

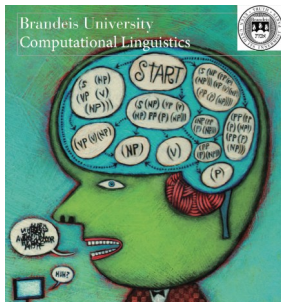
YAHOO!



ARL



Carnegie Mellon University
Language Technologies Institute



UNIVERSITY of
WASHINGTON



SIGIR
Special Interest Group
on Information Retrieval

Chōōsito!
search and learn



UNIVERSITY OF
MARYLAND

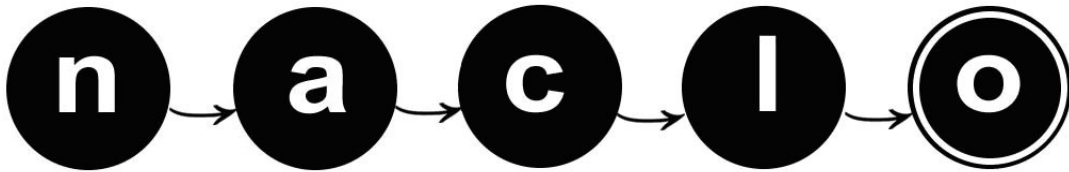
***The Tenth
Annual
North American
Computational
Linguistics
Olympiad
2016***
www.nacloweb.org

**Invitational
Round**

March 10, 2016

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:

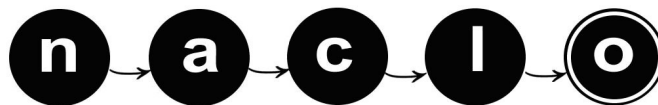
Jessica Sawyer	#850
----------------	------
8. The top students from each country (USA and Canada) will be invited to the next round, which involves online practices before the international competition in India.
9. Each problem has been thoroughly checked by linguists and computer scientists as well 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 Lovsted, 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)
Elysia 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

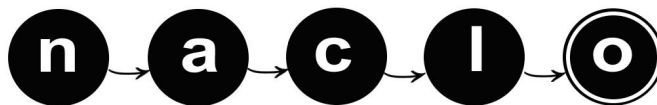
Stella Lau, University of Cambridge

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: Tamar Solorio
University of Illinois at Urbana-Champaign: Julia Hockenmaier, Ryan Musa
University of Maryland: Julia Buffinton, Kasia Hitczenko
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 Hematiam, Wlodek Zadrozny
University of North Texas: Genevieve 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.



NACLO 2016

Sites



As well as more than 120 high schools throughout the USA and Canada

(I) Deriving Enjoyment (1/2) [10 points]

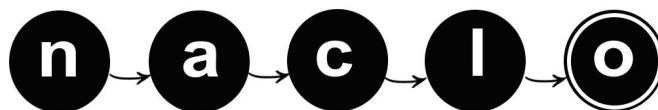
Slovenian is a South Slavic language spoken by approximately 2.5 million speakers worldwide, the majority of whom live in Slovenia.

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.

Adam	Adam	Adamič	Adams
baba	woman	babica	grandmother, little old lady
(a)	buffalo	bivolica	female buffalo
boben	drum	bobnič	small drum, eardrum
bog	god	(b)	small god
čokolada	chocolate	čokoladica	small chocolate
dekla	maid	deklica	young girl
Gregor	Gregory	Gregorič	Gregson
grm	bush	(c)	small bush
jama	cave	jamica	hole
knjiga	book	(d)	booklet
koklja	hen	kokljica	chicken
menih	monk	menišič	young monk
muha	fly	(e)	midge (small fly)
noga	leg	nožica	small leg
ogenj	fire	ognjič	small fire
orel	eagle	orlica	female eagle
		(f)	eaglet
osel	donkey	oslič	donkey foal
		(g)	jenny (i.e. she-donkey)
otrok	child	(h)	baby
oven	sheep	(i)	lamb



(I) Deriving Enjoyment (2/2)

Pavel	Paul	(j)	Paulson
Peter	Peter	Petrič	Peterson
pob	boy	pobič	small boy
Primož	Primus	Primožič	Primusson
(k)	crab	račič	baby crab
roka	arm	ročica	small arm
(l)	Stephen	Štefanič	Stephenson
šapa	paw	šapica	small paw
Tomaž	Thomas	(m)	Thomson
(n)	thorn	trnič	small thorn
Urh	Ulrik	Uršič	Ulrikson
veter	wind	(o)	draft (current of air)
volk	wolf	volčič	wolf cub
vrh	peak	(p)	small peak
zid	wall	(q)	small wall
žep	pocket	(r)	small pocket

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*)

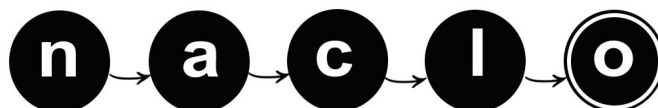
Many speakers of American English, particularly younger generations, can insert the syllable “ma” into a word (like “edumacation” or “saxomophone”) to produce a humorous variant. For many words, everyone agrees on how the “edumacated” variant should be formed, but there’s a lot of disagreement, too.

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.

	Alan	Barbara	Chris
Alabama	AlamaBAma	AlamaBAma	AlamaBAma
capital	CApimaTAL	CApimaTAL	CApimaTAL
captain	CApamaTAIN	CAPtamaTAIN	Uh... I’m not sure.
congratulations	conGRAtumaLAtions	conGRAtumaLAtions	conGRAtumaLAtions
hypothermia	HYpomaTHERmia	HYpomaTHERmia	HYpomaTHERmia
oboe	ObamaBOE	OboemaBOE	OOOmaBOE
octagon	OCTamaGON	OCTamaGON	OCTamaGON
octet	(a)	(b)	I dunno...
purple	(c)	(d)	(e)
tuba	(f)	TUbamaBA	(g)
wonder	WONdamaDER	WONdermaDER	WONNNmaDER?
wonderful	WONdermaFUL	WONdermaFUL	WONdermaFUL

- (A) PURpamaPLE
- (B) OCTemaTET
- (C) TUbamaBA
- (D) TUUUmaBA
- (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.

	Alan	Barbara	Chris
antiseptic	(a)	(b)	(c)
Canada	(d)	(e)	(f)
feudalism	(g)	FEUdamaLISm	(h)
optics	(i)	(j)	(k)
party	PARtamaTY	(l)	(m)
table	(n)	(o)	(p)
water	(q)	(r)	WAAAmATER

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?”

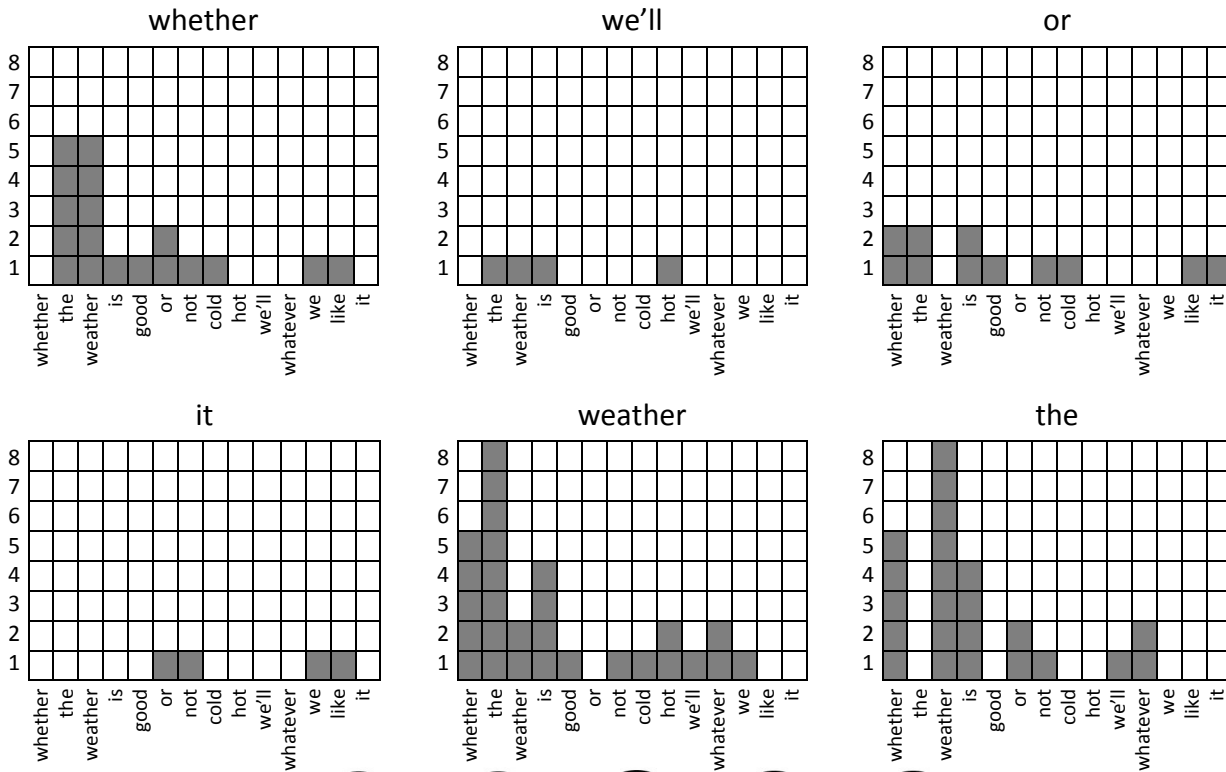
Even though you’ve simplified the problem, it’s still not very easy. Language users have so much background knowledge wrapped up in their mental definitions of words that it is tough to teach a computer all the shades of meaning encompassed by each word.

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:

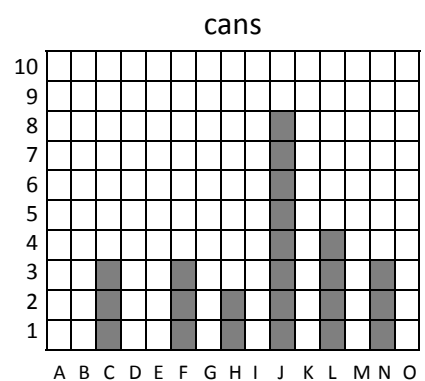
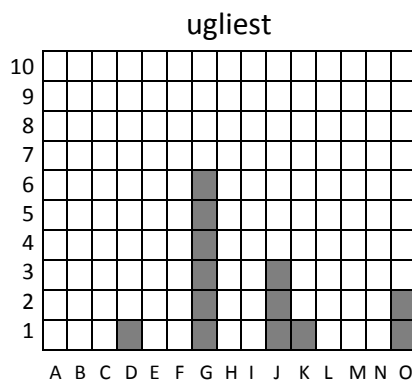
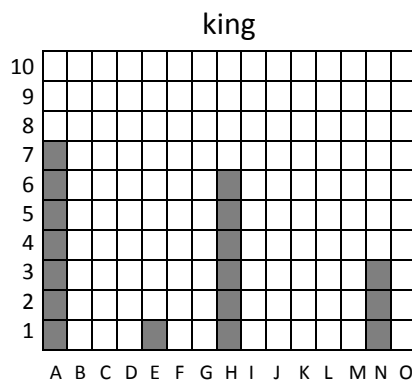
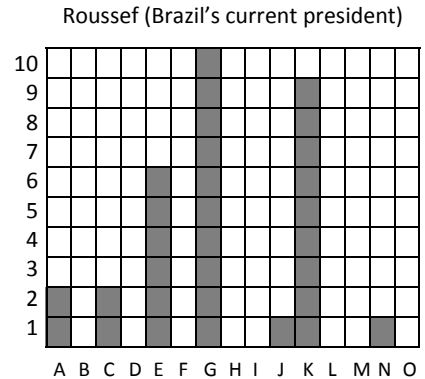
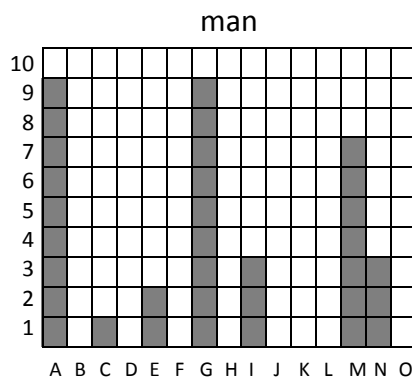
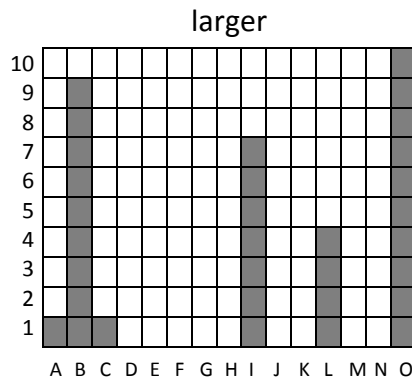
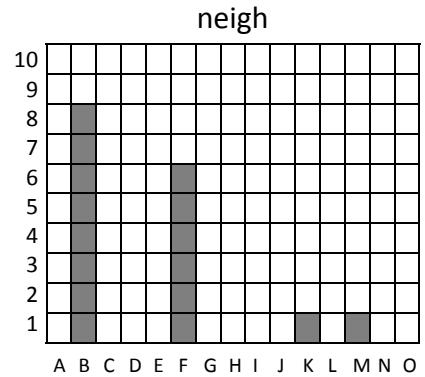
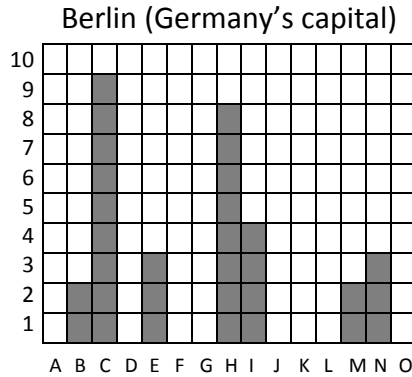
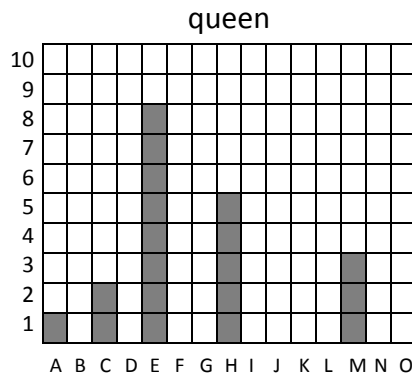
*Whether the weather is good, or whether the weather is not,
Whether the weather is cold, or whether the weather is hot,
We’ll weather the weather—whatever the weather—
Whether we like it or not.*

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*.

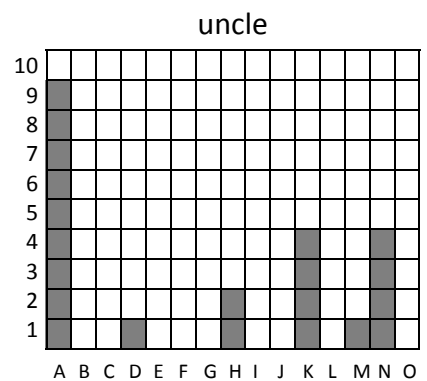
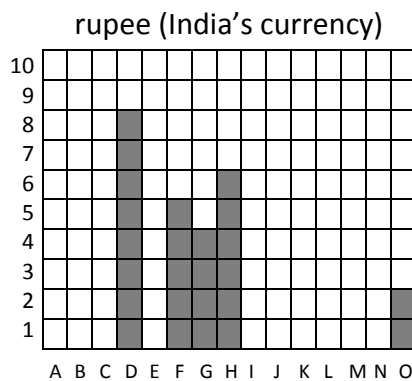
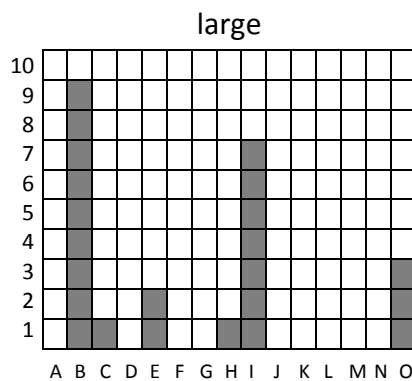
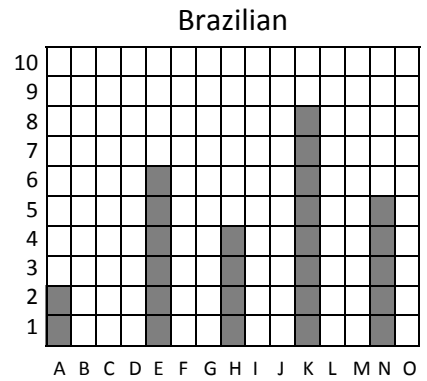
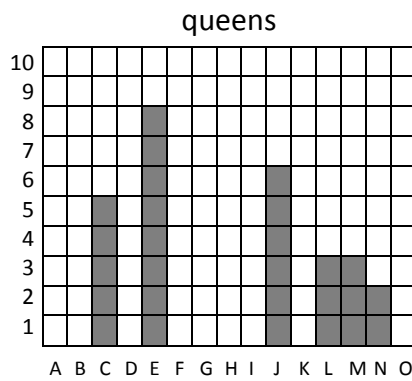
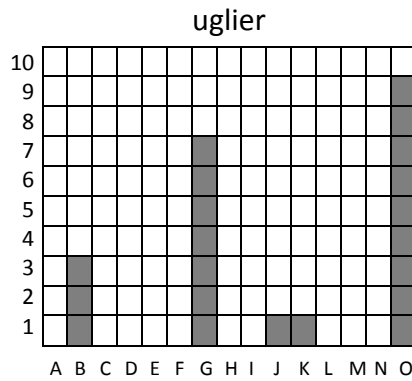
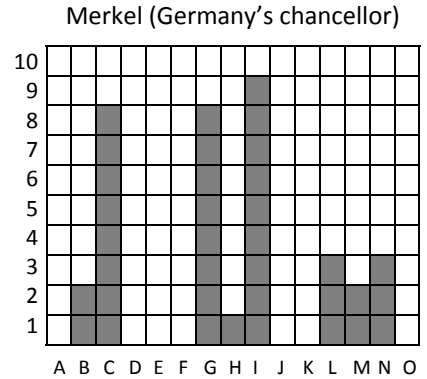
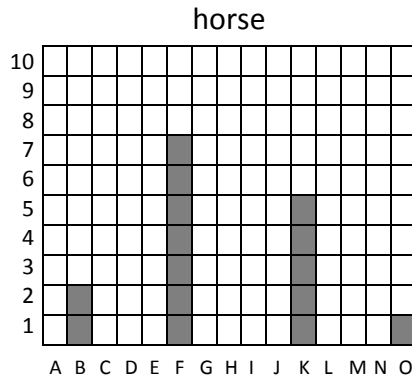
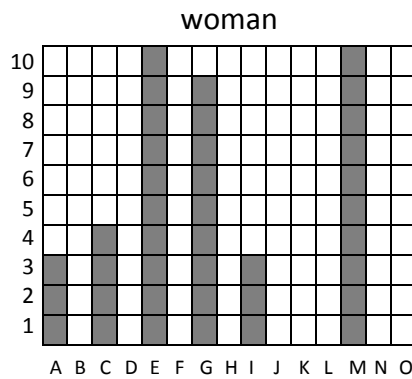
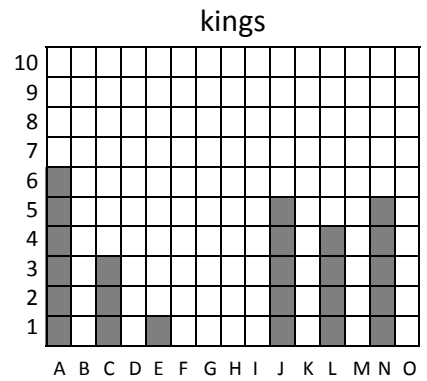
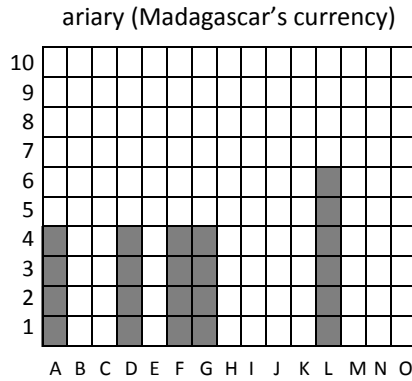
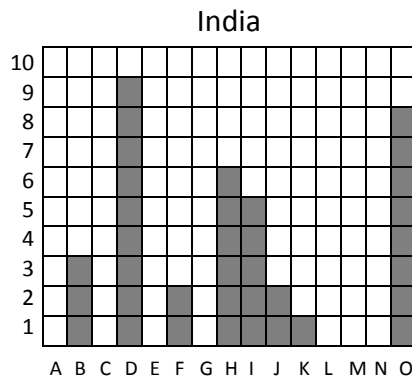


(K) Kings, Queens, and Counts (2/5)

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.

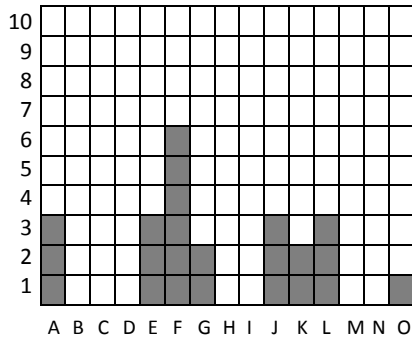


(K) Kings, Queens, and Counts (3/5)

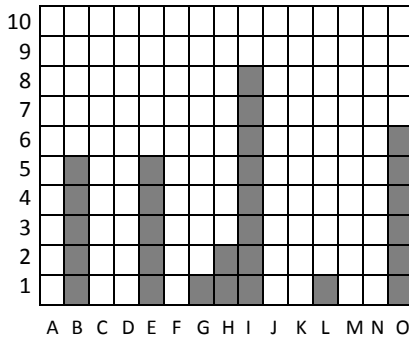


(K) Kings, Queens, and Counts (4/5)

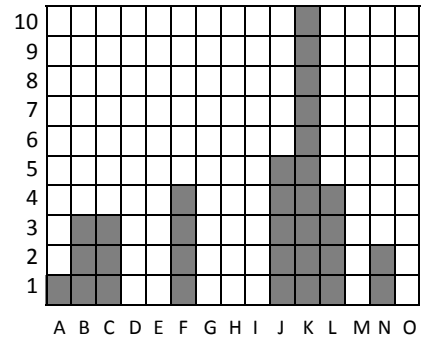
Antananarivo (Madagascar's capital)



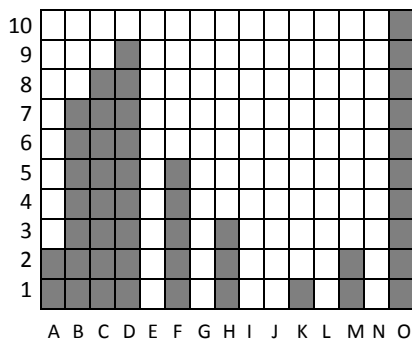
mystery word #1



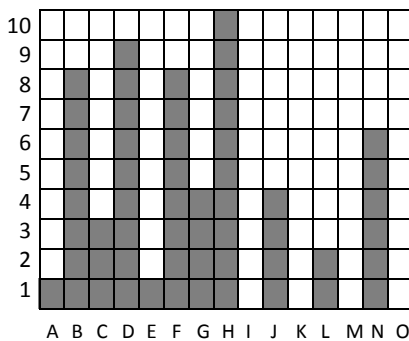
mystery word #2



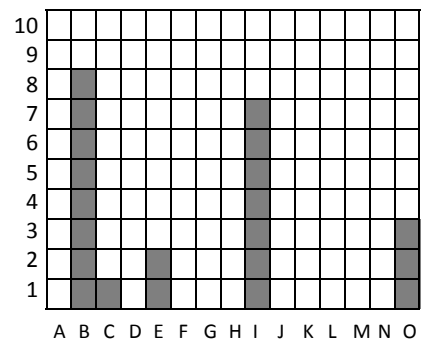
mystery word #3



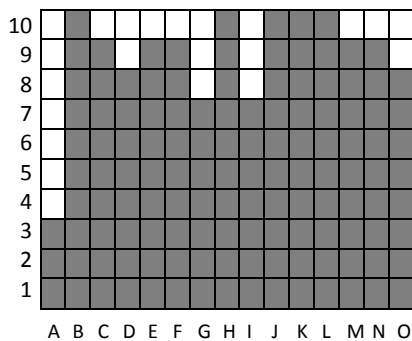
mystery word #4



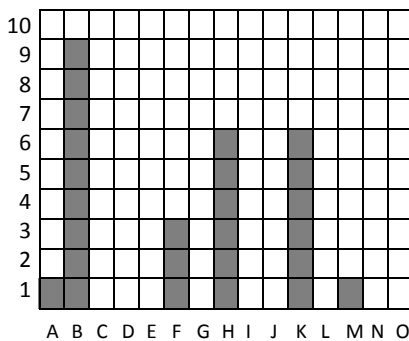
mystery word #5



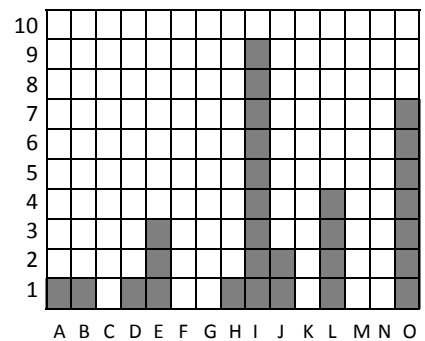
mystery word #6



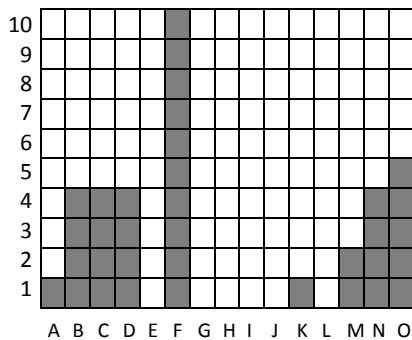
mystery word #7



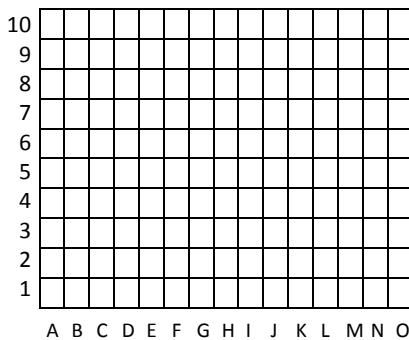
mystery word #8



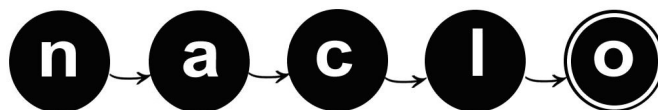
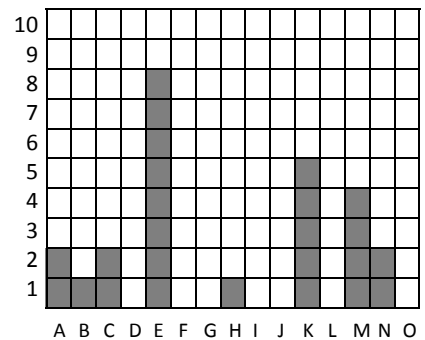
mystery word #9



mystery word #10



mystery word #11



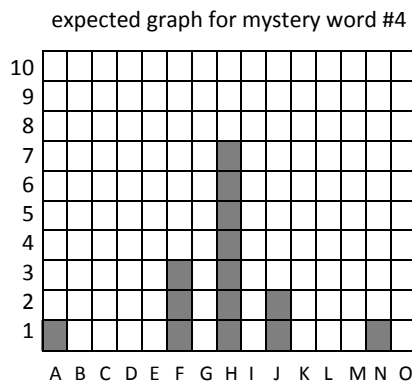
(K) Kings, Queens, and Counts (5/5)

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

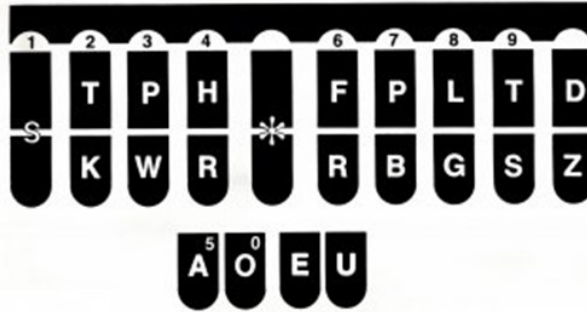
On your Answer Sheet, write the number of the mystery word corresponding to each definition.

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)

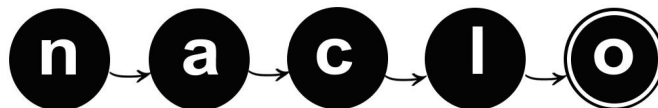
							O							P	B						
S	T	K	P	W	H	R								F	R	P	B	L	G	T	S
					H		O			U											
	T	K					O			U											
			P		H	R	A	O		E											D

(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							



(L) The Short Hand of the Law (2/2)

(D)

T				O				
	K			O	U	P B	T S	
		W		O		P B		
			A			P B		D
T	W			O				

(E)

				E	P B		
T				E	R		
			A				
	P	H R	A O	E			
			A				T

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?

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

L3. Explain your answer.



(M) Sound Judgments (1/2) [5 points]

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: $p > \emptyset / _ \#$ 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.

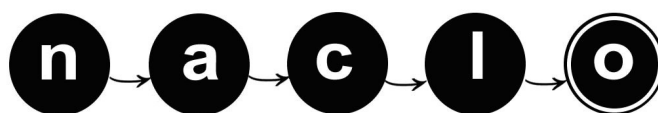
Because many ancient languages were never written down until recent millennia, linguists have to rely on clever deductions to work out the details of their early history. Our only records of Tocharian are some 9th century manuscripts around the Tarim Basin in western China, so our knowledge of its development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

stage	"share"	"row of teeth"	"knee"	"war"	"one hundred"	"dog"	"prop"
(A)	pákos	kómos	kónu	kóro-	kṃtóm	kuó	stema-
(B)	págos	gómos	gónu	kóro-	kṃtóm	kuó	stema-
(C)	b ^h ágos	jómb ^h os	jónu	kóro-	kṃtóm	kuó	stemb ^h a-
(D)	b ^h ágos	jómb ^h os	jónu	kóro-	cṃtóm	cuó	stemb ^h a-
(E)	páko	kómo	kónu	kóro-	kṃtóm	kuó	stema-
(F)	b ^h ágos	gómos	gónu	kóro-	kṃtóm	kuó	stema-
(G)	b ^h ágos	gómb ^h os	gónu	kóro-	kṃtóm	kuó	stemb ^h a-

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:

	Early Tocharian	Later Tocharian
“they drive”	*agonti	*akën
“they are driven”	*agontor	*akëntär
“ten”	*dékamt	*śákə
“one hundred”	*kəmtóm	*kənté
“stag hunter”	*kēruwos	*śerəwë
“father”	*patér	*pacér
“running” (later > “river”)	*ték ^w os	*cák ^w ë
“that”	*tód	*té
“twenty”	*wíkanti	*w ^ʷ íkən

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 > ə (D) or > ur / __ #
 (E) ē, ī, ū > 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 > ə (L) k > ś / __ e, ē
 (M) w > w^ʷ / __ i, ī (N) g > k (O) d > Ø / __ # (P) s > Ø / __ #
 (Q) ti > Ø / n __ #

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.



(N) What happened at the chess tournament? (1/1) [5 points]

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. |
| N6. Senkit nem vert meg a Petya. | (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. |
| N9. Mindenki megvert valakit. | (I) Everyone beat someone. |
| N10. Valaki megverte a Marcit. | (J) Who didn't beat anyone? |
| N11. Senkit nem vert meg senki. | (K) There's someone who didn't beat anyone. |
| N12. Kit nem vert meg senki? | (L) Peter beat no one. |
| N13. Nem vertem meg senkit. | (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.

Nung	English
Cáu ca vủhn nahng kíhn.	I was about to continue to eat it.
Cáu cháhn slờng páy mi?	Do I truly want to go?
Cáu mi slày kíhn.	I don't have to eat it.
Cáu ngám hẻt pehn tế.	I did it like that just now.
Cáu tan đohc hẻn mưhng.	I only saw you.
Cáu vủhn nahng bô sạhm tẻng hẻt hờn.	I also continue to build the house alone.
Da kíhn!	Don't eat it!
Da khải hờn!	Don't sell the house!
Mưhng chớng ca cháhn fải khải.	Then she truly was about to have to sell it.
Mưhng mi cháhn đày non.	She truly can't sleep.
Mưhng náhc-thày chớng bô sạhm kíhn.	Then she also just previously ate it.
Mưhng náhc-thày slờng tẻng páy.	You wanted to go alone just previously.

01. Cáu cháhn đày non.
02. Da páy non!
03. Mưhng bô sạhm mi slờng hẻt hờn mi?
04. Mưhng ngám bô sạhm páy hờn.
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?



(P) A Matter of Horn Clauses (1/2) [10 points]

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:

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

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:

$$S(\text{Mary eats}). \quad S(\text{Mary sleeps}). \quad S(\text{John sleeps}). \quad S(\text{John eats}).$$

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

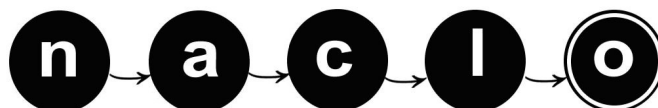
$$\begin{array}{ll} S(xy) :- Nsg(x), Vsg(Y). & S(xy) :- Npl(x), Vpl(Y). \\ Nsg(\text{Mary}). & Npl(\text{dogs}). \\ Vsg(\text{sleeps}). & Vpl(\text{sleep}). \end{array}$$

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

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.



(P) A Matter of Horn Clauses (2/2)

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 $S(\text{help John let Kim let John})$.

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

Jan säit das mer em Hans em Jan es huus hälfed hälfe aastriiche.

Jan says that we helped Hans help Jan paint the house.

Jan säit das mer em Hans em Jan em Hans es huus hälfed hälfe hälfe 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 hälfed hälfe hälfe hälfe 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 lönd laa aastriiche.

Jan says that we let Hans let Jan paint the house.

Jan säit das mer de Hans em Jan de Hans es huus lönd hälfe laa 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 lönd hälfe hälfe laa hälfe 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:

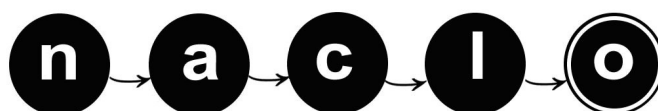
$$R(wy, xz) :- T(w, x), T(y, z).$$

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 $T(\text{the, cat})$. 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 $S(\text{de Hans em Jan em Hans laa hälfe hälfe})$.

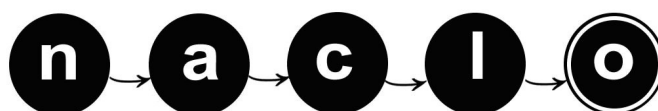


(Q) A Cup of Javanese (1/3) [15 points]

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.

Javanese script	Latin Script	Meaning
	penyakit	disease
	Inggris	England
	traktor	tractor
	panyumbang	donor
	rembulan	moon
	tansah	always
	Amérika	America (continent)



(Q) A Cup of Javanese (2/3)

ꦒꦸꦫꦸꦥ	ngrebut	to grab
ꦲꦧꦸꦏꦺ	ibukota	capital
ꦲꦂꦂꦺꦤꦠ	Argentina	Argentina
ꦱꦺꦤꦒꦺꦁꦺ	srengéngé	sun
ꦥꦱꦸ	palsu	false
ꦫꦺꦂꦺꦁꦒꦺꦤ	rerenggan	decoration
ꦲꦁꦱꦭ	angsal	to acquire
ꦲꦁꦒꦶ	inggih	yes
ꦲꦁꦱꦸ	(a)	often



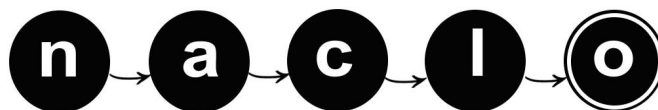
(Q) A Cup of Javanese (3/3)

ꦭꦏꦱꦏꦸꦁ	(b)	letter / script
ꦲꦸꦏꦸꦁ	(c)	to unload
ꦲꦸꦏꦸꦁꦲꦸꦏꦸꦁ	(d)	to examine
ꦲꦸꦏꦸꦁꦲꦸꦏꦸꦁ	(e)	to cancel

Q2. Write these words in the Javanese script

	Latin Script	Meaning
a.	nyolong	to steal
b.	sepalih	half
c.	trengginas	lively
d.	Antartika	Antarctica
e.	Istanbul	Istanbul

Q3. Explain your answers.



(R) Changing the Subject (1/2) [5 points]

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

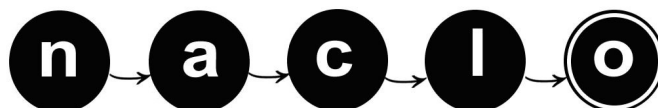
c: same as 'x', but voiced

r: a rolled 'r' as in Italian

': the consonant sound in the middle of *uh-oh*

1st person	
Somali word	English translation
akhriyay	I read
aragay	I saw
(a)	I taught
ba'ay	I was destroyed
baajiyay	I prevented
baaqay	I announced
baxay	I left
bi'iyay	I destroyed
bilaabay	I began
(e)	I ate
cabay	I drank
cararay	I ran away

3rd person singular	
Somali word	English translation
akhriday	He read
aragtay	He saw
bartay	He taught
ba'day	He was destroyed
(b)	He prevented
baaqday	He announced
baxday	He left
(c)	He destroyed
(d)	He began
cuntay	He ate
cabtay	He drank
carartay	He ran away

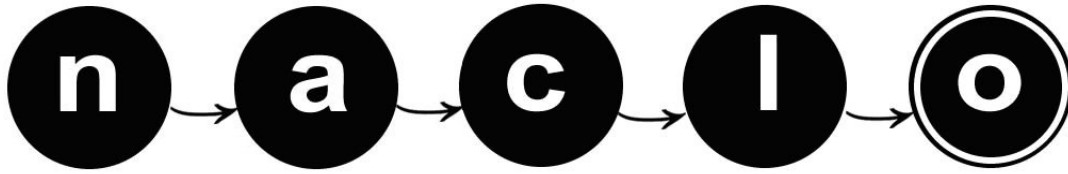


(R) Changing the Subject (2/2)

1st person	
Somali word	English translation
daaqay	I grazed
dhacay	I fell
dhisay	I built
diiday	I refused
dilay	I killed
faraxay	I was happy
gaadhay	I reached
galay	I entered
go'ay	I cut
(k)	I found
horjeeday	I opposed
kacay	I rose
keenay	I brought
korodhay	I increased
qaaday	I took
tagay	I went
xidhay	I closed
walaaqay	I stirred

3rd person singular	
Somali word	English translation
(f)	He grazed
(g)	He fell
dhistay	He built
diiday	He refused
dishay	He killed
(h)	He was happy
gaadhay	He reached
(i)	He entered
(j)	He cut
heshay	He found
horjeeday	He opposed
(l)	He rose
keentay	He brought
korodhay	He increased
(m)	He took
tagtay	He went
(n)	He closed
(o)	He stirred





The North American Computational Linguistics Olympiad

Contest Booklet

REGISTRATION NUMBER

--	--	--	--

Name: _____

Contest Site: _____

Site ID: _____

City, State: _____

Grade: _____

Start Time: _____

End Time: _____

YOUR NAME:

REGISTRATION #

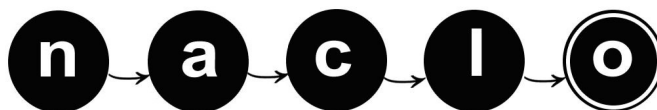
Answer Sheet (1/8)

(I) Deriving Enjoyment

1.	a.	<input type="text"/>	b.	<input type="text"/>	c.	<input type="text"/>	d.	<input type="text"/>
	e.	<input type="text"/>	f.	<input type="text"/>	g.	<input type="text"/>	h.	<input type="text"/>
	i.	<input type="text"/>	j.	<input type="text"/>	k.	<input type="text"/>	l.	<input type="text"/>
	m.	<input type="text"/>	n.	<input type="text"/>	o.	<input type="text"/>	p.	<input type="text"/>
	q.	<input type="text"/>	r.	<input type="text"/>				
2.	<input type="text"/>		<input type="text"/>					
3.	<input type="text"/>		<input type="text"/>					

(J) Get Edumacated!

1.	a.	<input type="text"/>	b.	<input type="text"/>	c.	<input type="text"/>	d.	<input type="text"/>	e.	<input type="text"/>	f.	<input type="text"/>	g.	<input type="text"/>
2.	a.	<input type="text"/>	b.	<input type="text"/>	c.	<input type="text"/>								
	d.	<input type="text"/>	e.	<input type="text"/>	f.	<input type="text"/>								
	g.	<input type="text"/>	h.	<input type="text"/>	i.	<input type="text"/>								
	j.	<input type="text"/>	k.	<input type="text"/>	l.	<input type="text"/>								
	m.	<input type="text"/>	n.	<input type="text"/>	o.	<input type="text"/>								
	p.	<input type="text"/>	q.	<input type="text"/>	r.	<input type="text"/>								



YOUR NAME:

REGISTRATION #

Answer Sheet (2/8)

(J) Get Educated! (cont.)

3.

(K) Kings, Queens, and Counts

1.

is

8																			
7																			
6																			
5																			
4																			
3																			
2																			
1																			

whether the weather is good or not cold hot we'll whatever we like it

2. a. b. c. d. e. f. g. h. i. j. k.



YOUR NAME:

REGISTRATION #

Answer Sheet (3/8)

(K) Kings, Queens, and Counts (cont.)

3.

(L) The Short Hand of the Law

1. First Last

2.

3.



YOUR NAME:

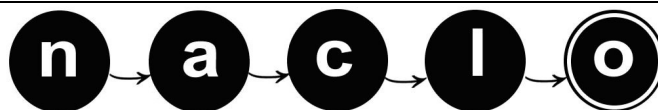
REGISTRATION #

Answer Sheet (4/8)

(M) Sound Judgments

1. Earliest Latest

2.



YOUR NAME:

REGISTRATION #

Answer Sheet (5/8)

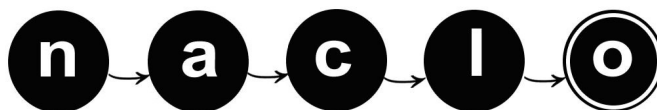
(N) What happened at the chess tournament?

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.



YOUR NAME:

REGISTRATION #

Answer Sheet (6/8)

(P) A Matter of Horn Clauses

1.

2.

(Q) A Cup of Javanese

1.

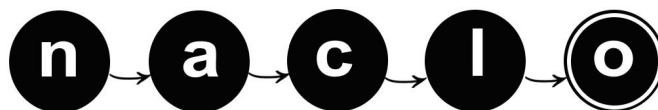
a.

c.

e.

b.

d.



YOUR NAME:

REGISTRATION #

Answer Sheet (7/8)

(Q) A Cup of Javanese (cont.)

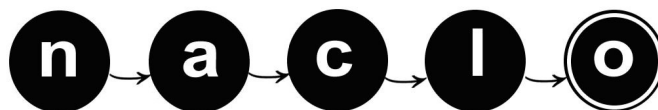
2. a.

b.

c.

d.

e.



YOUR NAME:

REGISTRATION #

Answer Sheet (8/8)

(Q) A Cup of Javanese (cont.)

3.

(R) Changing the Subject

1. a.	<input type="text"/>	b.	<input type="text"/>	c.	<input type="text"/>
d.	<input type="text"/>	e.	<input type="text"/>	f.	<input type="text"/>
g.	<input type="text"/>	h.	<input type="text"/>	i.	<input type="text"/>
j.	<input type="text"/>	k.	<input type="text"/>	l.	<input type="text"/>
m.	<input type="text"/>	n.	<input type="text"/>	o.	<input type="text"/>



YOUR NAME:

REGISTRATION #

Extra Page (1/2)



YOUR NAME:

REGISTRATION #

Extra Page (2/2)

