nacl→o

NAACL

YAHOO!

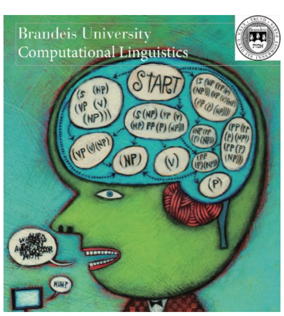
Linguistic Society of America

DARPA

NSF

ARL

Carnegie Mellon University
Language Technologies Institute



MICHIGAN

The Eleventh Annual

North American Computational Linguistics Olympiad

2017

www.nacloweb.org

Open Round
January 26, 2017

Serious language puzzles that are surprisingly fun!

-Will Shortz, Crossword editor of The New York Times and Puzzlemaster for NPR

Welcome to the eleventh 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 three hours long and includes eight problems, labeled A to H.
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 this year's open round, no points will be given for explanations. Instead, make sure to fill out all the answer boxes properly.
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 10% of participants (approximately) across the continent in the open round will be invited to the second round.
9. Each problem has been thoroughly checked by linguists and computer scientists as well as students like you for clarity, accuracy, and solvability. Some problems are more difficult than others, but all can be solved using ordinary reasoning and some basic analytic skills. You don't need to know anything about linguistics or about these languages in order to solve them.
10. If we have done our job well, very few people will solve all these problems completely in the time allotted. So, don't be discouraged if you don't finish everything.
11. **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 2017 Organizers

Program Committee:

Adam Hesterberg, Massachusetts Institute of Technology (co-chair)
Alan Chang, University of Chicago
Aleka Blackwell, Middle Tennessee State University
Alex Wade, Stanford University
Alexander Iriza, Princeton University
Andrew Lamont, University of Massachusetts, Amherst
Babette Newsome, Aquinas College
Ben Sklaroff, University of California, Berkeley
Caroline Ellison, Stanford Univeristy
Daniel Lovsted, McGill University
David McClosky, IBM
David Mortensen, Carnegie Mellon University
David Palfreyman, Zayed University
Dick Hudson, University College London
Dorottya Demszky, Princeton University
Dragomir Radev, Yale University (co-chair)
Elisabeth Mayer, Australian National University
Elysia Warner, University of Cambridge
Harold Somers, All Ireland Linguistics Olympiad
Harry Go, Washington University in St. Louis
Heather Newell, Université du Québec à Montréal
James Pustejovsky, Brandeis University
Jason Eisner, Johns Hopkins University
Jonathan Graehl, SDL International
Jonathan Kummerfeld, University of Michigan
Jonathan May, ISI
Jordan Boyd-Graber, University of Colorado
Jordan Ho, University of Toronto
Josh Falk, University of Chicago
Julia Buffinton, University of Maryland
Kai Low, University College London
Kevin Watson, University of Canterbury
Lars Hellan, Norwegian University of Science and Technology
Lori Levin, Carnegie Mellon University
Lynn Clark, University of Canterbury
Mary Laughren, University of Queensland
Oliver Sayeed, University of Cambridge
Patrick Littell, Carnegie Mellon University
Rachel McEnroe, University of Chicago
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

NACLO 2017 Organizers (cont'd)

Problem Credits:

Problem A: Tom McCoy
Problem B: James Hyett
Problem C: Jordan Ho, Patrick Littell, and Tom McCoy
Problem D: Deryle Lonsdale
Problem E: Kai Low Rui Hao
Problem F: Harold Somers
Problem G: Kai Low Rui Hao
Problem H: Tom McCoy

Organizing Committee:

Adam Hesterberg, Massachusetts Institute of Technology
Adrienne Ballsmeyer
Aleka Blackwell, Middle Tennessee State University
Alex Wade, Stanford University
Alexander Iriza, Princeton University
Andrew Lamont, University of Massachusetts, Amherst
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, Yale University
Haley Barbieri, Bennington College
Harry Go, Washington University in Saint Louis
Heather Newell, Université du Québec à Montréal
James Pustejovsky, Brandeis University
James Bloxham, Massachusetts Institute of Technology
Janis Chan, University of Western Ontario
Jordan Ho, University of Toronto
Josh Falk, University of Chicago
Julia Buffinton, University of Maryland
Lori Levin, Carnegie Mellon University
Mary Jo Bensasi, Carnegie Mellon University
Matthew Gardner, Allen Institute for Artificial Intelligence
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
Yilin Guo, University of Michigan
Yilu Zhou, Fordham University

NACLO 2017 Organizers (cont'd)

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

Canadian Coordinator and Team Coach:
Heather Newell, Université du Québec à Montréal

USA Contest Site Coordinators:
Brandeis University: James Pustejovsky
Brigham Young University: Deryle Lonsdale
Carnegie Mellon University: Mary Jo Bensasi, Lori Levin
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: Emma Manning
Goshen College: Peter Miller
Indiana University: Melinda Bristow-Meadows, Rachel Hertz
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, Louise Chan, Dean Kelley, Richard Roiger
Montclair State University: Anna Feldman
Northeastern Illinois University: J. P. Boyle, R. Hallett, Judith Kaplan-Weinger, K. Konopka
Ohio State University: Micha Elsner, Michael White
Oregon State University: Liang Huang
Princeton University: Dora Demszky, Yiwei Liu, Misha Khodak, Christiane Fellbaum
San Diego State University: Rob Malouf
Southern Illinois University: Vicki Carstens, Jeffrey Punske
SpringLight Education Institute: Sherry Wang
Stanford University: Chris Manning, Ignacio Cases
Stony Brook University: Lori Repetti, Sarena Romano
the Mathletics: Shirley Ma
Union College: Kristina Striegnitz, Nick Webb
University of Alabama, Birmingham: Steven Bethard
University of California, Irvine: Sameer Singh
University of Colorado at Boulder: Silva Chang
University of Houston: Thamar Solorio
University of Illinois at Urbana-Champaign: Julia Hockenmaier, Benjamin Leff, Greg Sigalov
University of Maryland: Kasia Hitczenko
University of Massachusetts, Amherst: Stacey Chobany, Andrew Lamont, Mina Puig, Andrew Wang, UMass Linguistics Club
University of Massachusetts, Lowell: Anna Rumshisky, David Donahue, Willie Boag
University of Memphis: Vasile Rus
University of Michigan: Steven Abney, Marcus Berger, Sally Thomason
University of Nebraska, Omaha: Parvathi Chundi
University of North Carolina, Charlotte: Hossein Hematialam, Wlodek Zadrozny
University of North Texas: Genene Murphy, Rodney Nielsen
University of Pennsylvania: Marianna Apidianaki, Chris Callison-Burch, Anne Cocos, Cheryl Hickey, Mitch Marcus, Derry Wijaya
University of Southern California: Jonathan Gordon, Nima Pourdamghani
University of Texas at Dallas: Jing Lu, Vincent Ng, Isaac Persing
University of Texas, Austin: Rainer Mueller
University of Utah: Heather Burkhart
University of Washington: Jim Hoard, Joyce Parvi
University of Wisconsin, Madison: Steve Lacy
University of Wisconsin, Milwaukee: Joyce Boyland, Hanyon Park, Gabriella Pinter, Anne Pycha
Western Washington University: Kristin Denham
Yale University: Tom McCoy, Raffaella Zanuttini

NACLO 2017 Organizers (cont'd)

Canada Contest Site Coordinators:
Dalhousie University: Magdalena Jankowska, Vlado Keselj, Dijana Kosmajac, Armin Sajadi
McGill University: Michael Wagner, Lisa Travis
Simon Fraser University: John Alderete, Marion Caldecott, Maite Taboada
University of Alberta: Herbert Colston, Sally Rice
University of British Columbia: Hotze Rullmann, Jozina Vander Klok
University of Calgary: Dennis Storoshenko
University of Ottawa: Diana Inkpen
University of Toronto: Jordan Ho, James Hyett
University of Western Ontario: Janis Chang

High school sites: Dragomir Radev

Booklet Editors:
Andrew Lamont, University of Massachusetts, Amherst
Dragomir Radev, University of Michigan

Sponsorship Chair:
James Pustejovsky, Brandeis University

Sustaining Donors
Linguistic Society of America
NAACL
NSF
Yahoo!

Major Donors
ARL
DARPA

University Donors
Brandeis University
Carnegie Mellon University
University of Michigan

Many generous individual donors

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

NACLO 2017 Sites

Brandeis
Johns Hopkins University
University of Colorado Boulder
UT Dallas — The University of Texas at Dallas
Carnegie Mellon
SFU Simon Fraser University
The University of Texas at Austin
BYU Brigham Young University
Massachusetts Institute of Technology
University of Houston
University of Toronto
College of William & Mary 1693
McGill
SIU Carbondale
SpringLight Education Institute
Illinois — University of Illinois at Urbana-Champaign
Middle Tennessee State University
Stanford University
University of Maryland
Columbia University
Minnesota State University Mankato
Stony Brook — State University of New York
UMass Lowell
University of Massachusetts Amherst 1863
The University of Utah
Cornell University Founded A.D. 1865
Montclair State University
The University of Memphis
University of Washington 1861
Western
Dalhousie University — Inspiring Minds
Northeastern Illinois University
Michigan
University of Nebraska Omaha
Dartmouth
Ohio State
Union College
UNC Charlotte
The University of Wisconsin Madison
Emory University
Oregon State University — OSU
UAB The University of Alabama at Birmingham
UWM University of Wisconsin Milwaukee
Georgetown University 1789
University of Lethbridge
University of Alberta
University of North Texas
Princeton University
UBC
uOttawa
Goshen College
San Diego State University
Penn — University of Pennsylvania
Western Washington University
University of Wisconsin Eau Claire 1916
University of Calgary
The University of California Irvine
University of Southern California 1880
Yale University

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

(A) A Little Tshiluba (1/1) [5 points]

Tshiluba, spoken by about 6 million people, is one of the official languages of the Democratic Republic of the Congo. Below are some sentences in Tshiluba, along with their English translations:

Tshiluba	*English*
mukaji uvwa mumona muana.	The woman saw the child.
bakaji bavwa bamona muana.	The women saw the child.
muluma uvwa mumona bakaji.	The man saw the women.
muluma uvwa mumona bambuji.	The man saw the goats.
banzolu bavwa bamona bantambwe.	The chickens saw the lions.
tubambwa tuvwa tumona baluma.	The small dogs saw the men.
mbwa uvwa mumona ntambwe.	The dog saw the lion.
ntambwe uvwa mumona tubanzolu.	The lion saw the small chickens.
kanzolu kavwa kamona tubantambwe.	The small chicken saw the small lions.
tubakulu tuvwa tumona mbwa.	The small adults saw the dog.
kamuntu kavwa kapeta kantambwe.	The small person found the small lion.

Answer these questions in the Answer Sheets.

A1. Translate the following into Tshiluba:

a. the dog
b. The man saw the child.
c. The chicken saw the dogs.
d. The adult found the goat.
e. The small goats found the small child.

A2. Tshiluba belongs to a group of languages known as the Bantu languages. What does *bantu* mean in Tshiluba?

A3. The Tshiluba word for "fruit" is *cimuma*, and the Tshiluba word for "fruits" is *bimuma*. Translate the following into English:

a. cimuma civwa cimona ntambwe.
b. ntambwe uvwa mumona tubimuma.

(B) Phở Bar (1/2) [15 points]

Menus often use non-English names for dishes that originate from other countries, along with descriptions in English. For example, a Vietnamese take-out menu might list Gỏi Cuốn: Salad rolls. Below are the names and descriptions of twenty dishes from a Vietnamese take-out menu in arbitrary order. Identify the correct correspondences; write your answers in the Answer Sheets.

Also note that two of the below dishes come from a section with the following description:

> Our Famous Vietnamese Noodle Soup. Choice of rice noodle "Phở" or yellow noodle "Mì" and your selected finest meat in an aromatic beef broth with scallions, onion, and cilantro. Soups are served with a plate of bean sprouts, fresh basil, sliced lime, jalapeno.

B1.	Soup Hoành Thánh
B2.	Chim Cút Rôti
B3.	Bánh Oai Vạc Chiên Hoặc Hấp (6 pcs.)
B4.	Bánh Bột Chiên Hành (8 pcs.)
B5.	Bún Xào Đo Biến
B6.	Bánh Xèo
B7.	Gỏi Tôm Hoặc Gỏi Gà
B8.	Gỏi Ngo Sen
B9.	Phở Hoặc Mì Gà
B10.	Bò Xào Cà Ry
B11.	Bún Bò Huế
B12.	Thịt Lụi
B13.	Bún Thịt Nướng
B14.	Bún Chả Giò
B15.	Chả Giò (2 pcs.)
B16.	Bún Thịt Nướng Chả Giò
B17.	Bún Tôm
B18.	Bún Tôm Thịt Nướng Chả Giò
B19.	Cá Salmon Hoặc Cá Bông Lau Hấp
B20.	Mì Xào

(B) Phở Bar (2/2)

(A)	Lotus Stem Salad.
(B)	Vietnamese Crepe. A traditional mixture of shrimp and pork, bean sprouts and a delicate sauce folded into a rice powder pancake.
(C)	Noodle Soup with shredded chicken.
(D)	Wonton Soup. Shrimp, pork dumplings, lettuce, onion and scallions in chicken broth.
(E)	Beef stew “Huế Style”. Spicy lemon grass beef noodle soup and shrimp.
(F)	Scallions Pancake. Fried sweet flour with scallions.
(G)	Vermicelli[1] with seafood sautéed with lemon grass sauce.
(H)	Vermicelli with crispy spring rolls.
(I)	Roasted Quail.
(J)	Filet of salmon or catfish steamed with ginger, scallions, and Chef’s special sauce.
(K)	Vermicelli with choice of grilled meat.
(L)	Beef sautéed in curry sauce.
(M)	Vermicelli with a choice of grilled meat with crispy spring rolls.
(N)	Vermicelli with shrimp, choice of grilled meat, and crispy spring rolls.
(O)	Saigon Ravioli Fried or Steamed. Homemade dumplings filled with a mixture of chicken, pork, and vegetable. Served with ginger dipping sauce.
(P)	Teriyaki with choice of meat.
(Q)	Vietnamese salad. Choice of poached shrimp, chicken, or combination, with shredded carrots, cabbage, fresh mints, roasted peanuts, onions and homemade dressing.
(R)	Crispy Spring Rolls. A savory mixture of ground pork, taro, carrots, onion, rice vermicelli, and mushroom wrapped in spring roll and fried golden brown.
(S)	Vermicelli with grilled shrimp.
(T)	Sautéed crispy yellow noodle.

[1]Vermicelli are long, slender noodles.

(C) LOLWUT (1/2) [10 points]

Recognizing words – like that the sequence of letters "c-a-t" represents the word "cat" – is the foundation to any technology that works on text… but it's not always so easy, especially on the internet where writers so often vary their spelling to express emphasis, emotion, surprise, etc. How can you recognize the word "what" when it could appear as "whaaaat" or "whaaa" or "wat" or "waaat" or "whut" or "wut" or…?

One way to recognize many variants at once is to use a *regular expression* (also called a "regex") – a special sequence of symbols where you can indicate that a letter is optional, that it can occur many times, that one of several letters might occur, etc. For example, the regular expression

`wh?a+t*`

means that the "h" can occur exactly once or not at all, the "a" can occur one or more times, and the "t" can occur any number of times (including zero times). The symbols used in this regex are defined as follows (in this example, *unit* refers to single letters, but as you will see below, a unit can be larger):

- `?` The previous unit can occur zero or one times
- `*` The previous unit can occur zero or more times
- `+` The previous unit can occur one or more times.

So the regular expression "`wh?a+t*`" would "recognize" the words below (and infinitely many more):

what wha whaaa wat waaaa watt waaaatttt waaaaaat

It won't, however, recognize the word "wht" (because there's no "a" in this word and it requires at least one "a"), "whut" (there's no "u" allowed here), "whhhat" (because it only allows zero or one "h" and this word has three), or "waaaah" (because the "h", if it occurs, must precede the "a"; no reordering is allowed).

There are many more symbols that can appear in regular expressions, but for this problem you only need the above three symbols and parentheses. Parentheses group letters into units that themselves can be operated on by symbols, so that `L(OL)+` would recognize any of the words below (and infinitely many more):

LOL LOLOL LOLOLOL LOLOLOLOL

It won't, however, recognize the word "L" (because it needs at least one instance of "OL"), or "LOOOOL" (because the entire "OL" part needs to be repeated, not just the "O"), or "LOOLL" (because the "OL" needs to repeated as a unit).

Answer the questions on the next page in the Answer Sheets.

(C) LOLWUT (2/2)

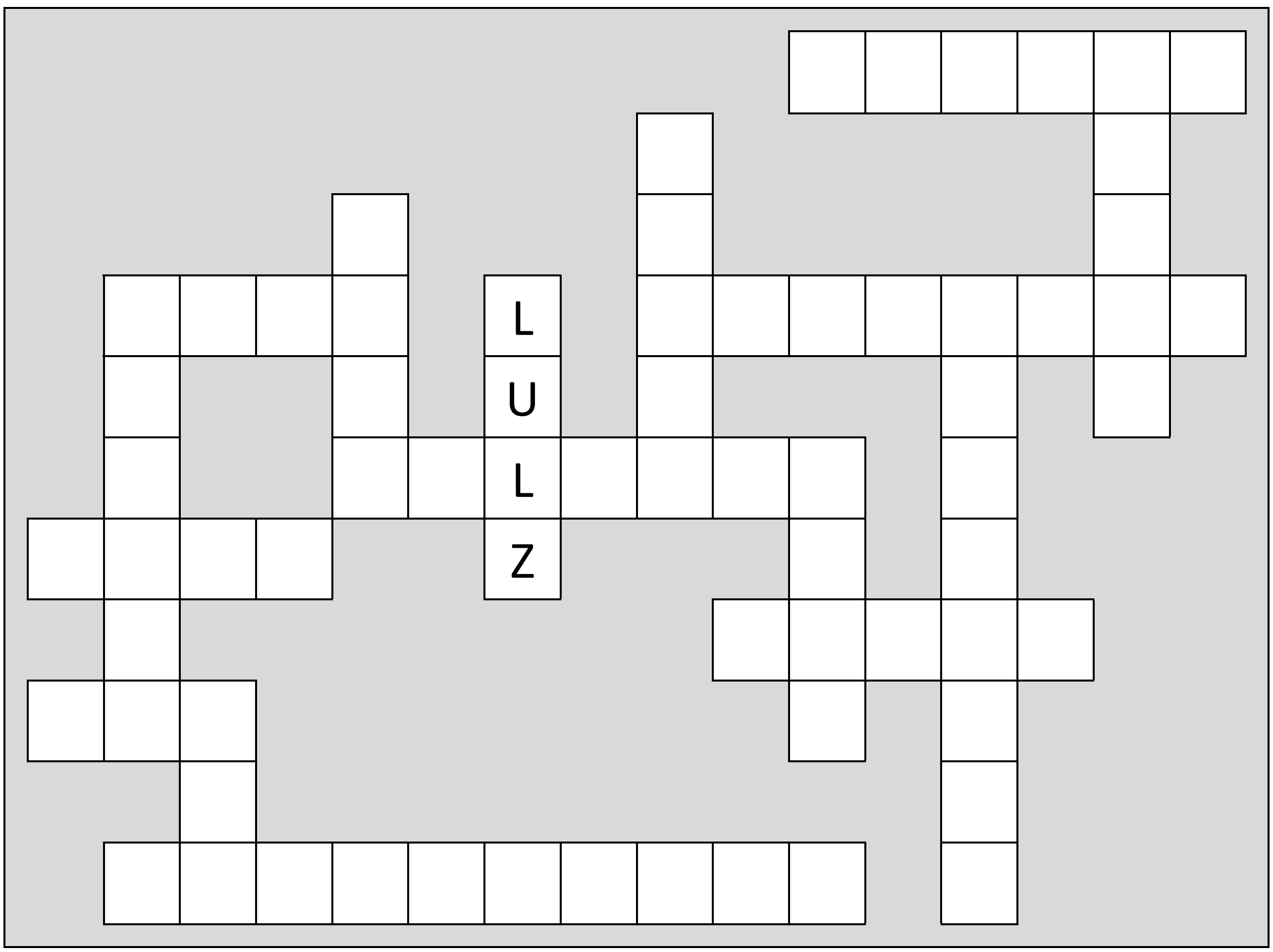
C1. We've put together a small crossword puzzle entirely of internet laughter, and clued each entry with a regular expression like "`L(OL)+`".

The clues do not appear in any particular order, so you'll have to work out for yourself where each entry goes, but each clue corresponds to only one entry. To help you get started, we've entered one answer into the grid already (there is no clue for this pre-entered answer). Be sure to write your answers in the Answer Sheets; nothing on this page will be graded!

Clues:

`L(OL)+`	`(HO)+`	`K(EK)*E`	`(HAR+)+`
`H(EH)+`	`ROT?FL`	`TE(HE+)+`	`LAW*L`
`MWA(HA)+`	`HE(HE)+`	`LO+L`	`HAHA*`
`(AH)+A+`	`HA+`	`(JA)+`	

Puzzle:



C2. For each of the following regexes, write the shortest word that it could describe.

a. `LOL*LO?(OL)?O+L`
b. `(A?HA)+(LO)?O*`

(D) Let's Roll! (1/1) [10 points]

Tavla is a variant of backgammon played in Turkey. In the game, two six-sided dice are thrown, resulting in two random numbers between 1 and 6. Each possible outcome has a name, and these names consist of a curious mixture of Persian and Turkish. Some letters used to write these languages are not used to write English, such as *ü*, *ş*, and *ı*, but you do not need to know how these are pronounced to solve this problem. The table on the left lists some of the possible outcomes, while the table on the right lists the names of those outcomes, but the outcomes and their names are not given in the same order. Answer these questions in the Answer Sheets.

D1. Match up which name goes with which outcome. Write your answers in the Answer Sheets (that is, next to each outcome in the Answer Sheet, write the letter corresponding to that outcome's name). Note that the order of the dice does not matter for the naming scheme: For example, 2-6 and 6-2 have the same name.

	Outcomes
a.	1-3
b.	4-6
c.	1-5
d.	3-5
e.	1-4
f.	1-6
g.	2-6

	Names
(A)	pencü yek
(B)	pencü se
(C)	şeşi yek
(D)	şeş cıhar
(E)	se yek
(F)	şeşi dü
(G)	cıharı yek

For Questions **D2** through **D4**, your answer should be an element from the following list of possible outcomes: 1-1, 1-2, 1-3, 1-4, 1-5, 1-6, 2-2, 2-3, 2-4, 2-5, 2-6, 3-3, 3-4, 3-5, 3-6, 4-4, 4-5, 4-6, 5-5, 5-6, 6-6

D2. What outcome has the name "pencü dü"?

D3. What outcome has the name "cıhari se"?

D4. What outcome has the name "düşeş"?

NOTE: There is significant variation in the exact names used across different Tavla players from different regions. Therefore, some Tavla players use different sets of names than the ones given in this problem.

(E) On the Right Track (1/1) [20 points]

Tamil is predominantly spoken by the Tamil people of Tamil Nadu, a state in southern India, and Sri Lanka. Tamil is also one of the official languages of Singapore, along with English, Malay, and Mandarin.

Singapore Mass Rapid Transit (SMRT), one of Singapore's train operators, has translated their station names into two other languages from English, namely Mandarin and Tamil.

You have been wrongly issued a list of North-South Line (NSL) train stations in Tamil. These Tamil names are listed below and numbered from 1 to 25. Match the Tamil names to their English names (which have been assigned the letters from (A) to (Y)). Write your answers in the Answer Sheets.

E1.	ஜூரோங் கிழக்கு
E2.	அங் மோ கியோ
E3.	அட்மிரல்டி
E4.	ஆர்ச்சர்ட்
E5.	இயூ டீ
E6.	இயோ சூ காங்
E7.	உட்லண்ட்ஸ்
E8.	காதிப்
E9.	கிராஞ்சி

E10.	சாமர்செட்
E11.	சுவா சூ காங்
E12.	செம்பாவாங்
E13.	டோபி காட்
E14.	தோ பாயோ
E15.	நகர மண்டபம்
E16.	நியூட்டன்
E17.	நொவீனா
E18.	பிரேடல்

E19.	பீஷான்
E20.	புக்கிட் கொம்பாக்
E21.	புக்கிட் பாத்தோக்
E22.	மரீனா பே
E23.	மார்சிலிங்
E24.	யீஷூன்
E25.	ராஃபிள்ஸ் பிளேஸ்

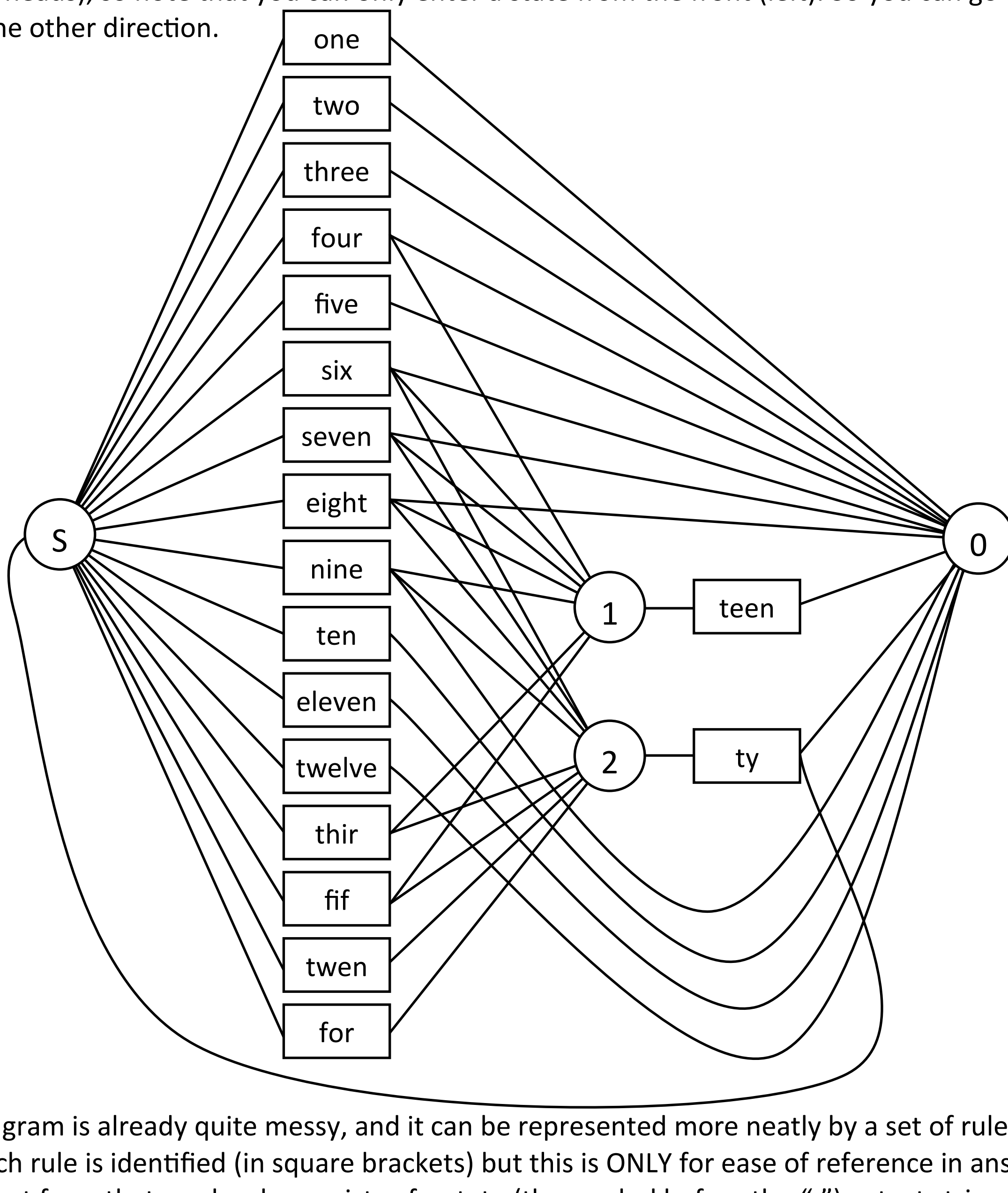
(A)	Jurong East
(B)	Bukit Batok
(C)	Bukit Gombak
(D)	Choa Chu Kang
(E)	Yew Tee
(F)	Kranji
(G)	Marsling
(H)	Woodlands
(I)	Admiralty

(J)	Sembawang
(K)	Yishun
(L)	Khatib
(M)	Yio Chu Kang
(N)	Ang Mo Kio
(O)	Bishan
(P)	Braddell
(Q)	Toa Payoh
(R)	Novena

(S)	Newton
(T)	Orchard
(U)	Somerset
(V)	Dhoby Ghaut
(W)	City Hall
(X)	Raffles Place
(Y)	Marina Bay

(F) Transition(al) Numbers (1/3) [10 points]

The following diagram represents a “transition network” (also known as a “finite state automaton”). The circles represent “states” while the boxes represent letter sequence that can be “generated” from any given state, as indicated by the lines (the “transitions”). The aim is to start at “S” and get to the end state “0”. For some boxes there is a choice of transition. The lines are directional (it would have been even more messy to add the arrowheads), so note that you can only enter a state from the front (left). So you can go from “ty” to “S”, but not the other direction.



The above diagram is already quite messy, and it can be represented more neatly by a set of rules as on the next page. Each rule is identified (in square brackets) but this is ONLY for ease of reference in answering the questions. Apart from that, each rule consists of a state (the symbol before the “:”), a text string, and then, after the arrow (“->”), a list of states to which you can then move. Starting at position “S”, you generate the text indicated, and then continue to any ONE of the states listed after the arrow. State “0” is a special case meaning “finish”.

(F) Transition(al) Numbers (2/3)

[a]	S: one -> 0
[b]	S: two -> 0
[c]	S: three -> 0
[d]	S: four -> 0,1
[e]	S: five -> 0
[f]	S: six -> 0,1,2
[g]	S: seven -> 0,1,2
[h]	S: eight -> 0,1,2
[i]	S: nine -> 0,1,2
[j]	S: ten -> 0
[k]	S: eleven -> 0
[l]	S: twelve -> 0
[m]	S: thir -> 1,2
[n]	S: fif -> 1,2
[o]	S: twen -> 2
[p]	S: for -> 2
[q]	1: teen -> 0
[r]	2: ty -> S,0

So for example, we can generate “fourteen” by taking rule [d] then rule [q]. We cannot generate “twelveteen” because rule [l] only allows one way to progress, namely to finish.

Answer the following questions in the Answer Sheets.

F1. Write out the sequence of rules and states followed to generate the following words. For example, for “fourteen,” you would write “d 1 q 0”.

a. sixteen
b. ninetythree
c. twentyeight
d. fifteen

F2. The network above “overgenerates”, that is, it allows us to create words which are not valid numbers. For each of the following words, write Y if the word can be generated by the network, or write N if the word cannot be generated by the network.

a. oneten
b. fiftytwelve
c. sixteensix
d. twentyfourteen
e. fortythirty
f. eleventythree
g. fivety

(F) Transition(al) Numbers (3/3)

F3. The above network currently generates a misspelling in the case of "eighteen" as well as any number beginning with "eighty." This can be fixed by removing rule [h] and replacing it with two new rules (both of which will be similar to rule [h]). In your answer sheet, write the two new rules that need to replace rule [h]. NOTE: For full points, make sure that the modified network still generates "eight" without misspelling it. Also, you do not need to use up all of the boxes in the answer sheet.

(G) Magik Yup'ik (1/1) [15 points]

Central Alaskan Yup'ik belongs to the Eskimo-Aleut language family. It is spoken in western and southwestern Alaska by around 20,000 speakers. Two other Yup'ik languages are still spoken: the Alutiiq language and the Siberian Yup'ik language.

Yup'ik people have an interesting concept when it comes to counting – the words for the numbers can be broken down into meaningful parts which may be related to their body parts. For example, the word for five, *talliman*, means *an arm* and the word for six, *arvinlegen*, means *cross over*, as you need to change hand to go on counting.

The Yup'ik people often include geometry in their Yup'ik parkas, often having border patterns. One such pattern comes in the form of a 3 by 3 square:

1 (a)	2 9	3 (b)
2 (c)	(d)	(e)
3 (f)	(g)	(h)

A magic square can be constructed by placing the digits 1 to 9 within the cells such that the sum of all the digits in every row, column, and diagonal is the same.

To help you fill in the magic square, the following clues are given in Yup'ik. The numbers are spelled (i.e. 123 is spelled as "One hundred and twenty-three" in English). HINT: The Yup'ik name for the number 294 is yuinaat qula cetaman qula cetaman.

Across	
1	Yuinaat yuinaq cetaman qula malruk
2	Yuinaat akimiaq malruk akimiaq malruk
3	Yuinaat yuinaak malruk akimiaq atauciq

Down	
1	Yuinaat yuinaq atauciq akimiaq pingayun
2	Yuinaat yuinaak malruk yuinaat malrunglegen qula atauciq
3	Yuinaat qula pingayun akimiaq atauciq

Answer these questions in the Answer Sheets.

G1. Fill in the numbers missing from the magic square above.

G2. In Yup'ik, write the number given in 1-Diagonal (Top left cell to Bottom right cell, shaded).

(H) Nothing But Net(works) (1/3) [15 points]

You have just crashed your spaceship at the Viterbi Spaceport. Being unfamiliar with spaceship repair, you're very much at a loss--but then a friendly-looking being from Rigel sidles up to you and says:

“ζ Ψ δ ξ   ϖ Ξ N ϕ   A ϕ Ω υ   Π Π α Σ”

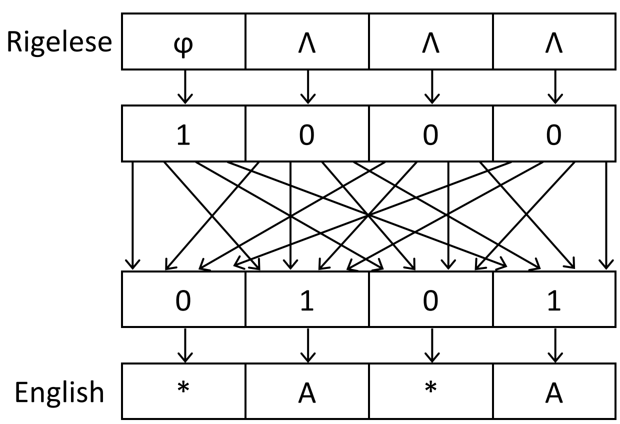
(Okay, maybe that's not so helpful after all.)

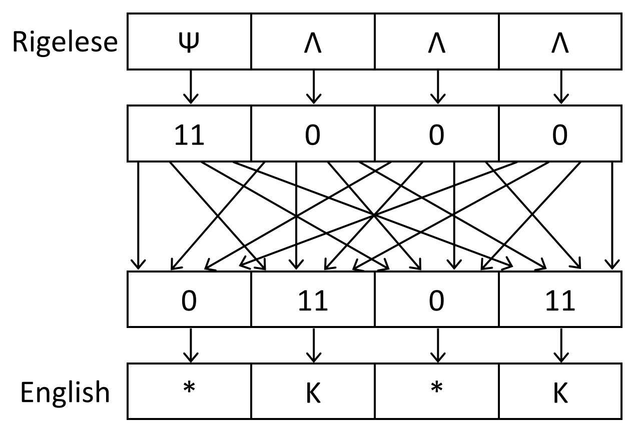
Luckily for you, English and Rigelese are related languages, and you own a GalactiLang translation device that can translate from the Rigelese sound system into more familiar English. This translator first turns the Rigelese word into a sequence of 4 numbers, then uses a neural network to transform those 4 numbers in some way (more about this in a minute), and then it transforms those final numbers into English letters using the following table:

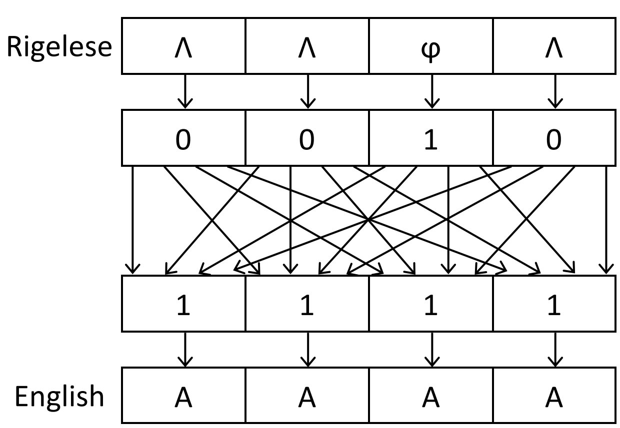
0	1	2	3	4	5	6	7	8	9	10	11	12	13
*	A	B	C	D	E	F	G	H	I	J	K	L	M

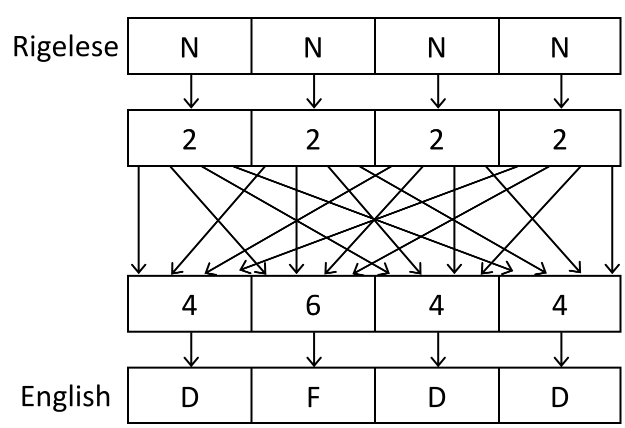
14	15	16	17	18	19	20	21	22	23	24	25	26
N	O	P	Q	R	S	T	U	V	W	X	Y	Z

Here are a few examples of the translator in action:

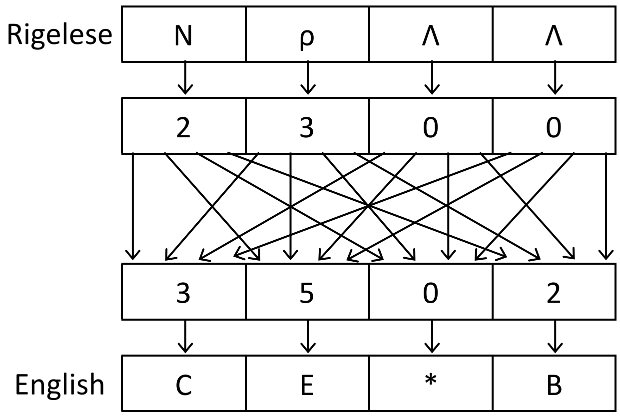


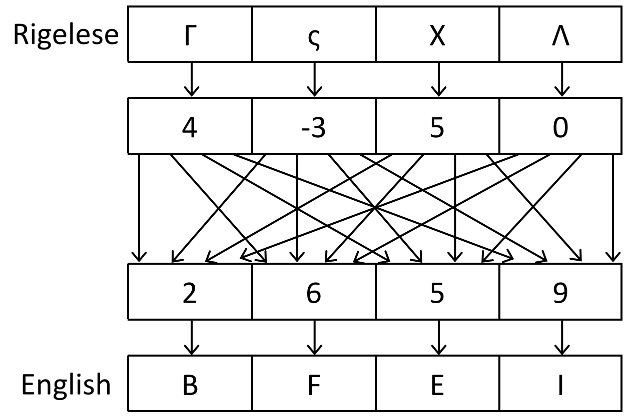




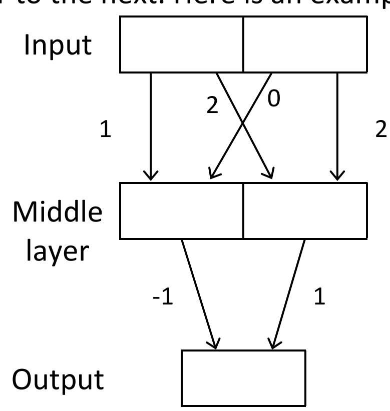


(H) Nothing But Net(works) (2/3)

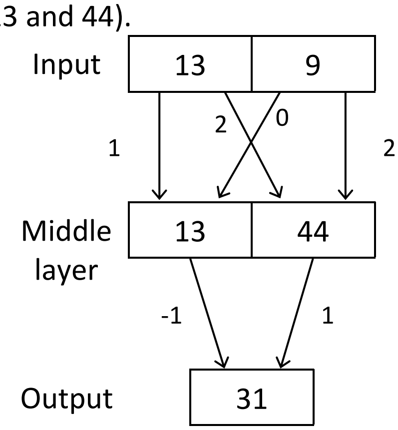




"But wait," you ask, "what's that big jumble of arrows in the middle of each translation?" To which we respond: The jumble of arrows stands for a neural network, which is an abstract computational structure that can be used to approximate any function. The network consists of several layers, including an input layer (the data to be processed), an output layer (the result of the computation), and potentially some middle layers in between the input and output layers. The network is trained on real data, and from this training process it learns how to transition from one layer to the next. Here is an example of a neural network:



This network takes two numbers as its input, then transitions from those two numbers to another two numbers in the middle layer, and then those two middle numbers get turned into a single output. The transitions between the layers are governed by the numbers written next to the arrows (these numbers are called weights). Here is an example of this network in action: Given the inputs 13 and 9, it yields the output 31 (after computing the middle layer of 13 and 44).

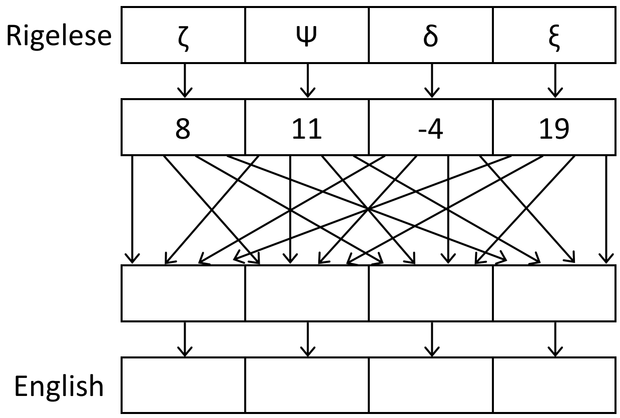


(H) Nothing But Net(works) (3/3)

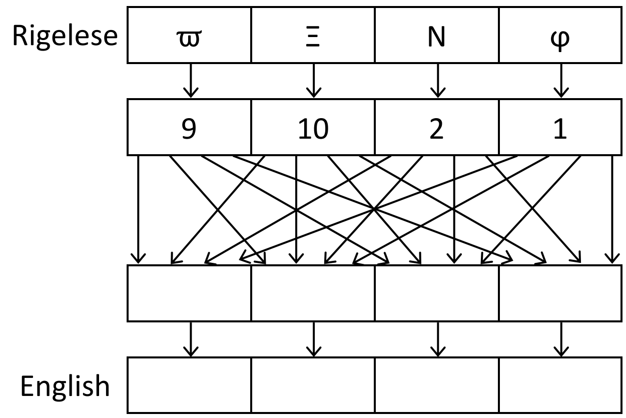
It is left to you to figure out exactly how the transitions are computed. In this case, if we call the inputs *a* and *b*, the output can be easily represented as *a + 2b*. However, neural networks can also represent many other more complex calculations that cannot be as easily expressed otherwise, and these other calculations have proven to be extremely useful in computational linguistic applications.

Now, returning to the Rigel example: When you try to translate the message from the Rigelian, your translator runs out of power after only computing one step of the translation. As a result, this is all that it gives you (each diagram represents the translation process for a single word):

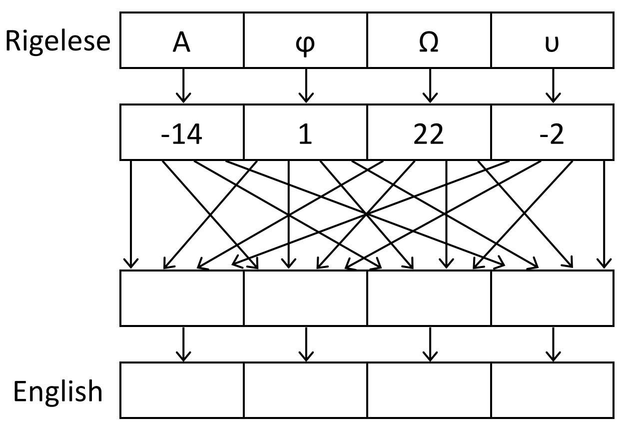
Word 1



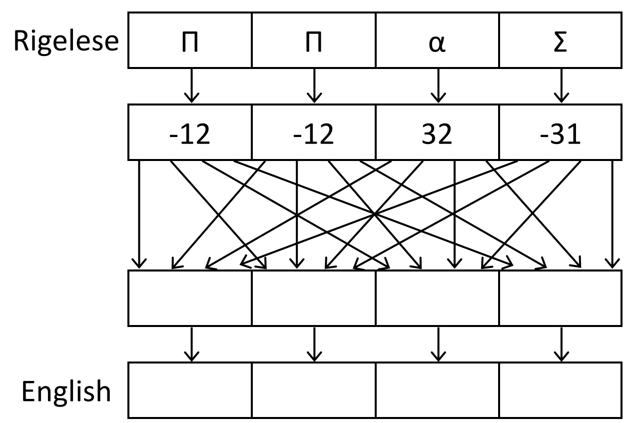
Word 2



Word 3



Word 4



H1. Finish the translation that the translator started. Write your answers in the Answer Sheets. Although you can see the six example translations at the start of this problem, you do not know what weights are attached to the arrows in the diagram (although you do know that the weights are the same across the translations for all four words). Therefore, you will have to use those diagrams to figure out the exact inner workings of the translator.

The North American Computational Linguistics Olympiad
www.nacloweb.org

Contest Booklet

REGISTRATION NUMBER			

Name: ____

Contest Site: ____

Site ID: ____

City, State: ____

Grade: ______

Start Time: ____
End Time: ____

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

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

Signature: ____

YOUR NAME: REGISTRATION #

Answer Sheet (1/3)

(A) A Little Tshiluba

1. a.

   b.

   c.

   d.

   e.

2.

3. a.

   b.

(B) Phở Bar

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

9. 10. 11. 12. 13. 14. 15. 16.

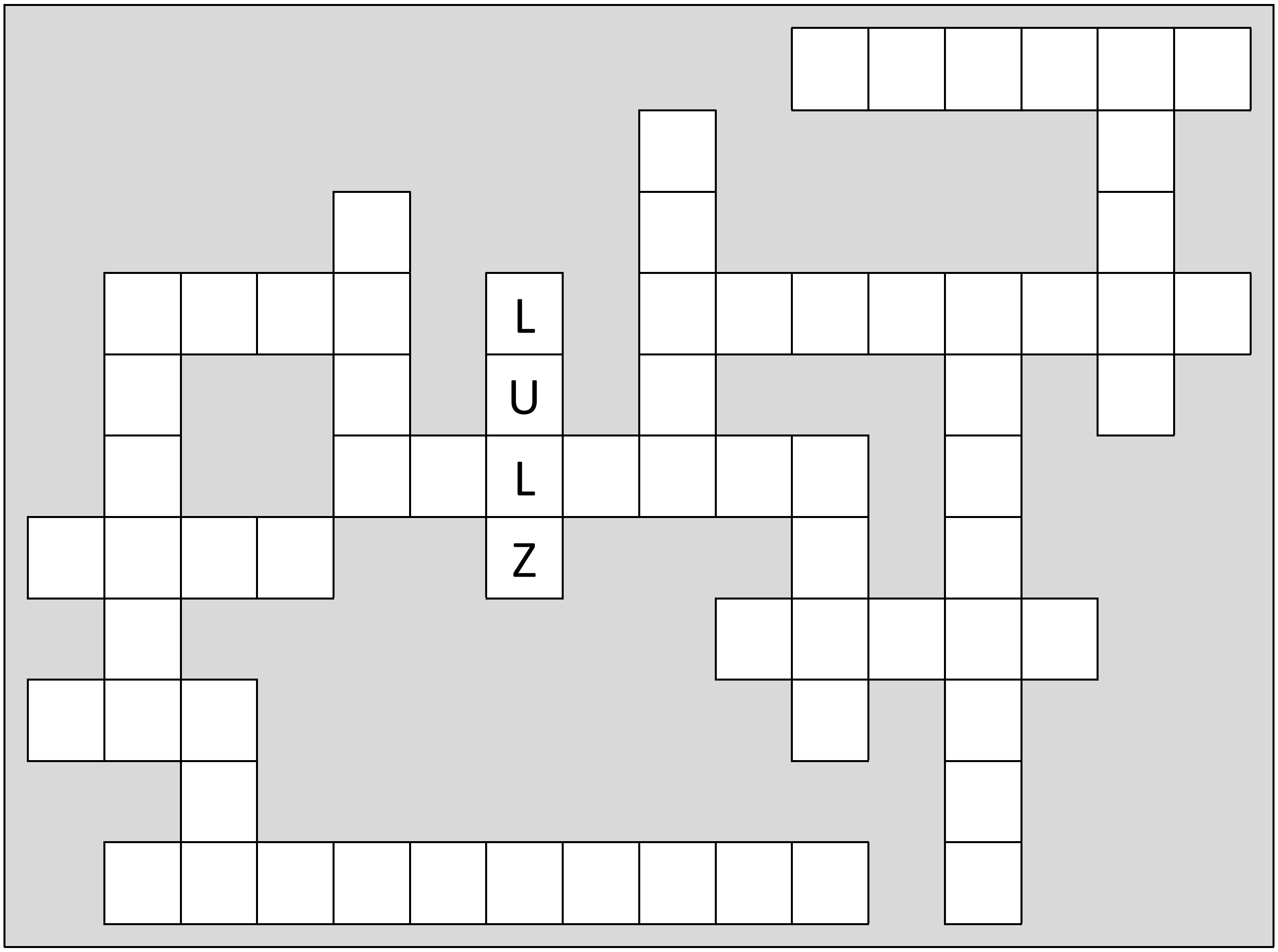
17. 18. 19. 20.

YOUR NAME: REGISTRATION #

Answer Sheet (2/3)

(C) LOLWUT

1.



2. a.

b.

(D) Let's Roll!

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

2.

3.

4.

YOUR NAME: REGISTRATION #

Answer Sheet (3/3)

(E) On the Right Track

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

9. 10. 11. 12. 13. 14. 15. 16.

17. 18. 19. 20. 21. 22. 23. 24.

25.

(F) Transition(al) Numbers

1. a.

b.

c.

d.

2. a. b. c. d. e. f. g.

3. : ->

: ->

(G) Magik Yup'ik

1. a. b. c. d. e. f. g. h.

2.

(H) Nothing But Net(works)

1. *Word 1*

Word 2

Word 3

Word 4