lines long, and divided it into five pieces.

(A)

```
O  P  B
S  T  K  P  W  H  R  F  R  P  B  L  G  T  S  H  O  U
T  K  O  U
P  H  R  A  O  E  D
```

(B)

```
S  T  K  P  W  H  R  F  R  P  B  L  G  T  S  R  U
R  E
T  K  A  O  E
T  O
```

(C)

```
T  H
T  A  O  E  U  P  L
T  K  F  R  P  B  L  G  T  S  K  W  R  E
U  R
```

Below, we have taken an example of court stenography, 25 lines long, and divided it into five pieces.
The original dialogue went like this:

THE COURT: Are you ready to enter a plea at this time?
THE DEFENDANT: Yes, your honor.
THE COURT: How do you plead to counts one and two?

Answer these questions in the Answer Sheets.

L1. Put the pieces in their original order.

L2. Below are the next nine lines of transcription. What do they say?

L3. Explain your answer.
The Tocharian languages were an extinct branch of the Indo-European language family (including English, French, German, Greek, and many other languages in Europe and Asia). Linguists have reconstructed the ancestor language, called Proto-Indo-European, from which all the descendant branches descended.

A major part of language change is sound change, where a language’s sounds shift around over time. Importantly, sound change is regular, and can be encapsulated neatly by writing down rules to describe how one stage of the language proceeds to the next. For example, a rule like t > d / _r means that all instances of ‘t’ change to ‘d’ before ‘r’, so tree would become dree, while: p > Ø / _# means that all instances of ‘p’ disappear (change to ‘zero’) at the end of a word (represented as the hash #), so stop would become sto. Sound changes apply to all sounds, in all words, that fit their criteria.

Because many ancient languages were never written down until recent millennia, linguists have to rely on clever deductions to work out the details of their early history. Our only records of Tocharian are some 9th century manuscripts around the Tarim Basin in western China, so our knowledge of its development comes from inferences of this type.

Answer these questions in the Answer Sheets.

**M1.** Here are some Tocharian words, meaning “share”, “row of teeth”, “knee”, “war”, “one hundred”, “dog”, and “prop” respectively. These groups of words represent seven stages in the very early history of the language, in a random order:

<table>
<thead>
<tr>
<th>stage</th>
<th>“share”</th>
<th>“row of teeth”</th>
<th>“knee”</th>
<th>“war”</th>
<th>“one hundred”</th>
<th>“dog”</th>
<th>“prop”</th>
</tr>
</thead>
<tbody>
<tr>
<td>(A)</td>
<td>pákos</td>
<td>kómos</td>
<td>kónu</td>
<td>kóro-</td>
<td>kṃtóm</td>
<td>kuō</td>
<td>stema-</td>
</tr>
<tr>
<td>(B)</td>
<td>págos</td>
<td>gómos</td>
<td>gónu</td>
<td>kóro-</td>
<td>kṃtóm</td>
<td>kuō</td>
<td>stema-</td>
</tr>
<tr>
<td>(C)</td>
<td>bʰágos</td>
<td>jómbʰos</td>
<td>jónu</td>
<td>kóro-</td>
<td>kṃtóm</td>
<td>kuō</td>
<td>stembʰa-</td>
</tr>
<tr>
<td>(D)</td>
<td>bʰágos</td>
<td>jómbʰos</td>
<td>jónu</td>
<td>kóro-</td>
<td>cm̥tóm</td>
<td>cuō</td>
<td>stembʰa-</td>
</tr>
<tr>
<td>(E)</td>
<td>páko</td>
<td>kómo</td>
<td>kónu</td>
<td>kóro-</td>
<td>kṃtóm</td>
<td>kuō</td>
<td>stema-</td>
</tr>
<tr>
<td>(F)</td>
<td>bʰágos</td>
<td>gómos</td>
<td>gónu</td>
<td>kóro-</td>
<td>kṃtóm</td>
<td>kuō</td>
<td>stema-</td>
</tr>
<tr>
<td>(G)</td>
<td>bʰágos</td>
<td>gómbʰos</td>
<td>gónu</td>
<td>kóro-</td>
<td>kṃtóm</td>
<td>kuō</td>
<td>stembʰa-</td>
</tr>
</tbody>
</table>

As we can see, between these stages of Tocharian, some sound changes have occurred. Put the stages in historical order, and write down rules describing the sound changes that happened in between each stage. If you can find different orders, explain which you think is the most likely. (The accent ‘ on a vowel can be ignored.)
### (M) Sound Judgments (2/2)

**M2.** Here in alphabetical order are some roots from a slightly later point in the early development of Tocharian, and their descendants later on in the history of the family:

<table>
<thead>
<tr>
<th>Early Tocharian</th>
<th>Later Tocharian</th>
</tr>
</thead>
<tbody>
<tr>
<td>“they drive”</td>
<td>*agonti</td>
</tr>
<tr>
<td>“they are driven”</td>
<td>*agontor</td>
</tr>
<tr>
<td>“ten”</td>
<td>*dékamt</td>
</tr>
<tr>
<td>“one hundred”</td>
<td>*kamtóm</td>
</tr>
<tr>
<td>“stag hunter”</td>
<td>*kēruwos</td>
</tr>
<tr>
<td>“father”</td>
<td>*patēr</td>
</tr>
<tr>
<td>“running” (later &gt; “river”)</td>
<td>*tékʷos</td>
</tr>
<tr>
<td>“that”</td>
<td>*tód</td>
</tr>
<tr>
<td>“twenty”</td>
<td>*wíkanti</td>
</tr>
</tbody>
</table>

Between these two stages of Tocharian, some further sound changes have occurred. Here they are in a random order. (Some changes apply to multiple sounds at once; a comma indicates this.)

(A) o > ë  
(B) nt > Ø / __ #  
(C) e > a  
(D) or > ur / __ #  
(E) ē, ī, ũ > e, i, u  
(F) m > n / __ #  
(G) t > c / __ e, ē  
(H) d > ś / __ e, ē  
(I) n > Ø / __ #  
(J) m > n / __ t  
(K) u > a  
(L) k > ś / __ e, ē  
(M) w > wʸ / __ i, ī  
(N) g > k  
(O) d > Ø / __ #  
(P) s > Ø / __ #

Using any diagrams or notation you wish, write down as much as you can deduce about the order in which these changes happened, and *explain* how you have reached your answer.
Hungarian is a Finno-Ugric language spoken in Hungary by about 10 million speakers and about 2.5 million speakers in the surrounding countries, as well as the diaspora. Hungarian is often called a nonconfigurational language, which means that a) the words are unambiguously marked for their role in the sentence and b) the word order is not rigid but often determined by the conversational context the sentences appear in.

Match the Hungarian sentences with their English translations. Write your answers in the Answer Sheets.

N1. Valaki megvert valakit.  
   (A) No one beat everyone (at e.g. chess).
N2. Kit vert meg valaki?  
   (B) Who wasn't beaten by anyone?
N3. Senki nem verte meg a Petyát.  
   (C) No one got beaten.
N4. Valakit senki nem vert meg.  
   (D) Someone beat Martin.
N5. Senki nem vert meg mindenkit.  
   (E) I didn't beat anyone.
   (F) No one beat Peter.
N7. Ki nem vert meg senkit?  
   (G) Who got beaten by someone?
N8. Valaki senkit nem vert meg.  
   (H) Someone beat someone.
   (I) Everyone beat someone.
N10. Valaki megverte a Marcit.  
    (J) Who didn't beat anyone?
    (K) There's someone who didn't beat anyone.
N12. Kit nem vert meg senki?  
    (L) Peter beat no one.
    (M) There is someone who didn't get beaten.

N14. Explain your answers.
(O) Don’t Sell the House! (1/1) [10 points]

This problem involves the Nung language of northeastern Vietnam, spoken by about a million people and related to the Thai, Lao, Isan, Shan, and Zhuang languages of Southeast Asia in the Tai-Kadai family. It is not related to Chinese, Vietnamese, Khmer, Hmong, Malay, or Burmese, so far as we know. In this problem, the Nùng Phạn Slinh variety of Nung will be used. In Nùng Phạn Slinh as seen here, word order is fixed: that is, for every sentence containing certain words, there is only one way to properly order those words.

Here is a list of sentences in Nung and their English translations. Find the sentences without English or Nung equivalents and write down the missing translation in the Answer Sheets. Note: The marks above vowels indicate tone and the length of the vowel. đ and sl are consonants. You do not need to know how to pronounce Nung in order to solve the problem.

<table>
<thead>
<tr>
<th>Nung</th>
<th>English</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cáu ca vửhn nhahng kihn.</td>
<td>I was about to continue to eat it.</td>
</tr>
<tr>
<td>Cáu cháhn slờng páy mi?</td>
<td>Do I truly want to go?</td>
</tr>
<tr>
<td>Cáu mi slày kihn.</td>
<td>I don’t have to eat it.</td>
</tr>
<tr>
<td>Cáu ngám héht pehn té.</td>
<td>I did it like that just now.</td>
</tr>
<tr>
<td>Cáu tan dohc hánh mưhng.</td>
<td>I only saw you.</td>
</tr>
<tr>
<td>Cáu vừhn nhahng bô sạhm tåhng héht hòn.</td>
<td>I also continue to build the house alone.</td>
</tr>
<tr>
<td>Da kihn!</td>
<td>Don’t eat it!</td>
</tr>
<tr>
<td>Da khảo hòn!</td>
<td>Don’t sell the house!</td>
</tr>
<tr>
<td>Mưhn chóng ca cháhn fải khảo.</td>
<td>Then she truly was about to have to sell it.</td>
</tr>
<tr>
<td>Mưhn mi cháhn dáy non.</td>
<td>She truly can’t sleep.</td>
</tr>
<tr>
<td>Mưhn náhc-thày chóng bô sạhm kihn.</td>
<td>Then she also just previously ate it.</td>
</tr>
<tr>
<td>Mưhng náhc-thày slông tåhng páy.</td>
<td>You wanted to go alone just previously.</td>
</tr>
</tbody>
</table>

01. Cáu cháhn dây non.
02. Da páy non!
03. Mưhn bó sạhm mi slông héht hòn mi?
04. Mưhn ngám bó sạhm páy hòn.
05. I wasn’t about to eat it just previously.
06. She didn’t have to eat it alone like that just now.
07. The house truly can’t eat you.
08. Then were you also about to go just previously?
A Horn clause, named for logician Alfred Horn, is a notation used in mathematics and in logic programs such as Prolog. Horn clauses offer a flexible way to write the rules of grammar for a language. This problem will introduce you to Horn clause notation and ask you to use the notation to describe English and Swiss German.

Let's start with English, since you already know it. To capture a simple fragment of English, we might say that a sentence consists of a noun followed by a verb. If we write $S$ to mean sentence, $N$ to mean noun, and $V$ to mean verb, the following Horn clause captures this intuition:

$$S(xy) : - N(x), V(y).$$

This rule says that if we have a noun $x$ and a verb $y$, we can make a sentence by putting $x$ and $y$ together in that order. Horn clauses with the ‘:-’ symbol are called rules, and they tell us how to derive the thing on the left side of the ‘:-’ from the things on the right side of the ‘:-’. Note that the labels $S$, $N$, and $V$ don't affect how the rule behaves; they are simply chosen to help us remember what we're representing.

However, so far we haven't given ourselves any nouns or verbs, so we can't make a sentence. The following Horn clauses give us nouns and verbs to work with:

$$N(Mary). \quad N(John). \quad V(eats). \quad V(sleeps).$$

For example, the first clause says that “Mary” is a noun. Horn clauses without the ‘:-’ symbol are called facts, because they tell us things that we know are true without doing any work.

Using our facts and our lone rule, we can derive the following sentences:

$$S(Mary eats). \quad S(Mary sleeps). \quad S(John sleeps). \quad S(John eats).$$

We can extend our grammar to account for subject-verb agreement in English. $sg$ means singular and $pl$ means plural.

$$S(xy) : - Nsg(x), Vsg(y). \quad S(xy) : - Npl(x), Vpl(Y).$$

$$Nsg(Mary). \quad Npl(dogs). \quad Vsg(sleeps). \quad Vpl(sleep).$$

Note that we can derive the sentences $S(Mary sleeps)$ and $S(dogs sleep)$, but because we have no way to put an $Nsg$ together with a $Vpl$, we can't derive $S(Mary sleep)$.

Answer these questions in the Answer Sheets.

P1. The rules above can only generate a fixed, finite number of sentences, but there is no clear upper limit on the length of grammatical English sentences. For example, consider the following sentences:

We helped Mary help John paint the house.
We helped Mary help John help Kim paint the house.
We helped Mary help John help Kim help John paint the house.
We let Mary let John let Kim paint the house.
We let Mary help John let Kim paint the house.
We let Mary help John help Kim let Mary help John let Mary paint the house.
Clearly we can keep extending these sentences as long as we want; they will still be grammatical, even if they are a bit awkward.

To make things easier for you, we only want you to account for the underlined parts of the sentences. It's easy but tedious to extend the grammar to account for the entire sentences. Write a set of rules and facts that will generate all the possible combinations of “help”, “let”, “John”, and “Kim” that will fit in the sentences above. For example, you should be able to derive S(help John let Kim let John).

P2. Let's look at similar sentences in Swiss German:

Jan säit das mer em Hans em Jan es huus halfed halfe aastriiche.
Jan says that we helped Hans help Jan paint the house.

Jan säit das mer em Hans em Jan em Hans es huus halfed halfe halfe aastriiche.
Jan says that we helped Hans help Jan help Hans paint the house.

Jan säit das mer em Hans em Jan em Hans em Jan es huus halfed halfe halfe halfe aastriiche.
Jan says that we helped Hans help Jan help Hans help Jan paint the house.

Jan säit das mer de Hans de Jan es huus laa halfe aastriiche.
Jan says that we let Hans let Jan paint the house.

Jan säit das mer de Hans em Jan de Hans es huus laa halfe halfe aastriiche.
Jan says that we let Hans help Jan let Hans paint the house.

Jan säit das mer de Hans em Jan em Hans de Jan em Hans de Jan es huus laa halfe halfe laa halfe laa aastriiche.
Jan says that we let Hans help Jan help Hans let Jan help Hans let Jan paint the house.

It turns out that the formalism described above cannot generate the Swiss German data. However, a simple extension can. Instead of manipulating a single phrase or sentence, we allow ourselves to manipulate a pair of phrases or sentences:

\[ R(wy, xz) : T(w, x), T(y, z). \]

This says that if the pair \((w, x)\) is a \(T\) (whatever that may be), and the pair \((y, z)\) is also a \(T\), then the pair \((wy, xz)\) is an \(R\) (whatever that may be). At the end of the day, we can smash the pair into a single sentence:

\[ S(xy) : R(x, y). \]

For example, suppose we add the fact \(T(\text{the}, \text{cat})\). Then the first rule lets us derive \(R(\text{the, the, cat cat})\), and the second rule lets us derive \(S(\text{the, the cat cat})\).

Use this extension to describe the Swiss German data. Again, to make your life easier, you only need to generate the underlined part of the sentences. For example, you should be able to derive \(S(\text{de Hans em Jan em Hans laa halfe halfe})\).
Javanese is an Austronesian language spoken by nearly 100 million people in Indonesia and worldwide. Answer these questions about its script in the Answer Sheets.

**Q1.** Here are some Javanese words in the Javanese script, Latin script, and their meanings. ‘ny’ and ‘ng’ are consonants. ‘é’ is a vowel. Write the missing words in Latin script.

<table>
<thead>
<tr>
<th>Javanese Script</th>
<th>Latin Script</th>
<th>Meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>penyakit</td>
<td>penyaKIt</td>
<td>disease</td>
</tr>
<tr>
<td>Inggris</td>
<td>InggrIs</td>
<td>England</td>
</tr>
<tr>
<td>traktor</td>
<td>traktor</td>
<td>tractor</td>
</tr>
<tr>
<td>panyumbang</td>
<td>panyumbang</td>
<td>donor</td>
</tr>
<tr>
<td>rembulan</td>
<td>rembulan</td>
<td>moon</td>
</tr>
<tr>
<td>tansah</td>
<td>tansah</td>
<td>always</td>
</tr>
<tr>
<td>Amérika</td>
<td>Amérika</td>
<td>America (continent)</td>
</tr>
<tr>
<td>Javanese</td>
<td>English</td>
<td></td>
</tr>
<tr>
<td>----------</td>
<td>---------</td>
<td></td>
</tr>
<tr>
<td>ngrebut</td>
<td>to grab</td>
<td></td>
</tr>
<tr>
<td>ibukota</td>
<td>capital</td>
<td></td>
</tr>
<tr>
<td>Argentina</td>
<td>Argentina</td>
<td></td>
</tr>
<tr>
<td>srengéngé</td>
<td>sun</td>
<td></td>
</tr>
<tr>
<td>palsu</td>
<td>false</td>
<td></td>
</tr>
<tr>
<td>rerenggan</td>
<td>decoration</td>
<td></td>
</tr>
<tr>
<td>angsal</td>
<td>to acquire</td>
<td></td>
</tr>
<tr>
<td>inggih</td>
<td>yes</td>
<td></td>
</tr>
<tr>
<td>(a)</td>
<td>often</td>
<td></td>
</tr>
</tbody>
</table>
**Q2.** Write these words in the Javanese script

<table>
<thead>
<tr>
<th>Latin Script</th>
<th>Meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>a. nyolong</td>
<td>to steal</td>
</tr>
<tr>
<td>b. sepalih</td>
<td>half</td>
</tr>
<tr>
<td>c. trengginas</td>
<td>lively</td>
</tr>
<tr>
<td>d. Antartika</td>
<td>Antarctica</td>
</tr>
<tr>
<td>e. Istanbul</td>
<td>Istanbul</td>
</tr>
</tbody>
</table>

**Q3.** Explain your answers.
Somali is a Cushitic language spoken by approximately 16.6 million speakers, of which about half live in Somalia, the remainder living in Djibouti (where it is an official language), Ethiopia, and in the Somali diaspora.

**R1.** In the table below are given the inflected forms of 1st conjugation verbs in the 1st person ('I') and 3rd person singular ('he/she/it') past tense. In the Answer Sheets, fill in the missing words.

Pronunciation: Vowel sounds are much like in English. A double vowel indicates that the vowel is long. Consonants are also as in English except as follows:

- dh: a retroflex ‘d’ pronounced with the tongue tip curled backward like the ‘dr’ in *drive*
- q: a voiced uvular plosive, like a ‘g’ but pronounced at the back of the throat like the sound of drinking water
- kh: a bit like the ‘ch’ in Scottish *loch* but pronounced at the back of the throat
- x: a voiceless pharyngeal fricative, like an ‘h’ pronounced deep in the throat
- c: same as ‘x’, but voiced
- r: a rolled ‘r’ as in Italian
- ‘’: the consonant sound in the middle of *uh-oh*

<table>
<thead>
<tr>
<th>Somali word</th>
<th>English translation</th>
<th>Somali word</th>
<th>English translation</th>
</tr>
</thead>
<tbody>
<tr>
<td>akhriyay</td>
<td>I read</td>
<td>akhridaday</td>
<td>He read</td>
</tr>
<tr>
<td>aragay</td>
<td>I saw</td>
<td>aragtay</td>
<td>He saw</td>
</tr>
<tr>
<td>(a)</td>
<td>I taught</td>
<td>bartay</td>
<td>He taught</td>
</tr>
<tr>
<td>ba’ay</td>
<td>I was destroyed</td>
<td>ba’day</td>
<td>He was destroyed</td>
</tr>
<tr>
<td>baaggiyay</td>
<td>I prevented</td>
<td>(b)</td>
<td>He prevented</td>
</tr>
<tr>
<td>baaqay</td>
<td>I announced</td>
<td>baaqday</td>
<td>He announced</td>
</tr>
<tr>
<td>baxay</td>
<td>I left</td>
<td>baxday</td>
<td>He left</td>
</tr>
<tr>
<td>bi’iyyay</td>
<td>I destroyed</td>
<td>(c)</td>
<td>He destroyed</td>
</tr>
<tr>
<td>bilaabay</td>
<td>I began</td>
<td>(d)</td>
<td>He began</td>
</tr>
<tr>
<td>(e)</td>
<td>I ate</td>
<td>cuntay</td>
<td>He ate</td>
</tr>
<tr>
<td>cabay</td>
<td>I drank</td>
<td>cabtay</td>
<td>He drank</td>
</tr>
<tr>
<td>cararay</td>
<td>I ran away</td>
<td>carartay</td>
<td>He ran away</td>
</tr>
</tbody>
</table>
### (R) Changing the Subject (2/2)

<table>
<thead>
<tr>
<th>Somali word</th>
<th>English translation</th>
</tr>
</thead>
<tbody>
<tr>
<td>daaqay</td>
<td>I grazed</td>
</tr>
<tr>
<td>dhacay</td>
<td>I fell</td>
</tr>
<tr>
<td>dhisay</td>
<td>I built</td>
</tr>
<tr>
<td>diiday</td>
<td>I refused</td>
</tr>
<tr>
<td>dilay</td>
<td>I killed</td>
</tr>
<tr>
<td>faraxay</td>
<td>I was happy</td>
</tr>
<tr>
<td>gaadhay</td>
<td>I reached</td>
</tr>
<tr>
<td>galay</td>
<td>I entered</td>
</tr>
<tr>
<td>go‘ay</td>
<td>I cut</td>
</tr>
<tr>
<td>(k)</td>
<td>I found</td>
</tr>
<tr>
<td>horjeeday</td>
<td>I opposed</td>
</tr>
<tr>
<td>kacay</td>
<td>I rose</td>
</tr>
<tr>
<td>keenay</td>
<td>I brought</td>
</tr>
<tr>
<td>korodhay</td>
<td>I increased</td>
</tr>
<tr>
<td>qaaday</td>
<td>I took</td>
</tr>
<tr>
<td>tagay</td>
<td>I went</td>
</tr>
<tr>
<td>xidhay</td>
<td>I closed</td>
</tr>
<tr>
<td>walaaqay</td>
<td>I stirred</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Somali word</th>
<th>English translation</th>
</tr>
</thead>
<tbody>
<tr>
<td>(f)</td>
<td>He grazed</td>
</tr>
<tr>
<td>(g)</td>
<td>He fell</td>
</tr>
<tr>
<td>dhistay</td>
<td>He built</td>
</tr>
<tr>
<td>diiday</td>
<td>He refused</td>
</tr>
<tr>
<td>dishay</td>
<td>He killed</td>
</tr>
<tr>
<td>(h)</td>
<td>He was happy</td>
</tr>
<tr>
<td>gaadhay</td>
<td>He reached</td>
</tr>
<tr>
<td>(i)</td>
<td>He entered</td>
</tr>
<tr>
<td>(j)</td>
<td>He cut</td>
</tr>
<tr>
<td>heshay</td>
<td>He found</td>
</tr>
<tr>
<td>horjeeday</td>
<td>He opposed</td>
</tr>
<tr>
<td>(l)</td>
<td>He rose</td>
</tr>
<tr>
<td>keentay</td>
<td>He brought</td>
</tr>
<tr>
<td>korodhay</td>
<td>He increased</td>
</tr>
<tr>
<td>(m)</td>
<td>He took</td>
</tr>
<tr>
<td>tagtay</td>
<td>He went</td>
</tr>
<tr>
<td>(n)</td>
<td>He closed</td>
</tr>
<tr>
<td>(o)</td>
<td>He stirred</td>
</tr>
</tbody>
</table>
Case Booklet

REGISTRATION NUMBER

Name: ______________________________________________________

Contest Site: _______________________________________________

Site ID: ___________________________________________________

City, State: ________________________________________________

Grade: _____

Start Time: ________________________________________________

End Time: _________________________________________________
(I) Deriving Enjoyment

1. a. [ ] b. [ ] c. [ ] d. [ ]
   e. [ ] f. [ ] g. [ ] h. [ ]
   i. [ ] j. [ ] k. [ ] l. [ ]
   m. [ ] n. [ ] o. [ ] p. [ ]
   q. [ ] r. [ ]

2. [ ]
3. [ ]

(I) Get Edumacated!

1. a. [ ] b. [ ] c. [ ] d. [ ] e. [ ] f. [ ] g. [ ]

2. a. [ ] b. [ ] c. [ ]
   d. [ ] e. [ ] f. [ ]
   g. [ ] h. [ ]
   i. [ ] j. [ ] k. [ ]
   l. [ ]
   m. [ ] n. [ ]
   p. [ ] q. [ ]

\[\text{Image} n \rightarrow a \rightarrow c \rightarrow l \rightarrow o\]
(J) Get Edumacated! (cont.)

3.

(K) Kings, Queens, and Counts

1.

2. a. [ ] b. [ ] c. [ ] d. [ ] e. [ ] f. [ ] g. [ ] h. [ ] i. [ ] j. [ ] k. [ ]
(K) Kings, Queens, and Counts (cont.)

3.

(L) The Short Hand of the Law

1. First □ □ □ □ □ Last

2.

3.
(M) Sound Judgments

1. Earliest
   
   Latest

2.
(N) What happened at the chess tournament?

1. 
2. 
3. 
4. 
5. 
6. 
7. 
8. 
9. 
10. 
11. 
12. 
13. 
14. 

(O) Don’t Sell the House!

1. 
2. 
3. 
4. 
5. 
6. 
7. 
8. 

n a c l o
(P) A Matter of Horn Clauses

1.

(Q) A Cup of Javanese

1. a. [Blank]  b. [Blank]
   c. [Blank]  d. [Blank]
   e. [Blank]
(Q) A Cup of Javanese (cont.)

2. a.

b.

c.

d.

e.
(Q) A Cup of Javanese (cont.)

3.

(R) Changing the Subject

1. a. __________________________  b. __________________________  c. __________________________
   d. __________________________  e. __________________________  f. __________________________
   g. __________________________  h. __________________________  i. __________________________
   j. __________________________  k. __________________________  l. __________________________
   m. __________________________  n. __________________________  o. __________________________