n a c l o

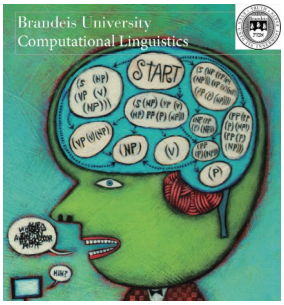
Linguistic Society of America

NAACL

Carnegie Mellon University
Language Technologies Institute

Massachusetts Institute of Technology

MIDDLE TENNESSEE STATE UNIVERSITY



NATIONAL SCIENCE FOUNDATION

Yale University
LUX ET VERITAS

The Twelfth Annual

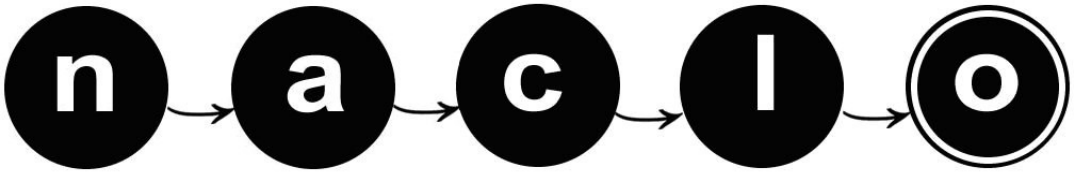
North American Computational Linguistics Olympiad 2018

www.nacloweb.org

Invitational Round March 8, 2018

Serious language puzzles that are surprisingly fun!

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



Welcome to the twelfth 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 team practices before the international competition in the Czech Republic.
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 2018 Organizers

Program Committee:

Adam Hesterberg, Massachusetts Institute of Technology
Alan Chang, University of Chicago
Aleka Blackwell, Middle Tennessee State University
Ali Sharman, University of Michigan
Andrés Salanova, University of Ottawa
Andrew Lamont, University of Massachusetts
Annie Zhu
Babette Newsome, Aquinas College
Daniel Lovsted, McGill University
David Mortensen, Carnegie Mellon University
David Palfreyman, Zayed University
Dick Hudson, University College London
Dorottya Demszky, Stanford University
Dragomir Radev, Yale University
Egor Tsedryk, St. Mary's University
Elisabeth Mayer, Australian National University
Elysia Warner, University of Cambridge
Harold Somers, AILO
Harry Go, Washington University in Saint Louis
Heather Newell, Université du Québec à Montréal
James Pustejovsky, Brandeis University
James Hyett, University of Toronto
Jane Li, Simon Fraser University
Jason Eisner, Johns Hopkins
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 Pennsylvania
Patrick Littell, University of British Columbia
Sam Ahmed, University of Cambridge
Sonia Reilly, Massachusetts Institute of Technology
Susan Barry, Manchester Metropolitan University
Tom McCoy, Johns Hopkins University
Tom Roberts, University of California, Santa Cruz
Verna Rieschild, Macquarie University
Vlado Keselj, Dalhousie University

NACLO 2018 Organizers (cont'd)

Problem Credits:

Round 2:

Problem I: Aleka Blackwell
Problem J: Babette Newsome
Problem K: Ali Sharman and Tom McCoy
Problem L: Ali Sharman
Problem M: Heather Newell
Problem N: Vica Papp
Problem O: Babette Newsome
Problem P: Harold Somers and Dick Hudson
Problem Q: Tom McCoy
Problem R: Tom McCoy

Organizing Committee:

Adam Hesterberg, Massachusetts Institute of Technology
Aleka Blackwell, Middle Tennessee State University
Ali Sharman, University of Michigan
Andrew Lamont, University of Massachusetts
Annie Zhu
Daniel Lovsted, McGill University
David Mortensen, Carnegie Mellon University
Deven Lahoti, Massachusetts Institute of Technology
Dorottya Demszky, Stanford 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
Jordan Ho, University of Toronto
Josh Falk, University of Chicago
Julia Buffinton, University of Maryland
Lori Levin, Carnegie Mellon University
Margarita Misirpashayeva, Massachusetts Institute of Technology
Matthew Gardner, Carnegie Mellon University
Patrick Littell, Carnegie Mellon University
Simon Huang, University of Waterloo
Stella Lau, University of Cambridge
Tom McCoy, Johns Hopkins University
Tom Roberts, University of California, Santa Cruz
Wesley Jones, University of Chicago
Yilu Zhou, Fordham University

NACLO 2018 Organizers (cont'd)

US Team Coaches:

Aleka Blackwell, Middle Tennessee State University
Dragomir Radev, Yale University
Lori Levin, Carnegie Mellon University

Canadian Coordinators and Team Coaches:

Daniel Lovsted, McGill University
Patrick Littell, Carnegie Mellon University

USA Contest Site Coordinators:

Brandeis University: James Pustejovsky, Nikhil Krishnaswamy, Sarah Irwin
Brigham Young University: Deryle Lonsdale
California State University, Long Beach: Michael Ahland
Carnegie Mellon University: David Mortensen, Lori Levin, Mary Jo Bensasi
College of William and Mary: Dan Parker
Columbia University: Brianne Cortese, Kathy McKeown, Smaranda Muresan
Cornell University: Abby Cohn, Sam Tilsen
Emory University: Jinho Choi, Phillip Wolff
Georgetown University: Emma Manning
Indiana University: Melinda Bristow-Meadows, Rachel Hertz, Stephanie Klausing
Johns Hopkins University: Rebecca Knowles
Keiser University, Jacksonville: Deborah Williams
Middle Tennessee State University: Aleka Blackwell
Minnesota State University Mankato: Dean Kelley, Louise Chan, Rebecca Bates
Montclair State University: Anna Feldman
Northeastern Illinois University: J. P. Boyle, Judith Kaplan-Weinger, K. Konopka, R. Hallett
Oregon State University: Liang Huang
Princeton University: Christiane Fellbaum, Misha Khodak
Saginaw Valley State University: Natalia Knoblock
San Diego State University: Rob Malouf
Southern Illinois University: Jeffrey Punske, Vicki Carstens
SpringLight Education Institute: Sherry Wang
Stanford University: Masoud Jasbi
Stony Brook University: Jeffrey Heinz, Lori Repetti, Sarena Romano
Union College: Kristina Striegnitz, Nick Webb
University of California, Irvine: Kristen Salsbury, Sameer Singh, Zhengli Zhao
University of Colorado at Boulder: Silva Chang
University of Houston: Deepti Bhat, Giovanni Molina, Thamar Solorio
University of Illinois at Urbana-Champaign: Benjamin Leff, Greg Sigalov, Julia Hockenmaier
University of Kentucky: Andrew Byrd, Kevin McGowan
University of Maryland: Aaron Doliana, Jan Michalowski, Kasia Hitczenko, Laurel Perkins, Yogarshi Vyas
University of Massachusetts, Amherst: Anastasia Chobany, Andrew Lamont, UMass Linguistics Club

NACLO 2018 Organizers (cont'd)

USA Contest Site Coordinators (continued):
University of Massachusetts, Lowell: Anna Rumshisky, David Donahue, Willie Boag
University of Memphis: Vasile Rus
University of Michigan: Marcus Berger, Sally Thomason, Steven Abney
University of Nebraska, Omaha: Parvathi Chundi
University of North Carolina, Charlotte: Hossein Hematialam, Kodzo Wegba, Wlodek Zadrozny
University of North Texas: Genene Murphy, Rodney Nielsen
University of Notre Dame: David Chiang
University of Pennsylvania: Anne Cocos, Cheryl Hickey, Chris Callison-Burch, Derry Wijaya, Marianna Apidianaki, Mitch Marcus, Oliver Sayeed
University of Southern California, Information Sciences Institute: Jonathan Gordon, Nima Pourdamghani
University of Texas at Dallas: Isaac Persing, Jing Lu, Vincent Ng
University of Utah: Brendan Terry, Jessica Larsen, Joselyn Rodriguez, Justin Nistler
University of Washington: Jim Hoard, Joyce Parvi
University of Wisconsin, Milwaukee: Anne Pycha, Gabriella Pinter, Hanyon Park, Joyce Boyland
Western Washington University: Kristin Denham
Yale University: Raffaella Zanuttini

Canada Contest Site Coordinators:
Dalhousie University: Armin Sajadi, Dijana Kosmajac, Magdalena Jankowska, Vlado Keselj
McGill University: Lisa Travis, Michael Wagner
Medicine Hat College: Jeffrey Klassen
Simon Fraser University: John Alderete, Keir Moulton, Maite Taboada, Marion Caldecott
University of Alberta: Herbert Colston, Sally Rice
University of British Columbia: Hotze Rullmann, Jozina Vander Klok
University of Calgary: Dennis Storoshenko
University of Ottawa: Gustavo Beritognolo
University of Toronto: James Hyett, Lola Bradford, Minh-Tam Nguyen
University of Western Ontario: Janis Chang

High school sites:
Dragomir Radev

Booklet Editors:
Patrick Littell, Carnegie Mellon University
Tom McCoy, Johns Hopkins University
Dragomir Radev, Yale University
Ali Sharman, University of Michigan

Sponsorship Chair:
James Pustejovsky, Brandeis University

NACLO 2018 Organizers (cont'd)

Sustaining Donors

Linguistic Society of America
North American Chapter of the Association for Computational Linguistics
National Science Foundation

University Donors

Brandeis University
Carnegie Mellon University
Massachusetts Institute of Technology
Middle Tennessee State University
Yale University

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 200+ High School Sites

NACLO 2018 Sites

As well as more than 200 high schools throughout the USA and Canada

(I) A Menya Puzzle (1/2) [15 points]

Menya is a Papuan language spoken in the Morobe Province of Papua New Guinea. It has the second most speakers of the at least twelve languages that constitute the Angan language family. The data in this problem are presented in the writing system of the language.

Answer the following questions in the Answer Sheets:

I1. Each Menya word or phrase in the column on the left below has its English translation somewhere in the column on the right, but these English translations are in scrambled order. On the Answer Sheet, match each Menya entry to its English translation.

	Menya
1.	ɑi
2.	täŋgɑ
3.	yä nɑqänäŋä
4.	ymeqä wäŋqä
5.	moni nɑqäŋganji
6.	ämɑqä nɑqä
7.	yämbuɑyä
8.	ymeqä qokä
9.	äkewi yŋŋä nɑqä hmɑnji
10.	ɑiŋgɑ
11.	yä ɑŋä
12.	buɑyä
13.	ämɑqä qokä
14.	tä
15.	i
16.	tä sipqäti botqä äwitäti
17.	i täqueqä äŋi?

	English
A.	a very large tree
B.	an important person
C.	That is whose house?
D.	The äkewä is not a large bird.
E.	a Cassava plant
F.	long ago
G.	done
H.	a man
I.	I wonder if this is a ship or a boat.
J.	that
K.	a sweet potato
L.	this
M.	now
N.	a small child
O.	a wooden house
P.	a son
Q.	Fines are big these days.

A Cassava plant

I2. Translate the words in the tables on the next page into the language indicated at the top of each table. Note that *hikŋä* means ‘lad’ or ‘young man’.

(I) A Menya Puzzle (2/2)

Translate into Menya:	
a.	‘large’
b.	‘a small piece of wood’ or ‘a small stick’
c.	‘the boat’
d.	‘a very small bird’

Translate into English:	
e.	ɑŋä nɑqänäŋä
f.	iŋgɑ
g.	hikŋäŋgɑ

I3. Within one of the multiword Menya phrases in the data is a single word typically used by Menya speakers to mean ‘a husband.’ Which word is it?

I4. Explain your answer by briefly describing how Menya words and phrases are structured.

(J) It's True: The Truth About Chalcatongo Mixtec (1/2) [10 points]

Chalcatongo Mixtec is a language spoken by just under 6,000 people in Oaxaca State of South-Central Mexico. It is famous among linguists for its many unusual characteristics, but it is an endangered language, at risk of extinction.

Answer the following questions in the Answer Sheets:

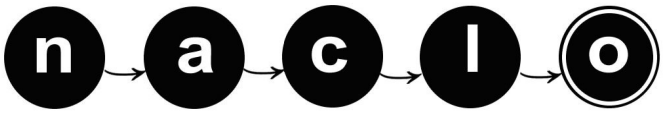
J1. Here are some sentences in Chalcatongo Mixtec. Their English translations have been given in scrambled order on the right. Match the Chalcatongo Mixtec sentences to their English translations. *Note:* Chalcatongo Mixtec (when written in the Roman alphabet) has some letter symbols and accents that we do not use in English—these are not relevant for solving this problem. Whenever a word or phrase is bold and underlined (**<u>like this</u>**), it means that the word or phrase is emphatic, but otherwise it is not emphatic. Finally, here is a hint to help you get started: Chalcatongo sentence (1) corresponds to English choice (a).

	Chalcatongo sentences
1.	Nduča kaa ñíʔní.
2.	Maria kúu ɨɨ xasúčí.
3.	Ñíʔní nduča.
4.	Juan kaa lúlí.
5.	Ndežu kaa žaʔu.
6.	Súčí Maria.
7.	Juan kúu xažiirí.
8.	Pedro kúu xalúlírí.
9.	Kaa kwíí.

	English translations
a.	**<u>The water</u>** is hot.
b.	Pedro is my child.
c.	Juan is my husband.
d.	**<u>It</u>** is green.
e.	Maria is a young person.
f.	The water is hot.
g.	Maria is young.
h.	**<u>The food</u>** is expensive.
i.	**<u>Juan</u>** is small / short.

Here are some words in Chalcatongo Mixtec. The English translation for each word is given on the right.

Chalcatongo	**English translation**
ndáa	true
bíší	sweet
bèè	heavy
tûû	black
ndeʔé	brave
kŭnú	deep
ñíʔní	hot
kwaʔá	red
súkú	tall

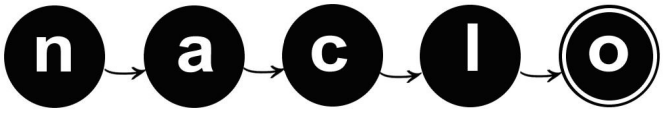


(J) It's True: The Truth About Chalcatongo Mixtec (2/2)

J2. Translate the following into Chalcatongo Mixtec:

a. depth
b. heat
c. Maria is brave.
d. **Pedro** is tall.
e. Pedro is a tall person.
f. The fruit is red.
g. My fruit is the green one.
h. **It** is true.
i. It is true.
j. It is the truth.

J3. Explain your answers by describing how Chalcatongo Mixtec words and sentences are structured.



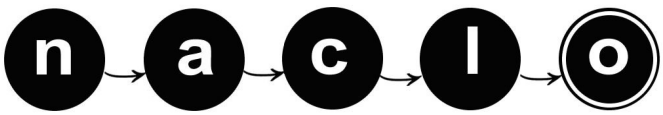
(K) Sri Lankan Names (1/2) [10 points]

Your job at the State Department is to educate people about the named entities (that is, the places and people) in Sri Lanka. From your reference book, you have learned that Sinhala is a language used in Sri Lanka and that it follows a subject-object-verb structure (which would, for example, rearrange the English sentence "I bought a book" to "I a book bought"). You have just arrived in Sri Lanka, where you were greeted by this sign containing the country's name (śrī laṁkā in transliterated Sinhala) written in the script of the Sinhala language:

ශ්‍රී ලංකා

You have also clipped a short snippet from a Sri Lankan newspaper, and have decided to perform a task called named entity recognition on this snippet. To make this task easier, you've added numbers between all of the words in the text. Here is what the snippet looks like with your numbers added:

1 ඔහු **2** කොළඹ **3** දිස්ත්‌රික්කයේ **4** ජීවත්වේ **5** . **6** සමහර **7** විට **8** කෑමට **9** සිරිසේන **10** මහත්මිය **11** ඔහුව **12** මුණගැසේ **13** . **14** ගුණසේන **15** මහතා **16** ඔවුන් **17** සමග **18** ආහාර **19** ගනී **20** . **21** ගිය **22** සතියේ **23** දෙහිවලට **24** පැමිණියා **25** . **26** වීරරත්න **27** මහතා **28** නිතරම **29** ඔහුව **30** එහිදී **31** මුණගැසේ **32** . **33** නමුත් **34** ඔහු **35** තවමත් **36** ශ්‍රී **37** ජයවර්ධනපුර **38** කෝට්ටේ **39** ජීවත්වේ **40** . **41** වීරරත්න **42** මහත්මිය **43** ඇයගේ **44** ඥාතියෙක් **45** සමග **46** නිතරම **47** ආහාර **48** ගනී **49** . **50**



(K) Sri Lankan Names (2/2)

Your task is to identify each named entity in this snippet and to indicate whether it is a person or a place. To identify a named entity, you need to indicate the number that appears right before the entity begins and the number that appears right after the entity ends. For example, if you were working with the following English sentence:

$_1$ the $_2$ great $_3$ King $_4$ Arthur $_5$ lived $_6$ in $_7$ Camelot $_8$. $_9$

Then the correct named entity identification would look like this:

Start	End	Type
3	5	Person
7	8	Place

The snippet contains seven named entities, one per sentence. You know that three of them are places: the capital city of Sri Lanka, which is Sri Jayawardenapura Kotte; the Colombo District; and the suburb Dehiwala. You don't know exactly what the names of these places look like in Sinhalese, but you do know that the Sinhalese names, when written in the English alphabet, look as follows: śrī jayavardhanapura kōṭṭēṭa, koḷaṁba distrikkayē, and dehivalaṭa. You can tell from their names that each of the four people are either husbands or wives. One of the men and one of the women are married to each other and share the same family name (vīraratna as written in English letters for both people). Note that no first names are listed and that all last names are one word long.

Answer the following questions in the Answer Sheets:

K1. Fill in the table in the Answer Sheet with the named entity information for the Sinhalese snippet above.

(L) Peeled Potato Act with Annie (1/3) [5 points]

Consider the following sentence:

1) The girl ate rice with shrimp.

There are multiple different things that this sentence could mean. It could mean that the girl ate rice which contained shrimp; with this meaning, the phrase *with shrimp* modifies the noun *rice*. On the other hand, it could instead mean that the girl used shrimp as tools that allowed her to eat the rice; with this meaning, *with shrimp* modifies the verb *ate*. For this sentence, it is obvious to a human reader that the intended meaning is the one where *with shrimp* modifies the noun rather than the verb, because it is common for shrimp to appear in rice while it is uncommon for people to use shrimp as eating implements. However, this changes if we change the word *shrimp* to *chopsticks*:

2) The girl ate rice with chopsticks.

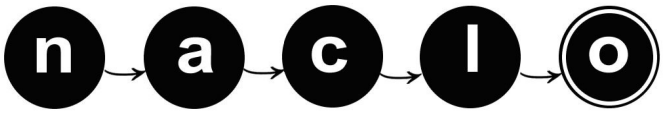
Now it is much more likely that *with chopsticks* modifies the verb *ate* rather than the noun *rice* because chopsticks are usually eating implements rather than ingredients.

With chopsticks and *with shrimp* are examples of prepositional phrases, which modify other phrases to provide more information about those phrases. Prepositional phrases always start with a preposition, such as *to*, *for*, *of*, or *with*. In ambiguous sentences (such as 1 or 2), it is usually easy for humans to tell whether the prepositional phrase is intended to modify the noun or the verb, but for computers this task can be much harder because the computers might not have all of the background knowledge that humans use to make this judgment. In the following article about Annie, a circus performer, each italicized component contains a prepositional phrase that could potentially be describing a noun or a verb:

> Annie [1] *joined the circus as tightrope walker*, but she [2] *is now head of the acrobats*. She [3] *performs her main act in a purple leotard*. Annie best [4] *likes the routine with the trapeze*, because she [5] *had liked that act as a child*. She usually [6] *performs this act with gusto*, but yesterday she [7] *performed this act with sadness*. She [8] *had just lost her lucky penny in the street*. But, at least she [9] *likes her new good luck charm with the inscription.* Because he was sick, Annie [10] *practiced her routine with Charley* alone yesterday. While Charley is an acrobat, he usually [11] *performs his main act with the clowns*. Charley [12] *likes the act with the bananas*, but Charley best [13] *likes the one with the peach pie*. Tonight, Charley gets to [14] *perform his main act with Annie*. Unfortunately, the peach pie act is not very popular with the audience, so they have to [15] *do the act with the peeled potatoes*.

Answer the following questions in the Answer Sheets:

L1. The table on the next page lists each instance of an ambiguously positioned prepositional phrase in the article about Annie. Some instances are already correctly labeled as either modifying the noun (N) or the verb (V). Fill in the remaining labels in the "Correct Label" column (on your Answer Sheet) with the most likely label given the context of the instance. Note that adjectives within the sentences are not listed in the table.



(L) Peeled Potato Act with Annie (2/3)

Instance	Preposition	Verb	Noun1	Noun2	Correct label
[1]	as	joined	circus	walker	
[2]	of	is	head	acrobats	N
[3]	in	performs	act	leotard	
[4]	with	likes	routine	trapeze	
[5]	as	liked	act	child	
[6]	with	performs	act	gusto	
[7]	with	perform	act	sadness	
[8]	in	lost	penny	street	
[9]	with	likes	charm	inscription	
[10]	with	practiced	routine	Charley	
[11]	with	performs	act	clowns	
[12]	with	likes	act	bananas	
[13]	with	likes	one	pie	
[14]	with	perform	act	Annie	V
[15]	with	do	act	potatoes	N

L2. A computer would struggle to predict the correct labels for these instances, but you can help it by giving it some rules to guide its decisions. These rules will be interpreted in a specified order such that, once an instance has been labeled by one rule, its label cannot be changed by any later rules, even if those rules would apply if it had not yet been labeled.

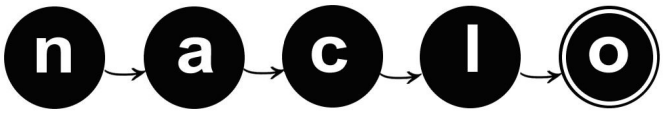
You are considering using the following rules:

A. If the Verb is a form of "to like" (e.g., likes, like, liked), label the instance as N.
B. If the Preposition is "with" and Noun2 contains a food item, label the instance as N.
C. If Noun1 is "act", label the instance as V.

For Algorithm1, you order the rules as follows:

1. __A__
2. __B__
3. __C__

On your answer sheet, fill in the label that Algorithm1 would assign to each instance. If the algorithm does not produce a label for an instance, leave that cell blank.

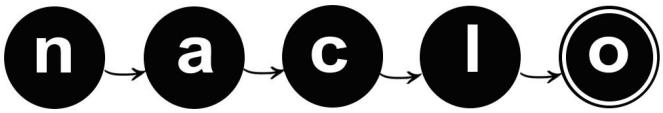


(L) Peeled Potato Act with Annie (3/3)

L3. A baseline is a starting point (the default predictions) that can be used to determine whether an algorithm improves labelling prediction. Because it is a default, the baseline consists of only one rule and applies to all instances (ie, there will be no blank labels). If you were a computer scientist developing Algorithm1, before testing Algorithm1, you should have already made a baseline algorithm for the article about Annie, and you found that Algorithm1 gets 10 labels right, gets 1 label wrong, and leaves 4 blank. The baseline gets 8 right and 7 wrong, leaving 0 blank. On your Answer Sheet, state a single rule that could have been the rule you used as your baseline. Then, next to the number of each instance, fill in the label for that instance given by your baseline.

L4. What order should the rules A, B, and C be placed in an algorithm in order to obtain the highest accuracy possible? Write the letter of the rule next to the number of the order that it is placed in on your Answer Sheets.

L5. Explain how you chose the baseline in L3 and why you ordered the rules the way that you did in L4.



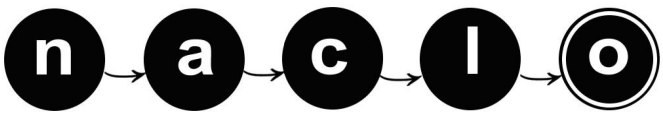
(M) Quests and Requests in Nivkh (1/2) [15 points]

Nivkh (also known as Gilyak) is a language isolate, spoken by approximately 1,000 people in Outer Manchuria. In the following problem, Balda needs supplies quickly because he must leave for a long journey immediately, where he must accomplish many quests. If he doesn't have all of these items, he will not be able to complete the quests he has been assigned. Figure out how to say what Balda is asking for so that he can get going in time! Note that p', t', and k' are single sounds in Nivkh and the clusters fx and sx are not permitted in the language.

Answer the following questions in the Answer Sheets:
M1. Match the following statements to their English translations:

1.	ñi ʂoxekiḍ
2.	ñi həhaqxekiḍ
3.	ñi itəfk'ekiḍ
4.	ñi ñeflaŋkxekiḍ
5.	ñi ñaqxekiḍ
6.	ñi p'ezŋaixekiḍ
7.	ñi ʈzɣəfk'ekiḍ
8.	ñi təpilaʂoxekiḍ
9.	ñi həzaqoxekiḍ
10.	ñi liɣsk'ekiḍ
11.	ñi təfk'ekiḍ
12.	ñi ʂaqxekiḍ
13.	ñi ñivrəxekiḍ
14.	ñi ñmuxekiḍ
15.	ñi ʈzŋaixekiḍ
16.	ñi aʂoxekiḍ
17.	ñi ñit'ulvhaqxekiḍ
18.	ñi ʈzaqoxekiḍ
19.	ñi ʈpilazɣəfk'ekiḍ
20.	ñi amuxekiḍ
21.	ñi haqxekiḍ
22.	ñi ʈeflaŋkxekiḍ
23.	ñi ʈliɣsk'ekiḍ

A.	I need this big fish
B.	I need my cap
C.	I need your branch
D.	I need my hut
E.	I need my winter cap
F.	I need my boat
G.	I need a fish
H.	I need your big bear
I.	I need a house
J.	I need your wolf
K.	I need my branch
L.	I need a cap
M.	I need my own picture
N.	I need a wolf
O.	I need that boat over there
P.	I need your bear
Q.	I need that cap
R.	I need that knife
S.	I need that fish over there
T.	I need his/her house
U.	I need your knife
V.	I need your picture
W.	I need your cap



(M) Quests and Requests in Nivkh (2/2)

The following are some other nouns:

həjk	hare
ʇif	track
ʇus	meat
fləŋg	ash
zrovs	nail

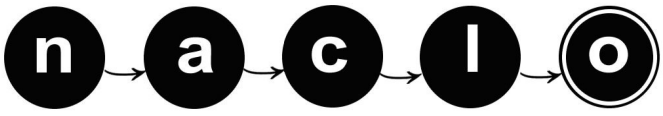
M2. Translate the following into English:

a.	ñi ʇpilamuxekiḍ
b.	ñi həzɣəfk’ekiḍ
c.	ñi təfləŋgxekiḍ
d.	ñi p’eʇusk’ekiḍ
e.	ñi ahaqxekiḍ

M3. Translate the following into Nivkh:

a.	I need your big track.
b.	I need my knife.
c.	I need that cap over there.
d.	I need your hare.
e.	I need my nail.

M4. Explain your answers by describing the structure of Nivkh words and sentences.



(N) You've Got This: Fijian Ownership (1/2) [10 points]

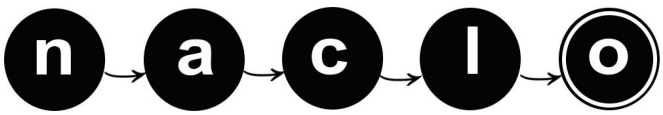
Below you can see a number of phrases in Fijian orthography. Note that 'our (incl.)' means 'belonging to me and you and other people' in contrast to 'our (excl.)', which would mean 'belonging to me and other people, not including you'.

Fijian	English
na uluqu	my head
na nona wau	his weapon[1] (he owns)
na memunī bia	your (pl.) beer
na kemudrau itukutuku	your (dual) story (about you two)
na nona motokaa	her car
na meda tī	our (incl.) tea
na kelemu	your (sing.) belly
na nona dio	her oyster (she'll sell)
na kequ uvi	my yam (edible starchy root)
na noqu itukutuku	my story (that I tell)
na watiqu	my spouse
na kemunī vuaka	your (pl.) pig (you'll eat)
na nomu kato	your (sing.) basket
na tamana	his father
na memudrau dio	their (dual) oyster (you'll slurp)
na nodra vuaka	their pig (they raise)
na keda wau	our (incl.) weapon (we'll be hit with)
na kedra raisi	their rice

[1]A club-like tool.

Answer the following questions on the Answer Sheets:

N1. Now the Fijian words are given for you. Your task is to translate the phrases in the table on the next page into Fijian.



(N) You've Got This: Fijian Ownership (2/2)

	Fijian	English	English phrase to translate
a.	uto	heart	my heart
b.	yaqona	kava[2]	her kava (she's drinking)
c.	draunikau	witchcraft	my witchcraft (used on / against me)
d.	dali	rope	your (sing.) rope (you own)
e.	ika	fish	your (dual) fish (for dinner)
f.	wai	water	your (pl.) water
g.	luve	child	her child
h.	yaqona	kava	his kava (drunk in his honor)
i.	waqa	canoe	our (incl.) canoe
j.	yapolo	apple	their apple (they're selling)
k.	draunikau	witchcraft	your (dual) witchcraft (you're making)
l.	dali	rope	your (pl.) rope (restraining you two)
m.	maqo	mango	their mango (for drinking)

[2]Ceremonial drink widely consumed in the Pacific.

N2.

a. Explain your translation of 'his kava (drunk in his honor)'.

b. The word for 'coconut' is 'niu'. List all the ways to say 'my coconut' and explain what they could mean.

(O) To Know or Not to Know Literary Tamil (1/2) [5 points]

Tamil is spoken in Southern India and Sri Lanka. It has a long history and a large literature dating back over two thousand years. This puzzle uses the literary form of Tamil, which is slightly different from the modern spoken language.

This problem uses the Roman alphabet with some modifications to write Tamil, but this is not relevant for solving the problem.

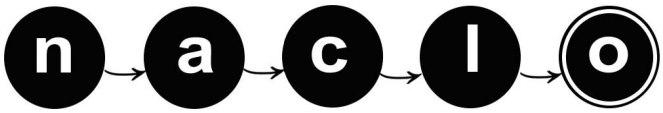
Study the following Tamil verb forms:

Tamil	English
paṭittēn	I learnt
paṭippīr	You (plural) will learn
ceyyān	He doesn't do
paṭiyēn	I don't learn
arippāy	You (singular) will know
paṭittāy	You (singular) learnt
arittān	He knew
aarambippāy	You (singular) will begin
aarambippōm	We will begin
aarambiyēn	I won't begin
paṭittāl	She learnt
ceyppēn	I will do
ceyyēn	I won't do
ceyppār	They will do

Answer the following questions in the Answer Sheets:

O1. Translate into Tamil (continued on the next page):

a. He doesn't begin
b. We will do
c. They didn't know
d. She won't begin
e. We won't do
f. You (plural) won't begin



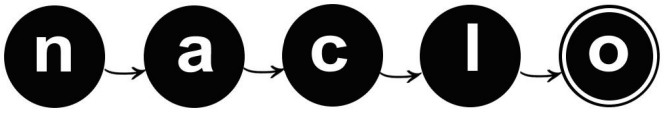
(O) To Know or Not to Know Literary Tamil (2/2)

g. He will learn
h. He didn't learn
i. We began
j. I didn't do
k. We didn't do
l. You (singular) don't know
m. You (singular) knew

O2. Translate the following into English. If there is more than one possible translation, give all the possibilities:

a. arittīr
b. aarambippāl
c. ariyōm
d. paṭippār
e. aarambiyēn
f. ceyyīr
g. paṭittōm

O3. Explain your answers by describing the structure of Tamil words.



(P) I Know What I Saw (1/2) [5 points]

This problem considers data from Beja (Bidhaawyeet), an Afroasiatic language of the Cushitic branch spoken on the western coast of the Red Sea by the Beja people. Its speakers number around 2 million, and inhabit parts of Egypt, Sudan and Eritrea.

Here are some Beja sentences and their translations, but in a random order:

1.	Tak rihan
2.	Yaas rihan
3.	Akra tak rihan
4.	Dabalo yaas rihan
5.	Tak akraab rihan
6.	Tak dabaloob rihan
7.	Tak akteen
8.	Rihane tak akteen
9.	Tak rihaneeb akteen

A.	I saw a man that is strong
B.	I know a man that I saw
C.	I saw a man that is small
D.	I saw a small dog
E.	I saw a strong man
F.	I saw a dog
G.	I saw a man
H.	I know a man

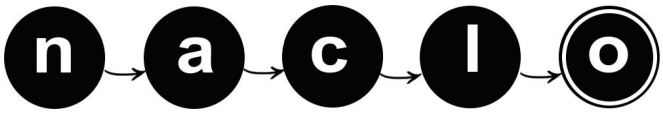
Answer the following questions in the Answer Sheets:
P1. Match up the Beja sentences with their translations. You will use one translation (letter) twice.

Here are some more words from the Beja language with their translations:

araw	friend
mek	donkey
kwati	happy

P2. Translate the following sentences into Beja. If there are different ways to translate the sentence, show all the alternatives.

a. I saw a donkey.
b. I saw a happy man.
c. I saw a friend that is happy.
d. I know a strong donkey.
e. I know a dog that is small.
f. I saw a donkey that I know.

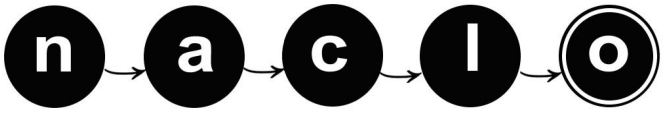


(P) I Know What I Saw (2/2)

P3. Translate the following sentences into English. One of them has a mistake. Write the correct version of this sentence in Beja next to (e) on your answer sheet.

a. Kwati mek rihan.
b. Akraab araw akteen.
c. Akteene yaas rihan.
d. Mek dabaloob akteen.

P4. Explain your answers by describing the structure of Beja words and phrases.



(Q) Better Left Unsaid (1/3) [10 points]

The following mini puzzles deal with five different languages from four different continents. They also deal with various sounds from world languages. In linguistics, sounds are grouped together based on some of their properties. For example, the following chart shows some linguistic sounds:

	Nasal	**Voiced stop**	**Voiceless stop**	**Voiceless fricative**
Bilabial	m	b	p	
Labiodental				f
Alveolar	n	d	t	s
Postalveolar				š
Velar	ŋ	g	k	

So, for example, n, d, t, and s are all alveolar sounds, while b, d, and g are all voiced stops. (The meanings of these linguistic terms are not relevant for this problem).

For this problem, there are five types of sounds that are especially relevant. These are listed in the chart below, along with the sounds used in this problem that fall into each category:

Category	**Sounds in the Category**
Vowels	a, e, i, o, u
Consonants	m, n, ŋ, b, d, g, j, r, l, ǰ, w, y, p, t, k, f, s, ʃ, h, q, č
Nasals	m, n, ŋ
Voiced Consonants	b, d, g, j, r, l, ǰ, w, y, m, n, ŋ
Voiceless Consonants	p, t, k, f, s, š, h, q, č

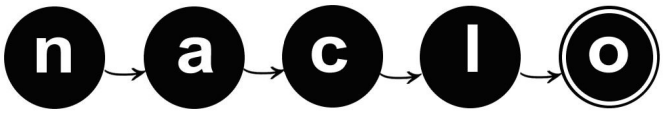
Answer the following questions in the Answer Sheets:

Q1. Indonesian, spoken by roughly 200 million people, is the official language of Indonesia. Below are some Indonesian words with their English translations. In the Answer Sheets, fill in the blanks from the chart.

Indonesian	**English**
meŋuji	to test
diuji	to be tested
meŋeja	to spell
dieja	to be spelled
meŋgaruk	to scratch
digaruk	to be scratched

Indonesian	**English**
mendapat	to obtain
didapat	to be obtained
memberi	to give
diberi	to be given
menulis	to write
ditulis	to be written

Indonesian	**English**
memutus	to cut off
diputus	to be cut off
(a)	to make
dibuat	to be made
(b)	to choose
dipilih	to be chosen



(Q) Better Left Unsaid (2/3)

Q2. Mandar is one of many local languages spoken in Indonesia. It has about 480,000 speakers. Below are some Mandar words with their English translations. In the Answer Sheets, fill in the blanks from the chart.

Mandar	English
mambatta	to split
dibatta	to be split
mandeŋŋeq	to carry on the back
dideŋŋeq	to be carried on the back
maŋidaŋ	to crave
diidaŋ	to be craved
mappasuŋ	to send out
dipasuŋ	to be sent out

Mandar	English
mattunu	to burn
ditunu	to be burnt
massiraq	to tie
disiraq	to be tied
(c)	to throw
ditimbe	to be thrown
(d)	to feed
dipande	to be fed

Q3. The Quechua languages are a group of languages spoken in South America. Below are some words from the variety of Quechua spoken in Puyo Pungo in Eastern Ecuador, along with their English translations. In the Answer Sheets, fill in the blanks from the chart.

Quechua	English
kam	you
kamba	yours
atam	frog
atambi	in the frog
hatum	the big one
(e)	the big one's
(f)	in the big one

Quechua	English
sinik	porcupine
sinikpa	porcupine's
čilispa	streamless region's
čilis	streamless region
sača	jungle
sačapi	in the jungle
punǰa	day
punǰapi	in the daytime

Q4. The Zoque languages are spoken in in southern Mexico. On the next page are some Zoque words with their English translations. In the Answer Sheets, fill in the blanks from the chart.

(Q) Better Left Unsaid (3/3)

Zoque	English
burru	burro
mburru	my burro
pama	clothing
mbama	my clothing
tatah	father
ndatah	my father
faha	belt
faha	my belt
sis	meat
sis	my meat

Zoque	English
flawta	harmonica
(g)	my harmonica
šapun	soap
šapun	my soap
disko	phonograph record
(h)	my phonograph record
kayu	horse
ŋgayu	my horse
kopak	head
(i)	my head

Q5. Below are some words from the language Lunyole, spoken in Uganda, along with their English translations. In the Answer Sheets, fill in the blanks from the chart.

Lunyole	English
oludaalo	day
endaalo	days
oluboyooboyo	hullabaloo
emboyooboyo	hullabaloos
olufudu	rainbow
efudu	rainbows
olukalala	list
ekalala	lists

Lunyole	English
olusosi	mountain
(j)	mountains
olubafu	rib
(k)	ribs
olupagi	spoke (of a bicycle wheel)
(l)	spokes (of a bicycle wheel)
olutambi	candle
(m)	candles

Q6. Tying it all together

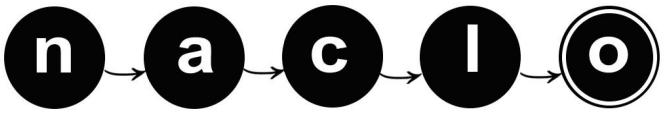
a. All five of the languages in this problem display processes that avoid a specific type of sound combination. In the Answer Sheets, fill in the blanks in the following sentence to describe this generalization:

> Avoid having a _____(n)_____ directly followed by a _____(o)_____.

HINT: The answers for (n) and (o) should be chosen from the following list: Vowel, Consonant, Nasal, Voiced Consonant, Voiceless Consonant.

b. English does not obey the rule from Part 6 (a). **In fact, there is an English word in this bolded sentence that violates this rule.** By space (p) in the Answer Sheets, write the English word from the previous sentence that violates the rule described in the box above in 6a.

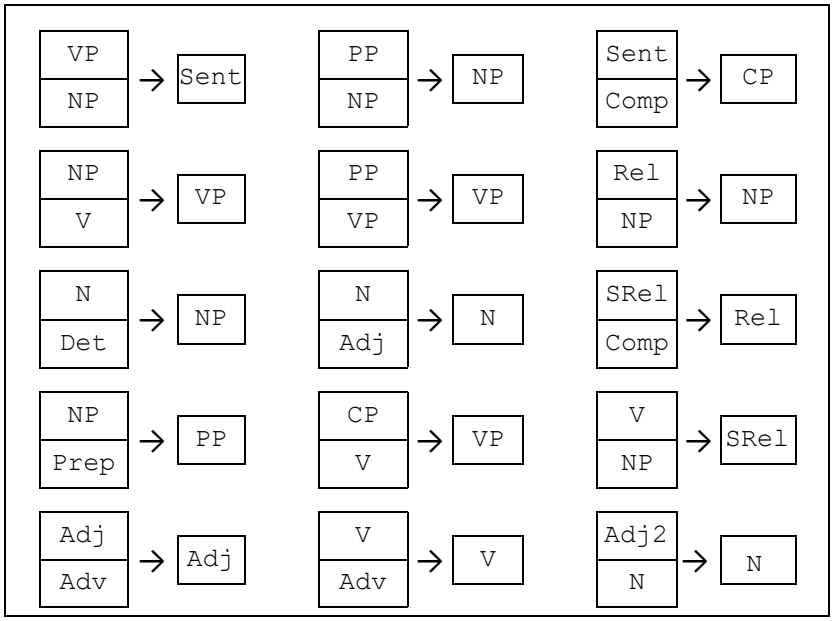
Q7. Explain your answer to (Q6a) by describing how this generalization applies to each of the five languages.



(R) A Make-Shift Code (1/4) [15 points]

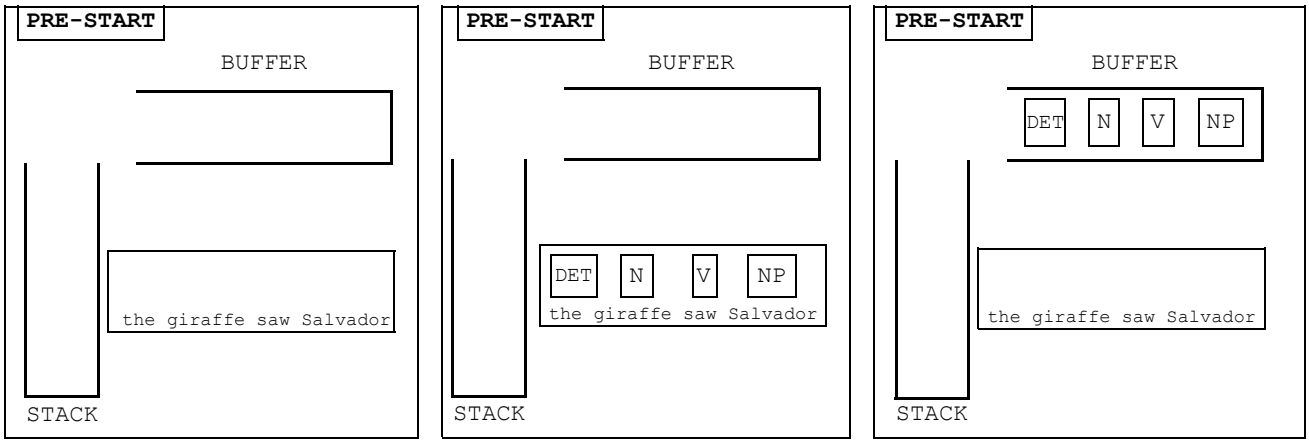
Alice and Bob belong to a secret organization called People Avoiding Really Sinister Eavesdroppers (PARSE). Unfortunately for them, all of their messages are intercepted by an enemy named Eve (who loves to eavesdrop), so Alice and Bob come up with a scheme for encrypting their messages to prevent Eve from reading them.

Their scheme is based on a device called a **shift-reduce parser**. This device makes use of a set of rules (below, left) and a table of part-of-speech tags (below, right):

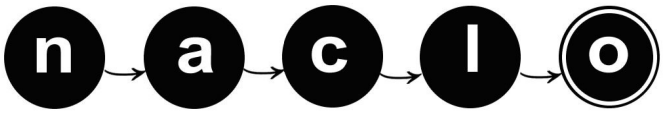


Tag	Words that can have that tag
NP	Salvador
N	giraffe, glass, saw, orange, felt, spots
V	saw, felt, heard
Comp	that
Det	the
Adj	glass, soft, orange
Adj2	extraordinaire
Adv	really
Prep	on
PP	here

The shift-reduce parser has two segments, called the **stack** and the **buffer**, and it takes a series of words as its input. To run the device, you first input the words, assign a tag from the part-of-speech tag table (above, right) to each word, and then move these tags (in order) onto the buffer. If the machine had the input sequence "the giraffe saw Salvador", these pre-processing steps would look like this:



Note that there might be more than one option for a word's tag–for example, "saw" could be N or V.

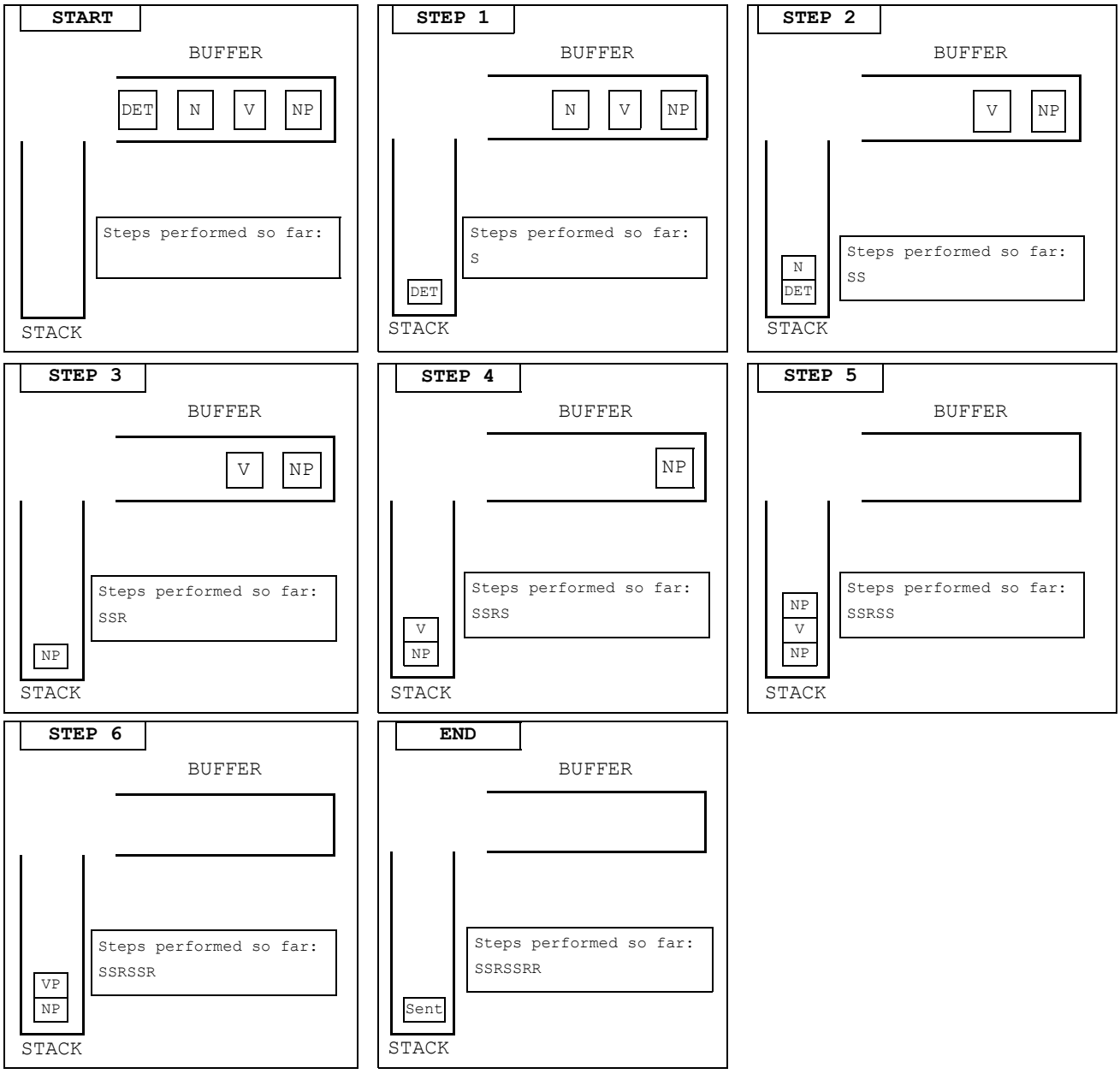


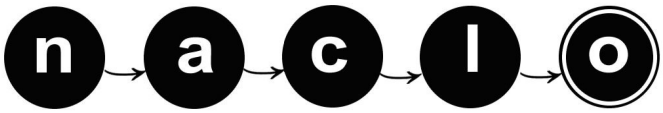
(R) A Make-Shift Code (2/4)

Now the main operation of the machine begins. At each step of its processing, the machine performs one operation, which is either a **shift** operation or a **reduce** operation. These two operations are defined as follows:

- **Shift**: Take the first item in the buffer and place it on top of the stack.
- **Reduce**: Combine the top two items in the stack using one of the rules from the rule set.

If the machine reaches a point where all the words have been removed from the buffer, and the only thing remaining on the stack is Sent, then the machine has created a successful parse for its input sentence. For our example, the shift-reduce processing would look as follows (we will use the abbreviations S for Shift and R for Reduce):





(R) A Make-Shift Code (3/4)

Because the parser ends up with nothing in the buffer and with only Sent in the stack, this was a successful parse of the input sentence. Note that the parser was not guaranteed to succeed in this case--if it had shifted at step 3 instead of reducing, it would have failed.

The way that Alice and Bob use this device to send messages is based on the sequence of Shift and Reduce operations that the shift-reduce parser performs (in the diagrams, this sequence is listed under "Steps performed so far"). Specifically, they break the final sequence of steps into chunks of 5 and then decode those chunks of 5 based on the following code:

A = RRRRR	F = RRRSS	K = RRRRS	P = RRSSR	U = SSSSS	Z = RRSRS
B = RRRSR	G = SSSSR	L = SRRSS	Q = SSRSR	V = SRSRS	
C = SSRSS	H = SSSRS	M = RSSSS	R = SSRRS	W = RRSRR	
D = SRSRR	I = SRRRS	N = SRSSR	S = RRSSS	X = RSRRS	
E = SSSRR	J = RSSRR	O = SRRRR	T = SSRRR	Y = SRRSR	

Therefore, the example worked out above would not actually yield a complete message because its final sequence of steps is SSRSSRR, so Alice or Bob could interpret the first 5 characters (SSRSS) as C, but there would still be two characters (RR) left dangling.

With this code, Alice and Bob successfully sent many messages to each other, where each message would name the country where their organization's next meeting would be. Eve continued to intercept these messages, and after years of careful analysis she managed to crack the code.

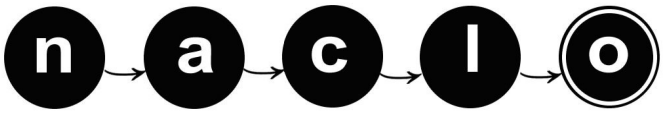
Answer the following questions in the Answer Sheets:

R1. The first message Eve intercepted after she broke the code was this sentence:

```
Salvador saw the orange spots on the giraffe
```

With mounting excitement, she decoded the message...but was utterly perplexed at the result. *Does this mean that PARSE owns some sort of spaceship?!?!*, she wondered to herself as she threw away the useless message.

a. What did Eve think the message said?
b. Where were Alice and Bob actually meeting?



(R) A Make-Shift Code (4/4)

R2. A few weeks later, Eve intercepted a second message between Alice and Bob:

```
the glass saw that the really soft orange felt really heard the giraffe
```

Once again, she decoded the message, and this time it actually made sense! As soon as she figured it out, she hopped on a plane to the country it named and waited there for Alice and Bob, but they never showed up! At least the trip wasn't a total waste--it was Eve's first visit to Africa.

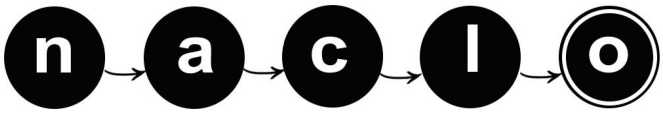
a. Where had Eve gone?
b. Where were Alice and Bob actually meeting?

R3. It's time for Alice to choose a new place to meet. When she and Bob last met, it took them a long time to find each other because all that Alice knew was the country where Bob wanted to meet, and countries are pretty big. To fix this, Alice has decided to send Bob the name of a city instead of a country. She's made a list of the six cities she most wants to visit: Ottawa, Quito, Oslo, Fez, Irkutsk, and Perth. The map below shows where these cities are located:

It turns out that only one of these six cities is a possible place for Alice and Bob to meet. On your answer sheet, please answer the following questions:

a. Which of these six cities is the one where they can meet?
b. Why wouldn't the other five cities work?
c. Give an example sentence that Alice could send to Bob to represent the name of the city that would work.

Note: The shift-reduce parsing algorithm is used in the real world to parse both human languages and programming languages. A parsing algorithm (such as this one) that builds up an analysis of a sequence by putting together individual units of the sequence is known as a bottom-up parsing algorithm. The opposite type of strategy, namely starting with the Sent symbol and trying to expand it until the desired sequence is created, is known as top-down parsing.



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 (1/10)

(I) A Menya Puzzle

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

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

   17. ☐

2. a.

   b.

   c.

   d.

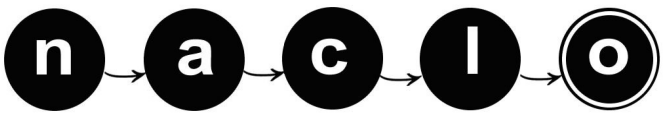
   e.

   f.

   g.

3.

4.



YOUR NAME: REGISTRATION #

Answer Sheet (2/10)

(J) It's True: The Truth About Chalcatongo Mixtec

1. 1. ☐ 2. ☐ 3. ☐ 4. ☐ 5. ☐ 6. ☐

   7. ☐ 8. ☐ 9. ☐

2. a.

   b.

   c.

   d.

   e.

   f.

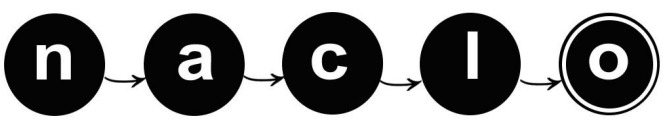
   g.

   h.

   i.

   j.

3.



YOUR NAME: REGISTRATION #

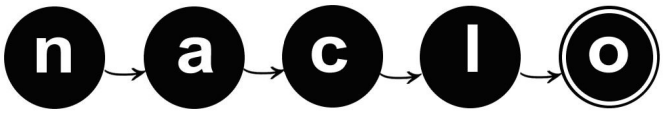
Answer Sheet (3/10)

(K) Sri Lankan Names

Start	End	Type

(L) Peeled Potato Act with Annie

1. 1. [ ] 2. [N] 3. [ ] 4. [ ] 5. [ ] 6. [ ] 7. [ ] 8. [ ]
   9. [ ] 10. [ ] 11. [ ] 12. [ ] 13. [ ] 14. [V] 15. [N]
2. 1. [ ] 2. [ ] 3. [ ] 4. [ ] 5. [ ] 6. [ ] 7. [ ] 8. [ ]
   9. [ ] 10. [ ] 11. [ ] 12. [ ] 13. [ ] 14. [ ] 15. [ ]
3. Baseline rule: [ ]
   1. [ ] 2. [ ] 3. [ ] 4. [ ] 5. [ ] 6. [ ] 7. [ ] 8. [ ]
   9. [ ] 10. [ ] 11. [ ] 12. [ ] 13. [ ] 14. [ ] 15. [ ]
4. 1. [ ] 2. [ ] 3. [ ]
5. [ ]



YOUR NAME: REGISTRATION #

Answer Sheet (4/10)

(M) Quests and Requests in Nivkh

1.

1.	2.	3.	4.	5.	6.	7.	8.
9.	10.	11.	12.	13.	14.	15.	16.
17.	18.	19.	20.	21.	22.	23.	

2.

a.

b.

c.

d.

e.

3.

a.

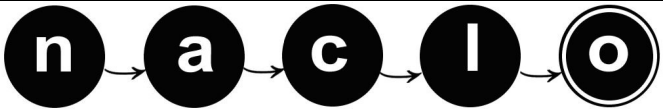
b.

c.

d.

e.

4.

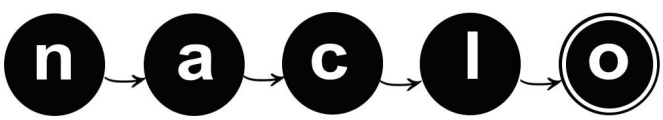


YOUR NAME: REGISTRATION #

Answer Sheet (5/10)

(N) You've Got This: Fijian Ownership

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



YOUR NAME: REGISTRATION #

Answer Sheet (6/10)

(N) You've Got This: Fijian Ownership (continued)

2. b.

(O) To Know or Not to Know Literary Tamil

1. a.

b.

c.

d.

e.

f.

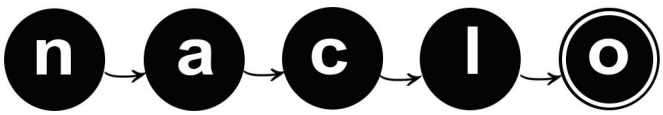
g.

h.

i.

j.

k.



YOUR NAME: REGISTRATION #

Answer Sheet (7/10)

(O) To Know or Not to Know Literary Tamil (continued)

1. l.

   m.

2. a.

   b.

   c.

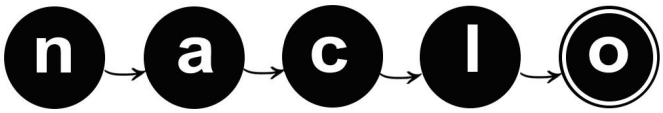
   d.

   e.

   f.

   g.

3.



YOUR NAME: REGISTRATION #

Answer Sheet (8/10)

(P) I Know What I Saw

1. 1. ☐ 2. ☐ 3. ☐ 4. ☐ 5. ☐ 6. ☐

   7. ☐ 8. ☐ 9. ☐

2. a.

   b.

   c.

   d.

   e.

   f.

3. a.

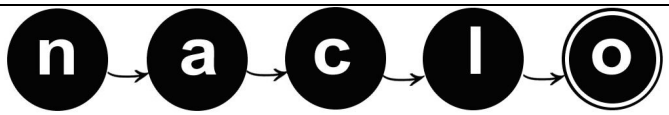
   b.

   c.

   d.

   e.

4.

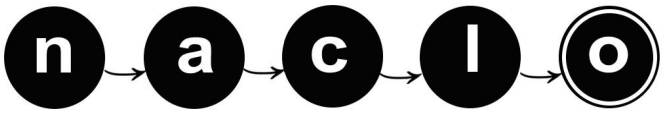


YOUR NAME: REGISTRATION #

Answer Sheet (9/10)

(Q) Better Left Unsaid

1. a.
   b.
2. c.
   d.
3. e.
   f.
4. g.
   h.
   i.
5. j.
   k.
   l.
   m.
6. n.
   o.
   p.
7. (box continues on the next page)



YOUR NAME: REGISTRATION #

Answer Sheet (10/10)

(Q) Better Left Unsaid (continued)

7.

(R) A Make-Shift Code

1. a.

   b.

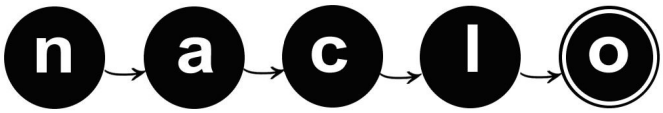
2. a.

   b.

3. a.

   b.

   c.



Extra Page for Problem …

Use as additional pages. **Make sure to write the problem number on each page**.