

***The Thirteenth
Annual***

**North American
Computational
Linguistics
Olympiad
2019**

www.nacloweb.org

**Invitational
Round**

March 7, 2019

Serious language puzzles that are surprisingly fun!

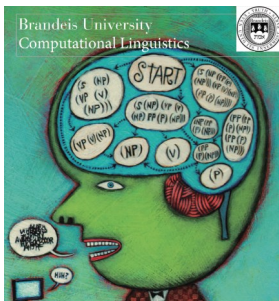
-Will Shortz, Crossword editor of The New York Times and Puzzlemaster for NPR



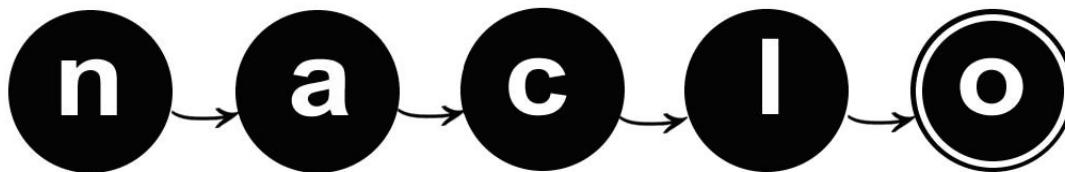
Carnegie Mellon University
Language Technologies Institute



**Massachusetts
Institute of
Technology**



Yale University



Welcome to the thirteenth 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 South Korea.
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 2019 Organizers

Program Committee:

Adam Hesterberg -- Massachusetts Institute of Technology
Aleka Blackwell -- Middle Tennessee State University
Ali Sharman -- University of Michigan
Andrés Salanova -- University of Ottawa
Andrew Lamont -- University of Massachusetts
Annie Zhu -- Massachusetts Institute of Technology
Babette Verhoeven -- Aquinas College
Daniel Lovsted -- McGill University
David Mortensen -- Carnegie Mellon University
Dick Hudson -- University College London
Dorottya Demszky -- Stanford University
Dragomir Radev -- Yale University
Elisabeth Mayer -- Australian National University
Elysia Warner -- University of Cambridge
Ethan Chi -- Stanford University
Harold Somers -- All Ireland Linguistics Olympiad
James Pustejovsky -- Brandeis University
James Hyett -- University of Toronto
Kevin Liang -- University of Pennsylvania
Liam McKnight -- University of Cambridge
Lori Levin -- Carnegie Mellon University
Mary Laughren -- University of Queensland
Oliver Sayeed -- University of Pennsylvania
Shuli Jones -- Massachusetts Institute of Technology
Sonia Reilly -- Massachusetts Institute of Technology
Susan Barry -- Manchester Metropolitan University
Tom Roberts -- University of California, Santa Cruz
Verna Rieschild -- Macquarie University
William Yang -- Yale University

Problem Credits:

Problem I: The Afrihili Word Machine - Michael Salter and Aleka Blackwell
Problem J: Polish These Nouns! - Milena Veneva
Problem K: Two Róngs Don't Make a Right - Ethan Chi
Problem L: We Were Born to Solve This Problem - Daniel Lovsted
Problem M: Colorless Green Concepts Scripting Furiously - Daniel Lovsted
Problem N: Fun With Witsuwit'en - Daniel Lovsted
Problem O: Infer a Surprise - Tom McCoy
Problem P: Do You Hear the Master's Moon? - Sam Ahmed
Problem Q: A Tale of Kieu-plets - Ethan Chi
Problem R: Disambiguate This! - Ethan Chi

Organizing Committee:

Adam Hesterberg -- Massachusetts Institute of Technology
Aleka Blackwell -- Middle Tennessee State University
Ali Sharman -- University of Michigan
Andrew Lamont -- University of Massachusetts



NACLO 2019 Organizers (cont'd)

Organizing Committee (cont'd):

Annie Zhu -- Massachusetts Institute of Technology
Caroline Ellison -- Stanford University
Daniel Lovsted -- McGill University
David Mortensen -- Carnegie Mellon University
Dragomir Radev -- Yale University
Ethan Chi -- Stanford University
James Pustejovsky -- Brandeis University
Janis Chang -- University of Western Ontario
Kevin Liang -- University of Pennsylvania
Lori Levin -- Carnegie Mellon University
Margarita Misirpashayeva -- Massachusetts Institute of Technology
Matthew Gardner -- Carnegie Mellon University
Patrick Littell -- University of British Columbia
Shuli Jones -- Massachusetts Institute of Technology
Simon Huang -- University of Waterloo
Sonia Reilly -- Massachusetts Institute of Technology
Tom McCoy -- Johns Hopkins University
Tom Roberts -- University of California, Santa Cruz
William Yang -- Yale University
Yilu Zhou -- Fordham University

US Team Coaches:

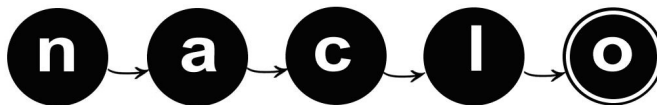
Aleka Blackwell, Middle Tennessee State University
Dragomir Radev, Yale University
Lori Levin, Carnegie Mellon University

Canadian Coordinators and Team Coaches:

Daniel Lovsted, McGill University

USA Contest Site Coordinators:

Austin Math Circle: Joshua Dever
Brandeis University: James Pustejovsky, Sarah Irwin, Nikhil Krishnaswamy
Brigham Young University: Deryle Lonsdale
California State University: Michael Ahland
Carnegie Mellon University: Mary Jo Bensasi, Lori Levin
Columbia University: Kathy McKeown, Brianne Cortese, Smaranda Muresan
College of William and Mary: Dan Parker
Cornell University: Abby Cohn, Sam Tilsen
Emory University: Jinho Choi, Phillip Wolff



NACLO 2019 Organizers (cont'd)

USA Contest Site Coordinators (cont'd):

Georgetown University: Emma Manning

Indiana University: Sandra Kuebler

Johns Hopkins University: Rebecca Knowles, Pamela Shapiro, Tom McCoy

Kansas State University: Natasha Rojkovskaia

Keiser University, Jacksonville: Deborah Williams

Long Island University Post: Rachel Szekely

Massachusetts Institute of Technology: Margarita Misirpashayeva, Sonia Reilly

Middle Tennessee State University: Aleka Blackwell

Minnesota State University Mankato: Rebecca Bates, Dean Kelley, Louise Chan

Montclair State University: Anna Feldman

Northeastern Illinois University: Judith Kaplan-Weinger, Richard Hallett

Northwestern University: Alaina Arthurs

Ohio State University: Michael White, Micha Elsner, Marie de Marneffe

Oregon Mathletics: Shirley Ma

Oregon State University: Liang Huang

Princeton University: Christiane Fellbaum, Misha Khodak

San Diego State University: Rob Malouf

Southern Illinois University: Vicki Carstens, Jeffrey Punske

SpringLight Education Institute: Sherry Wang

Stanford University: Daisy Leigh, Eli Strauss-Reis, Matthew Early

Stony Brook University: Lori Repetti, Sarena Romano, Jeffrey Heinz

UCLA: Sue Han, Lily Kawaoto

Union College: Kristina Striegnitz, Nick Webb

University of California, Irvine: Sameer Singh, Zhengli Zhao, Kristen Salsbury

University of Colorado at Boulder: Silva Chang

University of Houston: Tamar Solorio, Deepti Bhat, Giovanni Molina

University of Illinois at Urbana-Champaign: Benjamin Leff, Greg Sigalov, Julia Hockenmaier

University of Kentucky: Kevin McGowan, Andrew Byrd

University of Maryland: Kasia Hitczenko, Laurel Perkins, Aaron Doliana, Sigwan Thivierge, Nadiya Klymenko

University of Massachusetts, Lowell: Anna Rumshisky, David Donahue, Willie Boag

University of Massachusetts: Andrew Lamont, Anastasia Chobany

University of Memphis: Vasile Rus

University of Michigan: Steven Abney, Sally Thomason, Marcus Berger

University of Nebraska, Omaha: Parvathi Chundi

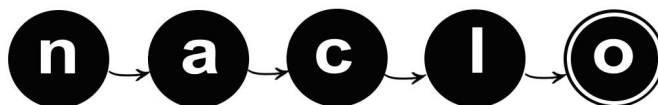
University of North Carolina Charlotte: Wlodek Zadrozny, Hossein Hematialam, Kodzo Wegba

University of North Texas: Rodney Nielsen

University of Notre Dame: David Chiang

University of Pennsylvania: Chris Callison-Burch, Mitch Marcus, Cheryl Hickey, Oliver Sayeed, Marianna Apidianaki, Derry Wijaya,
Anne Cocos

University of Southern California, ISI campus: Jon May, Nima Pourdamghani



NACLO 2019 Organizers (cont'd)

USA Contest Site Coordinators (cont'd):

University of Texas at Dallas: Vincent Ng, Jing Lu, Gerardo Ocampo Diaz

University of Utah: Brendan Terry, Jessica Larsen, Justin Nistler, Joselyn Rodriguez

University of Washington: Jim Hoard, Joyce Parvi

University of Wisconsin, Milwaukee: Joyce Boyland, Hanyon Park, Gabriella Pinter, Anne Pycha

Western Washington University: Kristin Denham

Yale University: Raffaella Zanuttini

Canada Contest Site Coordinators:

Dalhousie University: Vlado Keselj, Dijana Kosmajac

Simon Fraser University: Ashley Farris-Trimble

University of British Columbia: Yadong Liu, Jozina Vander Klok

University of Ottawa: Andrés Pablo Salanova

University of Toronto: Jordan Ho, Minh-Tam Nguyen, Lola Bradford

University of Western Ontario: Janis Chang

University of Calgary: Dennis Storoshenko

University of Alberta: Herbert Colston, Sally Rice

McGill University: Lisa Travis



NACLO 2019 Organizers (cont'd)

Booklet Editors:

Dragomir Radev, Yale University
William Yang, Yale University
Sonia Reilly, Massachusetts Institute of Technology
Daniel Lovsted, McGill University
Ethan Chi, Stanford University

Sponsorship Chair:

James Pustejovsky, Brandeis University

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:

The hosts of the 200+ High School Sites

All material in this booklet © 2019, North American Computational Linguistics Olympiad and the authors of the individual problems. Please do not copy or distribute without permission.



NACLO 2019

Sites



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

(I) The Afrihili Word Machine (1/2) [5 Points]

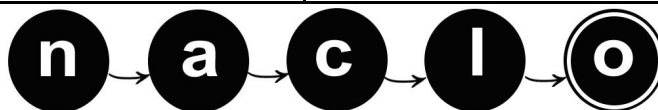
In 1970, a Ghanaian historian named Kumi Attobrah designed a Pan-African auxiliary language which he named Afrihili. As was the case with Esperanto (the most widely spoken constructed international auxiliary language devised in 1887 by L. L. Zamenhof), Attobrah decided to solve the problem of having to create a large number of unique words by developing an elaborate system of inflections and derivations for making new words.

Below are some words in Afrihili with their English translations:

adu	tooth	ikalini	pens
ajamuri	republic	ijamura	president
akalini	pen	ilengi	horses
amadu	dentist	imukazi	girls
amkate	bread	isabamatu	cobbler/shoemaker
amola	children	ishule	school
amukamo	kingdom	omuntu	man
aturesine	bouquet	olengi	horse
afidi	machine	uruzi	river
emeli	ship	omola	child
uruzindi	stream	omukazi	girl
eshuli	principal	emelisini	fleet
eture	flowers	oluganda	dialect
enti	date tree	omuntundu	dwarf

11. Fill in the gaps in the table below, marking your answers on the answer sheets:

English	Afrihili
machinist	
ships	
	ajamura
flower	
	amkamate
group of girls	
date fruit	
	oluga
shoe	
king	



(I) The Afrihili Word Machine (2/2)

I2. Below are three more Afrihili words and three options for a likely translation of the word. Pick the most likely translation to be correct, and explain your choice—make sure to mark these in your answer sheets.

A. imulenzi (a) fruit (b) boys (c) bridge

Explain your choice in your answer sheets:

B. aposino (a) baggage (b) classroom (c) parent

Explain your choice in your answer sheets:

C. iwelemase (a) book (b) library (c) librarian

Explain your choice in your answer sheets:

I3. Describe what you have learned about the structure of words in Afrihili in your answer sheets.



(J) Polish These Nouns! (1/3) [10 points]

The Polish language is a West Slavic language spoken by approximately 50 million people in Poland and around the world. Here are some phrases in Polish in the singular and plural and their translations in English.

Singular	Plural	Translation
<i>adwokat</i>	<i>adwokaci</i>	advocate (a person who supports a cause)
<i>ambaras</i>	<i>ambarasy</i>	problem
<i>autor</i>	<i>autorzy</i>	author
<i>biedny napis</i>	<i>biedne napisy</i>	poor subtitle
<i>blady indyk</i>	<i>blade indyki</i>	pale turkey
<i>blady optymista</i>	<i>bladzi optymiści</i>	pale optimist (someone with a positive outlook)
<i>ślotnisty artysta</i>	<i>ślotniści artyści</i>	muddy artist
<i>chiński antropolog</i>	<i>chińscy antropolodzy</i>	Chinese anthropologist
<i>cienki gnat</i>	<i>cienkie gnaty</i>	weak bone
<i>dochód</i>	<i>dochody</i>	income
<i>drobniuski granat</i>	<i>drobniuskie granaty</i>	tiny pomegranate
<i>dystrybutor</i>	<i>dystrybutorzy</i>	distributor
<i>egipski adwersarz</i>	<i>egipscy adwersarze</i>	Egyptian adversary
<i>kacyk</i>	<i>kacykowie</i>	cacique (the chief of an indigenous tribe)
<i>kardynał</i>	<i>kardynałowie</i>	cardinal (an official in the Church)
	<i>kardynały</i>	cardinal (the bird)
<i>kolega</i>	<i>koledzy</i>	colleague
<i>emir</i>	<i>emirowie</i>	emir (a title for the ruler of a Muslim country)
<i>nijaki zasób</i>	<i>nijakie zasoby</i>	unremarkable resource
<i>notariusz</i>	<i>notariusze</i>	notary (someone who certifies legal documents)



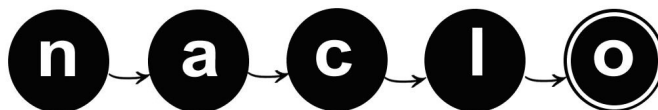
(J) Polish These Nouns! (2/3)

Singular	Plural	Translation
<i>objazd</i>	<i>objazdy</i>	detour
<i>obojczyk</i>	<i>obojczyki</i>	collarbone
<i>odcinek</i>	<i>odcinki</i>	episode (of a TV show)
<i>pancerz</i>	<i>pancerze</i>	armor
<i>Polak</i>	<i>Polacy</i>	Pole (someone from Poland)
<i>szacunek</i>	<i>szacunki</i>	respect

Note that **sz**, **rz**, **cz**, **dz**, **ń**, and **ł** are consonants and that **a**, **ó**, and **y** are vowels.

- J1.** The Polish word **darmozjad** means "a loafer," or a person who lives on others' expense. You might expect the plural of **darmozjad** to be **darmozjadzi**. However, the plural is actually **darmozjady**. Explain why this is the case in your answer sheets.
- J2.** **wielkolud** ('giant') has two plurals. What are they, and when would each be used? Mark your answers in your answer sheets.
- J3.** Give the plural forms of the following phrases in your answer sheets.

	Singular	Plural	Translation
A	<i>kalendar</i>		calendar
B	<i>jarski kapelus</i>		vegetarian hat
C	<i>łotr</i>		villain
D	<i>robotnik</i>		worker
E	<i>chorowity chirurg</i>		sickly surgeon
F	<i>partner</i>		partner
G	<i>zwalisty golfista</i>		stocky golfer
H	<i>zachodnioeuropejski akordeonista</i>		Western European accordionist
I	<i>chybki pilot</i>		swift pilot
J	<i>czepek</i>		bonnet



(J) Polish These Nouns! (3/3)

	Singular	Plural	Translation
K	<i>grzejnik</i>		heater
L	<i>rudy grób</i>		red tomb
M	<i>szympan</i>		chimpanzee
N	<i>pan</i>		lord
O	<i>zalqzek</i>		seed

J4. Explain how Polish plurals are formed in your answer sheets.



(K) Two Róngs Don't Make a Right (1/1) [10 Points]

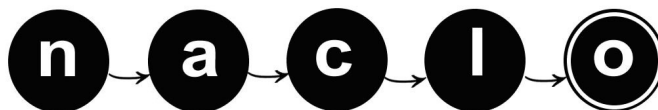
According to legend, the *Lepcha* script was devised in the early 18th century by Prince Chakdor Namgyal of Tibet. Today, it is still used to write the *Róng* (ᱫᱷᱟ) language, which is spoken by around 50,000 people in Sikkim, India. *Róng* is distantly related to Tibetan, Burmese, and Chinese. Below are some words written in *Lepcha* script, with their transcriptions in a different order. Note that *ó* and *ú* are distinct vowels (unrelated to *o* and *u*).

1. ᱮᱪᱟ	A. <i>bakto</i>	'grain'
2. ᱫᱷᱟ	B. <i>bi</i>	'vegetable'
3. ᱛᱟᱱ	C. <i>bik</i>	'cow'
4. ᱫᱷᱟ	D. <i>chik</i>	'to weigh'
5. ᱫᱷᱟ	E. <i>chung</i>	'little'
6. ᱫᱷᱟ	F. <i>dú</i>	'umbrella'
7. ᱫᱷᱟ	G. <i>ka</i>	'overseer'
8. ᱫᱷᱟ	H. <i>kajú</i>	'dog'
9. ᱫᱷᱟ	I. <i>khek</i>	'freeze'
10. ᱫᱷᱟ	J. <i>lali</i>	'a kind of laurel tree'
11. ᱫᱷᱟ	K. <i>món</i>	'pig'
12. ᱫᱷᱟ	L. <i>radi</i>	'blanket'
13. ᱫᱷᱟ	M. <i>thúk</i>	'season'
14. ᱫᱷᱟ	N. <i>tsung kóng</i>	'a kind of sorghum'
15. ᱫᱷᱟ	O. <i>tsúk</i>	'to bite'
16. ᱫᱷᱟ	P. <i>ót</i>	'to pick fruit'
17. ᱫᱷᱟ	Q. <i>úng</i>	'water'
18. ᱫᱷᱟ	R. <i>út</i>	'otter'

K1. Determine the correct correspondences in your answer sheets.

K2. Write *chóng* 'hand', *jik* 'native land' and, *thikúng* 'great-grandfather' in Lepcha in your answer sheets.

K3. Transcribe the following Róng words in your answer sheets: ᱫᱷᱟ 'to crack', ᱫᱷᱟ 'to read', ᱫᱷᱟ elephant.



(L) We Were Born to Solve This Problem (1/1) [15 Points]

Here are some words in Cupeño¹ and their translations in English:

- | | |
|---------------------|--|
| 1. pulinchemyax | We were born. (we were given birth to) |
| 2. pinehamanin | I embarrassed him. |
| 3. empulinwen | You (pl.) used to bear children. |
| 4. chimichungpenqal | He used to kiss us. |
| 5. ewel | You (sg.) grew. |
| 6. wixpemyax | They were stepped on. |
| 7. pemhamawen | They used to feel ashamed. |
| 8. piwixeminwen | You (pl.) used to step on him. |
| 9. nepuyqal | I used to dine |
| 10. pichempuynin | We fed him |
| 11. chakweyax | You (sg.) were caught. |
| 12. nichakwpeminwen | They used to catch me |
| 13. imichakwnen | I caught you (pl.) |

L1. Translate into English on your answer sheets:

chungpemyax

L2. Translate into Cupeño on your answer sheets:

We used to raise you (sg.)

L3. Explain how verbs are formed in Cupeño in your answer sheets

1. Cupeño is an extinct language of the Uto-Aztecan family, formerly spoken in Southern California. The last native speaker, Roscinda Nolasquez, died in 1987.

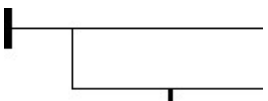
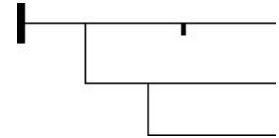
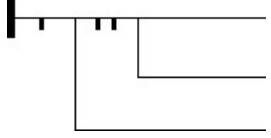
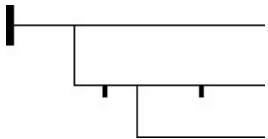
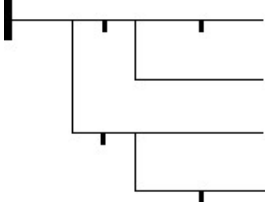


(M) Colorless Green Concepts Scripting Furiously (1/2)

[10 Points]

Modern logicians represent the logical relationships between statements with a straightforward notation. For example, if we represent the statement “Canada is beautiful” with p , then we can represent the statement “Canada isn’t beautiful” with $\neg p$ (read as “not p ”). If we have two statements represented by p and q , then we can represent “if p , then q ” as $p \rightarrow q$, and similarly we can represent “ p and q ” and “ p or q ” as $p \wedge q$ and $p \vee q$, respectively. Pretty easy, right?

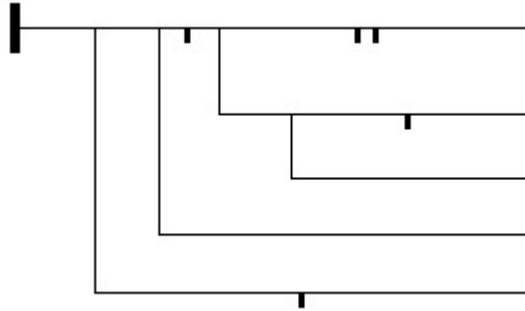
But things weren’t always this clear! In 1879, German logician Gottlob Frege published a seminal work on logic called *Begriffsschrift*, which literally translates to “concept script”. The notation he used, also called *Begriffsschrift*, confused many readers with its two-dimensional format and use of few symbols. That being said, the *Begriffsschrift* notation is a carefully thought-out system that adheres to formal rules. Here are some examples of *Begriffsschrift* formulas, with their translations into modern logical notation.

<i>Begriffsschrift</i>	Modern notation
	$B \vee A$
	$(C \rightarrow B) \rightarrow \neg A$
	$C \wedge \neg(B \rightarrow A)$
	$(C \rightarrow \neg B) \vee A$
	$(D \vee C) \vee (B \wedge A)$



(M) Colorless Green Concepts Scripting Furiously (2/2)

M1. Translate from *Begriffsschrift* into modern notation:

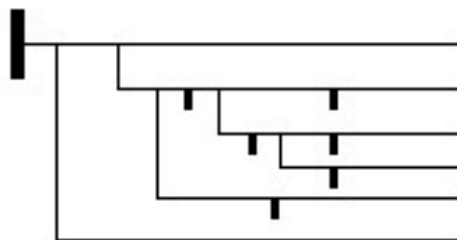


Begriffsschrift ain't just pretty to look at! It's also a meaningful logical language. As an example, consider the following facts which you may not have known about astrophysics:

All quaxors are galactions. Most of those pulsoids with a sateotrope are galactions, too. A pulsoid with a sateotrope is only not a galaction when it is dingly. (Of course, all this only holds true if the polyverse is Groop-normal.)

M2. Match these sentences to letters A to F to correctly complete the *Begriffsschrift* formula below (note that there are two possible correct answers) in your answer sheets:

- i. x is a galaction
- ii. the polyverse is Groop-normal
- iii. x is dingly
- iv. x has a sateotrope
- v. x is a quaxor
- vi. x is a pulsoid



M3: Explain how the *Begriffsschrift* notation works in your answer sheets



(N) Fun With Witsuwit'en (1/1) [15 Points]

Here are some words and phrases in Witsuwit'en and their English translations:

1. bət	a. abdomen
2. bətdex	b. bat
3. cəs	c. blanket
4. dex	d. broadleaf plantain
5. dexyəs	e. cold
6. dəlk ^w ax	f. dog
7. dəlk ^w axbət	g. dog harness
8. dəlk ^w axdətay	h. face
9. dəlk ^w axne+dəc	i. feather
10. dəni	j. female dog
11. dəninin	k. fine powder snow
12. təc	l. frog
13. təc'tl'ol	m. icicle
14. təc'yəs	n. light blue
15. təc'dəni	o. male dog
16. təc'nani	p. man
17. ne+dəc	q. penny
18. nin	r. snow
19. wəq'əz	s. snowflake
20. wəq'əz yəs	t. snow on branches or rooftops
21. wəq'əz yu	u. tooth
22. yəs	v. top
23. yəscəs	w. upper part of stomach
24. yu	x. wolf

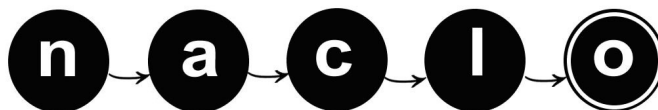
N1. Match the English translations to the Witsuwit'en words in your answer sheets.

N2. Translate to English in your answer sheets: dətay, tl'ol.

Note 1: The Witsuwit'en language is spoken by about 130 people in British Columbia, Canada.

Note 2: The broadleaf plantain (Plantago major) grows throughout British Columbia. Its broad leaves act as shelter for many small reptiles, insects, and amphibians.

Note 3: Many species of frog are found in British Columbia, varying in shape, size, and colour.



(O) Infer a Surprise (1/4) [5 Points]

A popular task in natural language processing is called natural language inference (NLI). This task involves training a model to take two sentences and label whether the first sentence *entails* the second sentence. (Sentence 1 is said to entail Sentence 2 if Sentence 2 is guaranteed to be true whenever Sentence 1 is true). Here are some examples of entailment and non-entailment:

Sentence 1	Sentence 2	Label
The judge is 6 feet tall and the lawyer is 5 feet tall.	The judge is taller than the lawyer.	entailment
Lichen grows on every continent.	Lichen grows on Antarctica.	entailment
The dentist was born in Illinois.	The dentist was born in Chicago.	non-entailment
Lichen grows on every continent.	The judge is taller than the lawyer.	non-entailment
Lichen grows on every continent.	Lichen does not grow on every continent.	non-entailment

To get a computer to solve this task, the standard approach is to train the computer on many examples like the ones above. Ideally, the computer will solve the task by learning to understand the sentences and therefore figure out which sentences have meanings that entail the meanings of other sentences. However, sometimes a computational model will place too much weight on certain coincidences in the training data, and this tendency will cause it to make incorrect predictions. Suppose a model is trained on these examples:

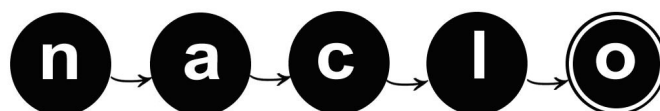
Sentence 1	Sentence 2	Label
Rhode island is the smallest US state.	Rhode Island is smaller than Connecticut.	entailment
Rice is a type of grass.	Rice is a plant.	entailment
Ottawa is the capital city of Canada.	Ottawa is in Alberta.	non-entailment
No human has ever set foot on Mars.	Several animals have been to Mars.	non-entailment

A model trained on these examples might learn the following generalization:

If both sentences start with R, the sentence pair should be labeled entailment. Else, label non-entailment.

However, this conclusion is incorrect. It makes the wrong predictions for the following sentences:

Sentence 1	Sentence 2	Correct label	Prediction
Regularization is useful.	Regularization is not useful.	non-entailment	entailment
No amphibians can echolocate.	Frogs cannot echolocate.	entailment	non-entailment



(O) Infer a Surprise (2/4)

A natural language inference model has been trained on the sentences in the following table (in practice, such a model would use a much larger training set, but we are displaying a small set to keep it manageable):

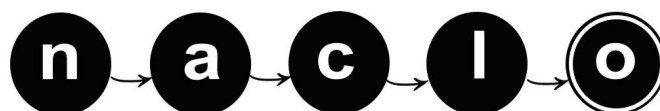
#	Sentence 1	Sentence 2	Label
1	the happiest walrus lives in Paris	the happiest walrus lives in Berlin	non-entailment
2	all dogs are mammals	all mammals are dogs	non-entailment
3	all dogs are mammals	no dog is not a mammal	entailment
4	we have stopped eating at Loretta's Diner	we once ate at Loretta's Diner	entailment
5	ice cream contains sugar	Chile is a narrow country	non-entailment
6	apples are red, and oranges are orange	oranges are orange	entailment
7	I love chocolate milk	I enjoy strawberry milk	non-entailment
8	the building is seventy feet tall	the building is over sixty-three feet in height	entailment
9	Charles Dickens was born in Portsmouth	the author of <i>A Tale of Two Cities</i> was born in a coastal city in the south of England	entailment
10	ice cream contains sugar	ice cream has sugar in it	entailment
11	Mary knows that the vase broke	the vase broke	entailment
12	all mammals are welcome here	Paul the walrus is welcome here	entailment
13	the editor read the submission	the submission was read by the editor	entailment
14	the squirrel chased the chipmunk	the chipmunk chased the squirrel	non-entailment
15	the folder containing my passport is in the filing cabinet	my passport is in the filing cabinet	entailment
16	I have never seen a walrus	I have never seen a manatee	non-entailment
17	Augustus was the first emperor	Augustus was an emperor	entailment
18	every walrus enjoys swimming	most walruses enjoy yoga	non-entailment
19	etiquette demands that one display a certain degree of respect toward one's elders	nothing in life is ever free	non-entailment
20	I like Baltimore very much	I like Baltimore	entailment
21	Wilhelmina has a cousin	Wilhelmina has at least one aunt or uncle	entailment



(O) Infer a Surprise (3/4)

The model was then tested on many examples, and it got the following examples wrong:

#	Sentence 1	Sentence 2	Correct Label	Model Prediction
22	while the painter painted the furniture was covered with a plastic sheet	the painter painted the furniture	non-entailment	entailment
23	I have never, ever seen a walrus	I have never seen a walrus	entailment	non-entailment
24	the book on the table is blue	the table is blue	non-entailment	entailment
25	fish swim	this is an example of a dummy sentence that is being used for demonstration purposes	non-entailment	entailment
26	the only animals in the aviary are birds	the aviary does not have a heron living in it	non-entailment	entailment
27	I have never seen a walrus	I have without a doubt seen a walrus	non-entailment	entailment
28	I like Baltimore very much	the moon shone like a burnished medallion	non-entailment	entailment
29	Alice believes Mary is lying	Alice believes Mary	non-entailment	entailment
30	every walrus loves oysters	Paul the walrus loves oysters	entailment	non-entailment
31	my aunt lives in Lagos with her pet walrus	my aunt lives in Lagos	entailment	non-entailment



(O) Infer a Surprise (4/4)

The model's behavior can be explained by a set of generalizations it could have learned from the training examples. On the answer sheet, fill in the blanks to describe these rules. For each rule, also write the numbers of 2 training example sentences (i.e. in the range 1-21) from which it might have learned the rule (there may be more than 2 training examples that could apply):

Rule 1: If __ (a) __ is more than __ (b) __ words long, label the sentences __ (c) __. (Evidence: example __ (d) __ and example __ (e) __)

Rule 2: If __ (f) __, label the sentences __ (g) __. (Evidence: example __ (h) __ and example __ (i) __)

Rule 3: If __ (j) __ contains the word __ (k) __, label the sentences __ (l) __. (Evidence: example __ (m) __ and example __ (n) __)

From the examples, it seems that the model has given different priorities to these three rules. Rank the rules in order of priority:

TOP PRIORITY: __ (o) __

MIDDLE PRIORITY: __ (p) __

BOTTOM PRIORITY: __ (q) __

Which test example(s) allow you to determine this ranking? __ (r) __



(P) Do You Hear the Master's Moon? (1/1) [15 Points]

Mayangna is a Misumalpan language spoken in Nicaragua and Honduras by about 8700 people. Here are some sentences in Mayangna along with their English translations. Note that in the English, 'you' refers to the singular; 'you (pl)' refers to the plural.

<i>kuring man yaihmat yak yamtasyang</i>	I do not make a <i>pipante</i> next to you.
<i>yang dakiwan ulamak balna kaswi yang mikit yak</i>	My master eats clams above me.
<i>kubamhni balna mukus balna yayamwaski witingna yaninamak taanit yak</i>	The owls do not make clouds in front of their farm.
<i>yangna dikitna yak witing palani dakaasman</i>	You do not hear his raft under us.
<i>man kammapani wainiku man mamat yak buiwi</i>	Your company shakes the moon above you.
<i>yang kuringki yaihnit yak wakurusni balna buknina kurahtayang</i>	I open the scorpions' book next to my <i>pipante</i> .
<i>man walamabis yang pinkisil talwaski witing daniwan balna yaihnitna yak</i>	Your son does not see my pencil next to his masters.
<i>mukusni balna witing sunipai balna dakakaawi palani dinit yak</i>	The clouds hear her spiders under the raft.
<i>man sumalu balna kisauri kakaswi yang dangkit yak</i>	Your dogs eat coriander behind me.
<i>man mumalalah yang yaihkit yak kastasman</i>	You do not eat your <i>guapote</i> next to me.
<i>was supai balna taanitna yak dakaaman</i>	You hear water in front of spiders.
<i>yamak yak yang titinghki wakisa talwi</i>	My grandmother sees a banana at a farm.

P1. Translate these sentences into English on your answer sheets:

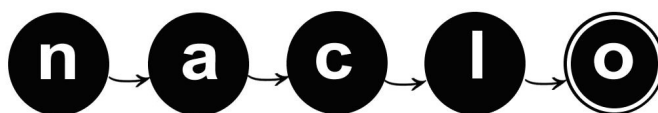
1. *waiku dakaayang man damat yak*
2. *ulamakni man kubamhma yaihnit yak buitaman*
3. *man wakimasa talasyangna yang walakibis balna taanitna yak*

P2. Translate these sentences into Mayangna on your answer sheets:

4. Our masters do not open the clouds in front of us.
5. You make the dog's raft next to the water.
6. I shake pencils above your (pl) grandmother.
7. You (pl) do not eat our bananas at a raft.

P3. Explain your solution in the answer sheets.

A *pipante* is a small wooden boat rowed with shovels and a *guapote* is a type of edible freshwater fish, sometimes called a wolf cichlid.)



(Q) A Tale of Kieu-plets (1/2) [10 Points]

The Tale of Kieu is an epic poem written in Vietnamese by author *Nguyễn Du* in 1820, which details the life of the talented young woman *Thúy Kiều*. The poem is composed of 1,627 couplets, or groups of two lines. Here are some couplets from *The Tale of Kieu* in no particular order:

1. Trăm năm trong cõi người ta,
Chữ tài chữ mệnh khéo là ghét nhau.
2. Đầu lòng hai ả tố nga,
Thúy Kiều là chị, em là Thúy Vân.
3. Nửa năm hơi tiếng vừa quen,
Sân ngô cảnh biếc đã chen lá vàng.
4. Vân rằng: Chị cũng nức cười,
Khéo dư nước mắt khóc người đời xưa.
5. Người đâu gặp gỡ làm chi,
Trăm năm biết có duyên gì hay không?

Each couplet of the *Tale* is written in the poetic meter known as *lục bát*¹, which has several rules. Even if we don't speak Vietnamese, we can still use these rules to identify whether a couplet has been corrupted. For example, here is a corrupted version of couplet #1:

- 1a. Trăm năm trong **dài** người ta,
Chữ tài chữ mệnh khéo là ghét nhau.

This couplet does not follow the rules of *lục bát* and consequently must have been corrupted. However, this is not always the case. Consider this other corruption to couplet #1:

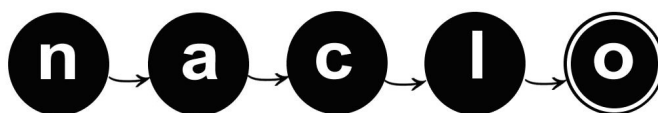
- 1b. Trăm năm trong **cá** người ta,
Chữ tài chữ mệnh khéo là ghét nhau.

This corrupted version of couplet #1 still follows the rules of *lục bát*; we would need to know Vietnamese in order to determine that there was a mistake.

In Vietnamese, the diacritics (accent marks) *á, à, â, ă, ã* indicate five of the six Vietnamese tones (known as *sắc, huyền, hỏi, nặng, and ngã* respectively); the absence of a diacritic indicates the sixth tone, *ngang*. Importantly, the diacritics *â, ă, ê, ơ, u* do not represent tones; instead, they represent slight changes to the pronunciation of the vowels. Traditionally, Vietnamese tones are divided into two groups: *sharp* tones, of which there are four, and *flat* tones, of which there are two.

Q1. Which tones belong to which groups? Write your answer on the answer sheets.

1. From the Middle Chinese words *liuk-pat*, or 'six-eight'.



(Q) A Tale of Kieu-plets (2/2)

While searching in your local library for linguistics papers, you find six more couplets from *The Tale of Kieu*. However, some have been damaged. Only two of them have remained uncorrupted—the other four have had **exactly two words changed or removed** so that they no longer follow the rules of *lục bát*. Fortunately, even if you don't know Vietnamese, you can use the rules of *lục bát* to determine which couplets are original.

6. Vâng trình hội chùa xem tường,
 Mà sao trong sổ đoạn sân có tên.

7. Âu đành quả tôn nhân duyên,
 Cùng người một hội, một thuyền đầu gái!

8. Thưa rằng: Chút phận ngây thơ,
 Dưỡng sinh đôi tiên tóc tơ chưa đền

9. Ngoài song thả thẻ oanh vàng,
 Nách tường bông liễu bay ngang trước mảnh.

10. Chàng Kim từ lại thư song,
 Nỗi nằng canh cánh bên lòng biếng khuây.

11. Mảnh Tương phát phát gió truyền,
 Hương gậy mùi nhớ, trà khan giọng.

Q2. Determine which two couplets are authentic and which four have been corrupted. For each couplet, circle “corrupted” or “authentic”. For each corrupted couplet, write the two words which have been corrupted, or write REMOVED if a word has been removed.

Here is one more couplet from *The Tale of Kieu*. The couplet has been transmitted in damaged form: for some of the words, two choices are possible (marked in parentheses).

Q3. For each combination in parentheses, determine which of the two words is correct. If both of the two words could occur and still follow the rules, circle the word BOTH. Write your answers on the answer sheets.

12. Nhấn (ở/từ) (quán/rạp) (khách/tươi) lân (la/khác),
 Tuần trắng thắm thoát nay (đà/ách) thềm hai.

Q4. Explain your observations about the structure of *lục bát* in your answer sheets.



(R) Disambiguate This! (1/3) [5 points]

One important (and often tricky) task in machine translation is *disambiguation*: identifying which sense of a word is being used in a sentence. Consider the following sentences:

- 1) The old sing.
- 2) The singers are old.

In 1), "old" is used as a noun, while in 2), "old" is used as an adjective. Computers must be able to deduce which meaning of the word is intended in order to properly label these sentences for translation. Here's an example:

1)	the	old	are	singing
	the.DET	old.N	be.VRB	sing.VRB
2)	the	singers	are	old
	the.DET	singer.N	be.VRB	old.ADJ

Here's a brief explanation of the above syntax:

- The lowercase word before the first period is the *lemma*—the base form of the word.
- The uppercase word after the period is a tag which marks the part of speech.
- The following tags are available: DET for determiners (broadly, words that come before a noun, like "this," "your," or "the"); N for nouns; PRN for pronouns (e.g. "I," "me," or "you"); VRB for verbs; ADJ for adjectives; ADV for adverbs; PREP for prepositions (words like "about").

Unlike us, computers are not automatically able to tell that the word *old* is a noun in the first sentence but an adjective in the second. Therefore, we must write rules to determine the correct tag for such words.

Here's an example rule, which is written in a syntax known as *constraint grammar*:

```
old: SELECT N if (+1 VRB)
```

This selects the noun (N) form of the word `old` if the next (+1) word is tagged as a verb (VRB), and does nothing otherwise. Note that negative numbers may be used to select previous words, as in this rule:

```
old: SELECT ADJ if (-1 DET)
```

Here's one more rule, which selects the verb form of the word "desert" in every case.

```
desert: SELECT VRB
```

Before using the rules, our computer system first tags all words that only have one possible part of speech. It then handles the rules in top-down order, applying each rule in turn to every word (from left to right) in the sentence that still has more than one possible tag. Beware: if no rule makes a decision, the system will crash!



(R) Disambiguate This! (2/3)

Below are some sentences in English containing the ambiguous word “her”, which can either serve as a determiner (‘her dress’, represented as *her*.DET) or as a pronoun (‘I saw her’, represented as *her*.PRN).

1. I see her now.
I.PRN see.VRB **her.PRN** now.ADV

2. Her son is tall.
her.DET son.N be.VRB tall.ADJ

3. The girl hears her daughter today.
the.DET girl.N hear.VRB **her.DET** daughter.N today.ADV

4. The dog looks at her.

5. The girl is her friend.

6. I am her daughter.

7. The cat saw her dog yesterday.

8. You walk with her.

9. The boy likes her.

10. A giraffe sees her now.

11. I give her flowers.

12. Her tall daughter is smart.

13. The cat examines her quizzically.

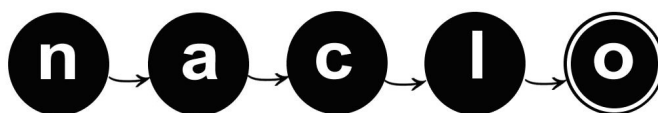
14. I am her older sister.

15. Her orange cat likes me.

R1. For sentences 4-15, indicate, on your answer sheets, whether “her” is being used as a determiner or as a pronoun.

R2. All of the above sentences—except one—can be disambiguated using three rules. List these three rules in your answer sheets, remembering that rule order matters. Assume that all words other than “her” only have one possible tag.

R3. Which sentence is disambiguated incorrectly? Explain, in your answer sheets, why it would be difficult to create a rule that would successfully tag this sentence.



(R) Disambiguate This! (3/3)

Below are some sentences in Sranan Tongo (an English-based creole language with influences from Dutch, Javanese, Hindustani, and Chinese, which is the national language of Suriname) with their translations in English.

Mi lobi den singi.	"I love the songs."
Den lobi yu singi.	"They love your songs."
Den lobi mi.	"They love me."
Mi singi abra yu lobi.	"I sing about your love."
Den lobi dati mi singi.	"They love that I sing."
Yu lobi mi sisa.	"You love my sister."
Mi lobi yu.	"I love you."

Note that no Sranan Tongo words change from their lemma forms. For example, the lemma form of *sisa* is *sisa*.N. Most of the words in these examples have parts of speech that were also present in the English examples on the previous page, but there are also two additions: “*abra*” should be tagged `PREP` (preposition) and “*dati*” should be tagged `COMP` (complementizer).

As you can see, disambiguation is much harder in Sranan Tongo than in English, as many words have multiple meanings. For example, “*lobi*” can mean “love” (noun) or “to love” (verb).

To deal with this level of difficulty, we need more powerful rules. Here's a rule that uses some additional syntax available for Sranan Tongo:

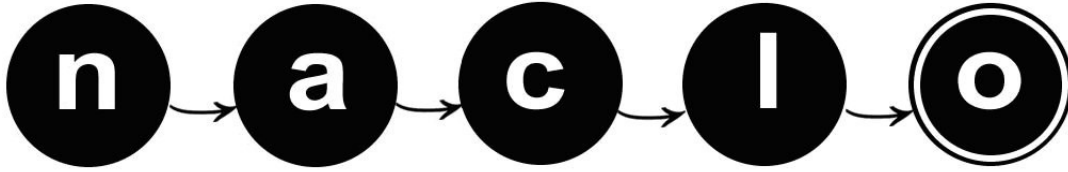
```
PRN/DET: SELECT DET if (-1 [VRB]) and (+1 PRN)
```

This selects the determiner (`DET`) form of a word that could be either a pronoun (`PRN`) or a determiner (`DET`) if the previous (`-1`) word could possibly be a verb (`[VRB]`) and the next word has been confirmed to be a pronoun (`PRN`) and does nothing otherwise. Specifically, the notation we are adding is the slash / (which can only be used before the colon, not after it); the brackets [] (which can only be used after the colon, not before it); and the word “and” (but not “or”)

Of course, you can still use the syntax given in the previous section.

- R4. Write a set of rules that would successfully disambiguate the above sentences in your answer sheets. Hint: you should need no more than 5 rules. Recall that, before the rules are applied, all words with only one possible part of speech are tagged with that part of speech.





The North American Computational Linguistics Olympiad

Contest Booklet

REGISTRATION NUMBER				

Name: _____

Contest Site: _____

Site ID: _____

City, State: _____

Grade: _____

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: _____

Answer Sheet (1/8)

(I) The Afrihili Word Machine

11.

English

machinist

ships

flower

group of girls

date fruit

shoe

king

Afrihili

ajamura

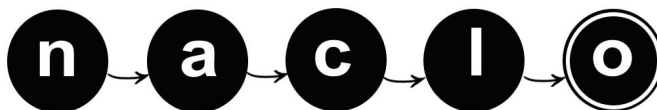
amkamate

oluga

12. A)

B)

C)



YOUR NAME:

REGISTRATION #

Answer Sheet (2/8)

13.

(J) Polish These Nouns!

J1.

J2.

J3.

A																				
B																				
C																				
D																				
E																				
F																				
G																				



YOUR NAME:

REGISTRATION #

Answer Sheet (3/8)

J3. H

I

--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--

J

--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--

K

--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--

L

--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--

M

--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--

N

--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--

O

--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--

J4.

--

(K) Two Róings Don't Make a Right

K1. 1. 2. 3. 4. 5. 6. 7. 8. 9.

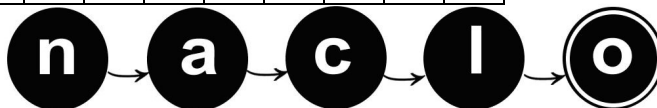
10. 11. 12. 13. 14. 15. 16. 17. 18.

K2. hand native land great-grandfather

K3.

--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--

--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--



YOUR NAME:

REGISTRATION #

Answer Sheet (4/8)

(L) We Were Born to Solve This Problem

L1.

L2.

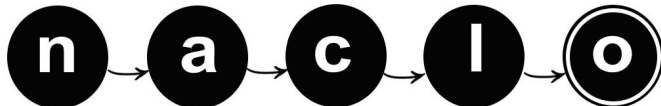
L3.

(M) Colorless Green Concepts Scripting Furiously

1.

2. i. ii. iii. iv. v. vi.

3.



YOUR NAME:

REGISTRATION #

Answer Sheet (5/8)

(N) Fun With Witsuwit'en

1. 1. 2. 3. 4. 5. 6. 7. 8.
9. 10. 11. 12. 13. 14. 15. 16.
17. 18. 19. 20. 21. 22. 23. 24.

2.
-

(O) Infer a Surprise

- (a) sentence (b) (c)
- (d) (e)

(f)

(g) (h) (i)

(j) sentence (k)

(l) (m) (n)

(o) (p) (q)

(r)

(P) Do You Hear the Master's Moon?

- P1. 1.
2.
3.



YOUR NAME:

REGISTRATION #

Answer Sheet (6/8)

P2. 4.

5.

6.

7.

P3.



Answer Sheet (7/8)

(Q) A Tale of Kieu-Plets

Q1. sharp tones

flat tones

Q2 6. corrupted / authentic

Corrupted words: #1 _____ #2 _____

7. corrupted / authentic

Corrupted words: #1 _____ #2 _____

8. corrupted / authentic

Corrupted words: #1 _____ #2 _____

9. corrupted / authentic

Corrupted words: #1 _____ #2 _____

10. corrupted / authentic

Corrupted words: #1 _____ #2 _____

11. corrupted / authentic

Corrupted words: #1 _____ #2 _____

Q3 a. ở từ BOTH

b. quán rạp BOTH

c. khách tươi BOTH

d. la khác BOTH

e. đà ách BOTH



YOUR NAME:

REGISTRATION #

Answer Sheet (8/8)

Q4

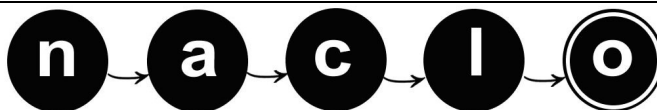
(R) Disambiguate This!

R1	4	DET	PRN	5	DET	PRN	6	DET	PRN
	7	DET	PRN	8	DET	PRN	9	DET	PRN
	10	DET	PRN	11	DET	PRN	12	DET	PRN
	13	DET	PRN	14	DET	PRN	15	DET	PRN

R2

R3

R4



YOUR NAME:

REGISTRATION #

Additional Space for Scrap Work

