naclo

The Fifteenth Annual

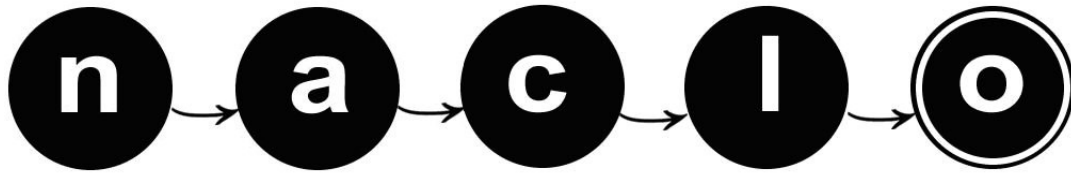# North American Computational Linguistics Open Competition

2021

www.nacloweb.org

Open Round
January 28, 2021

Serious language puzzles that are surprisingly fun!

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

Welcome to the fifteenth annual North American Computational Linguistics Open Competition! We (the NACLO organizers) are all excited for you to participate in this unique event. In order to be completely fair to all participants across North America, we need you to read, understand, and follow these rules completely.

Rules

1. The contest is three hours long and includes nine problems, labeled A to I. Note that this year's contest has one more problem than in previous years — 9 problems instead of 8.
2. If you want clarification on any of the problems, please post your question <u>privately</u> to the jury in the Piazza.
3. You may not discuss the problems with anyone except as described in items 2 & 12.
4. You may not access any information on the internet or elsewhere for the duration of the contest.
5. Each problem is worth a specified number of points, with a total of 100 points. In this year's open round, no points will be given for explanations. Instead, make sure to fill out all the answer boxes properly.
6. All your answers should be entered in the Online Answer Sheet. The link to it is provided in the Piazza.
7. Please follow the instructions in the Online Answer Sheet closely to ensure your submission is graded.
8. The top 10% of participants (approximately) across the United States and Anglophone Canada in the open round will be invited to the second round.
9. 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. Don't be discouraged if you don't finish everything. Don't give up! Very few people will solve all these problems completely in the time allotted.
11. Please shred this booklet at the conclusion of the contest.
12. **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 2021 Organizers

Organizing 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

Andrew Tockman — Massachusetts Institute of Technology

Annie Zhu — Harvard University

Ben LaFond — Harvard University

Brian Xiao — Massachusetts Institute of Technology

Corinne Soucy — Université de Montréal

Daniel Lovsted — McGill University

David Mortensen — Carnegie Mellon University

Dragomir Radev — Yale University

Duligur Ibeling — Stanford University

Ethan Chi — Stanford University

Heidi Lei — Massachusetts Institute of Technology

James Hyett — University of Toronto

James Pustejovsky — Brandeis University

Ji Hun Wang — Stanford University

Ken Jiang — University of Waterloo

Kevin Liang — University of Pennsylvania

Lori Levin — Carnegie Mellon University

Margarita Misirpashayeva — Massachusetts Institute of Technology

Matthew Gardner — Carnegie Mellon University

Mihir Singhal — Massachusetts Institute of Technology

Nathan Kim — Stanford University

Patrick Littell — University of British Columbia

Pranav Krishna — Massachusetts Institute of Technology

Rui Zhang — Yale University

Ryan Chi — Stanford University

Ryan Guan — Stanford University

Shuli Jones — Massachusetts Institute of Technology

Skyelar Raiti — University of Michigan

Sonia Reilly — Massachusetts Institute of Technology

Stella Lau — Massachusetts Institute of Technology
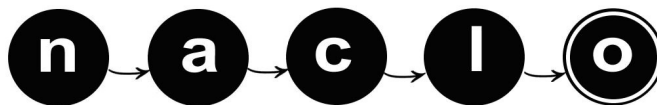
Tom McCoy — Johns Hopkins University

NACLO 2021 Organizers (cont'd)

Organizing Committee (continued):
Tom Roberts — University of California, Santa Cruz
Vanessa Hu — Harvard University
Yilu Zhu — Fordham University

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 — Harvard University
Babette Verhoeven — Aquinas College
Daniel Harbour — Queen Mary University of London
Daniel Lovsted — McGill University
David Mortensen — Carnegie Mellon University
Dick Hudson — University College London
Dragomir Radev — Yale University
Elisabeth Mayer — Australian National University
Erik Andersen — Brandeis University
Ethan Chi — Stanford University
Gordon Chi — Stanford University
Harold Somers — All Ireland Linguistics Olympiad
Heather Newell — Université du Québec à Montréal
James Hyett — University of Toronto
Jill Vaughan — University of Melbourne
Kevin Liang — University of Pennsylvania
Lori Levin — Carnegie Mellon University
Lynn Clark — University of Canterbury
Margarita Misirpashayeva — Massachusetts Institute of Technology
Mary Laughren — University of Queensland
Oliver Sayeed — University of Pennsylvania
Patrick Littell — University of British Columbia
Simi Hellsten — University of Oxford
Sonia Reilly — Massachusetts Institute of Technology

n a c l o

NACLO 2021 Organizers (cont'd)

(A) The Tortoise and the Rabbit (1/1) [10 Points]

Malagasy is a language with 25 million speakers, most of them living in Madagascar. Even though Madagascar is in Africa, Malagasy belongs to the Austronesian language family, making it a relative of Indonesian, Hawaiian, and many other languages spoken throughout the Pacific and Indian Oceans. Below are some sentences in Malagasy along with their English translations.

A1. Fill in the blanks in the table:

Nitsiky ny bitro kely.	The small rabbit smiled.
Mitsiky ny bitro kely.	The small rabbit smiles.
Mitsiky ny bitro mainty.	The black rabbit smiles.
Mihaona ny sokatra ny alika.	The dog meets the tortoise.
Niaro ny bitro ny sokatra.	The tortoise protected the rabbit.
(a)	The tortoise meets the dog.
Nitsiky ny sokatra mainty.	**(b)**

A2. Here are a few more Malagasy sentences. Fill in the blanks in this second table:

Nihaona ny vorona ity bitro lehibe ity.	This large rabbit met the bird.
Miaro ny sokatra lehibe iny alika iny.	That dog protects the large tortoise.
Miarahaba ity sahona maitso ity ny saka.	The cat greets this green frog.
Niteny ity alika kely ity.	This small dog spoke.
(a)	This small dog speaks.
Niaro ny saka ity bitro lehibe ity.	**(b)**
(c)	This green bird greeted that small frog.

Make sure you record your answers in your Answer Sheets!

(B) Junk Mail: Letters We Don't Need (1/1) [10 Points]

Recently, a group of researchers decided that the alphabet had too many letters. The letters could not fit conveniently on cell phone keyboards, there were too many options when filling in crosswords, and children were wasting time chanting all 26 letters in the alphabet song. Luckily, the researchers noticed that many letters were unnecessary. For example, *p* and *b* sound a lot alike, so they decided to replace every *p* with a *b* (while every *b* remained unchanged). For example, *plug* would be rewritten *blug*, but *bug* would stay written as *bug*. They continued this process until they narrowed down the alphabet to just 9 letters. This writing system was named "Duw Ulbubud," and the media went wild. Here is one of the headlines that ran that day:

WURDZ WULL BU ZBULLUD DUVVURUDDLU

BUGUUZU UV "DUW ULBUBUD"

The article ended with the following words:

Vur buddur ur vur wurzu, du drudzvurbud ulbubud wull bugu wrududg luug lugu u buzzludg zubblu uv

ruddub luddurz. Wu zuzd uvu udu buzur gwuuzduud: uz ud zuzduvuud? Ugzbugd ludz uv gubbluuddz

vrub buublu wu gruw ub wud du uld zuzdub.

**B1.** Rewrite each of the following words in Duw Ulbubud:

words, the, fifth, squeaky, jazz, bagpipes, vertex, calming

B2. One major problem with Duw Ulbubud is ambiguity. For example, after Duw Ulbubud was adopted, many people grew up assuming that the book *Moby Dick* was about a young bird. How is *Moby Dick* written in Duw Ulbubud? What were people mistakenly interpreting this Duw Ulbubud title to mean?

B3. TV shows began to use Duw Ulbubud's ambiguity as a source of humor. In one show, the main character texted two friends a message that was intended to say, "Give me a minute." But then one friend handed her a fruit, and the other handed her a vegetable. What was the fruit? What was the vegetable?

B4. Although the writing system was initially called Duw Ulbubud, people quickly shifted to a new name, *Zduvvu-Duzu Dulg*, a name based on how the words sound when read aloud. What does *Zduvvu-Duzu Dulg* translate to?

Make sure you record your answers in your Answer Sheets!
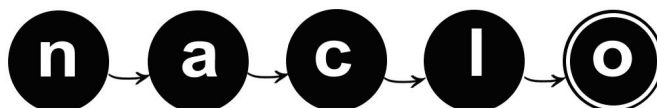
(C) Versatile Verb (1/1) [10 Points]

In the left column below appear sentences in Waama, or Yoabu, a Gur language of Benin spoken by roughly 50,000 people. These sentences appear in the writing system of the language. You do not need to know how the writing system's letters are pronounced to solve this problem. In the right column below, the translations of these sentences in English appear *in a scrambled order*.

1. *Cando dɛbite kpi, o ǹ faa o suka.*
2. *Tando dori.*
3. *N pe saaki ti yete.*
4. *Bika kɔɔsi kɔɔka.*
5. *Soosada kaate.*
6. *Suka kpi.*
7. *Ba kaate tiibu band.*
8. *N yeentire n daaso.*
9. *Bisu yɔkɔɔti.*
10. *Tiibu dori puŋa mii.*
11. *N taka n daaso yete.*
12. *Maari dikitifa pei, o ǹ fa piisi.*
13. *Suka miiki pɔmpɔmma.*
14. *Bika dori.*
15. *N kɔɔka taka Yooto yete.*

A. The tree fell in the forest.
B. A car passed by earlier.
C. I went to my friend's house.
D. The child fell.
E. Marie lost the money, but she found it.
F. It rained.
G. My hen went to Yooto's house.
H. My wife swept our house.
I. The children had fun.
J. Tchando's neighbor died, and he inherited his car.
K. They gathered under the tree.
L. I hurt my friend.
M. The soldiers assembled.
N. The car broke down.
O. The child sold the hen.

C1. Match sentences 1-15 to their English translations.

C2. One of the verbs you encountered in the sentences in the left column above would be used by speakers of this language in all 4 sentences below, each of which is translated into English. Which of the verbs you encountered above is that verb?

|	
*N tokore **verb***	'My shirt is torn.'
*Yaama **verb***	'The matter is settled.'
*O beere **verb***	'He lost his fame.'
*Yima **verb***	'The water is frozen.'

C3. Translate the following English sentences into Waama:
   (a) The children gathered under the house.
   (b) I sold my car.
   (c) Her friend played in the rain.

C4. Translate the following Waama sentences into English:
   (a) *Ba kɔɔsi ti kɔɔsu.*
   (b) *N susu kpi.*
   (c) *Maari daaso fa faa.*

Make sure you record your answers in your Answer Sheets!
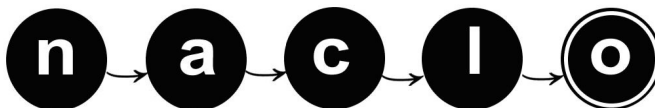
(D) Made You Look! (1/4) [10 Points]

How do humans process language? One way to study this question is with **eye tracking**, a technique that allows scientists to monitor where an experimental participant is looking while they listen to a sentence. Suppose that a person has been familiarized with this grid of images:
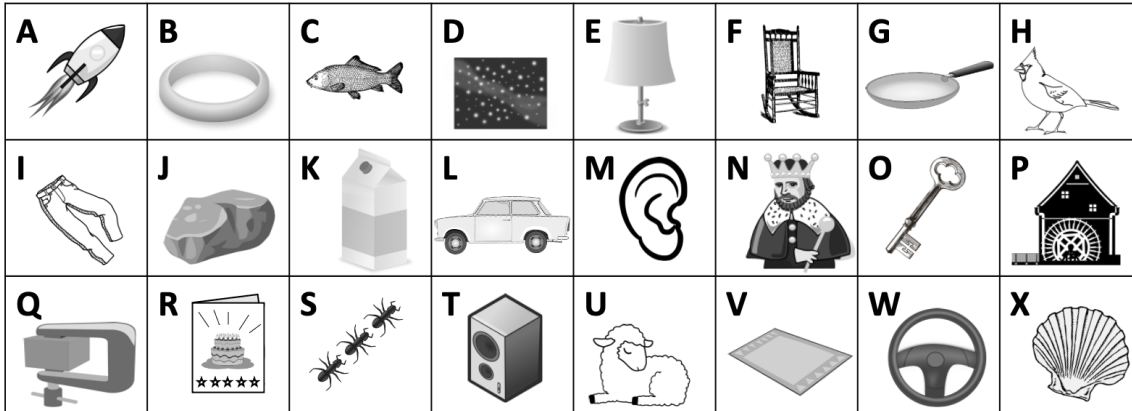
The person is then read the sentences in the table below, and we track where their eyes fall as they listen to the bolded words.[1] The table shows what the person's predicted eye trajectory is for each of five different hypotheses about how humans might process language (note that the question mark means "not looking at any particular image"). For example, for the first sentence, Hypothesis 4 predicts the listener starts out looking at no specific image, then looks at Image U, then T, then Image E.

	Hypothesis 1	Hypothesis 2	Hypothesis 3	Hypothesis 4	Hypothesis 5
Mary had a little **lamp**	?,E	U	?,U,E	?,U,T,E	**(a)**
The guitar needs an **amp**[2]	?,T	T	?,T	?,T	T
The flour is from a **mill**	?,P	P	?,P	?,P	P
I drive a **rocking chair**	?,F	L	?,J,F	?,J,N,F	L,A,F
This bird is a **cardinal**[2]	?,H	H	?,L,R,H	?,L,R,H	H
This fish is a **carp**[2]	?,C	C	?,L,C	?,L,C	C
I held the block tightly with the metal **pants**	?,I	Q	?,G,I	?,G,S,I	Q,G,I
I locked the door with a **steering wheel**	?,W	**(b)**	?,W	?,M,B,W	O,W
In the desert sky I could see the **Milky Way**[2]	?,D	D	?,P,K,D	**(c)**	D
This fish is a **clamp**[2]	**(d)**	**(e)**	**(f)**	**(g)**	C,X,Q
The floor needs a **carpet**	**(h)**	**(i)**	**(j)**	**(k)**	**(l)**

---

[1] In an actual experiment, these sentences would need to be designed differently to encourage participants to pay attention to the image. For instance, instead of saying "I drive a rocking chair", we might say, "In a minute, we will ask you to point at what I drive. I drive a rocking chair." In this problem, to save space, we use the shorter prompts shown here.

[2] A cardinal is a type of bird. A carp is a type of fish. A lamb is a young sheep. An amp is a device that makes instruments louder. A clamp is a tool for holding items together. The Milky Way is the galaxy containing the Earth.

(D) Made You Look! (2/4)

D1. Fill in the missing slots in the table on the previous page.

For the rest of this problem we will assume that **Hypothesis 5** is correct. Of course, actual human sentence processing is more complex than Hypothesis 5, but Hypothesis 5 is still a better description of human sentence processing than the other four hypotheses.

Suppose that we now use this grid of images for an experiment:

Participants are read the following sentence:

My cousin has a pet ***dolphin***

D2. Image HH is ambiguous: it can be viewed as a rabbit or as a duck. What trajectory will a participant's eyes follow if they view Image HH as a rabbit? What trajectory will their eyes follow if they view Image HH as a duck?

Another participant is read this sentence, and her eyes follow the trajectory EE, DD, FF, LL:

Alice was thirsty after her run, so she ordered some ***sushi***
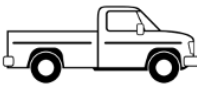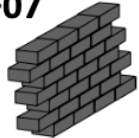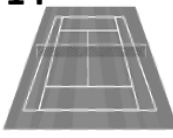
D3. Image DD depicts a carbonated beverage. Does this participant refer to carbonated beverages as *pop, soda, coke,* or *fizzy drink*?

D4. After the experiment, we somehow lost Image FF! What is Image FF? Your answer should be one word.
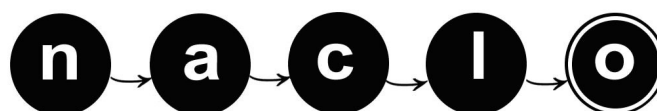
(D) Made You Look! (3/4)

In a final experiment, participants are familiarized with this array of images:

F01 truck	**F02** tractor	**F03** car	**F04** campfire		**F05** pencil
F06 bicycle	**F07** bricks	**F08** camera		**F09** cat	**F10** tennis ball
F11 sand		**F12** boat	**F13** candle	**F14** tennis court	
F15 bowtie	**F16** sticks	**F17** cake	**F18** camel		**F19** tent
F20 straws	**F21** traffic cone		**F22** present	**F23** tower	**F24** pen
	F25 strawberries	**F26** dog			**F27** castle

naclo

(D) Made You Look! (4/4)

D5. The table below, when completed, encodes a secret message which could serve as an alternate title for this problem. This message is 7 letters long (a 3-letter word followed by a 4-letter word). Fill in the missing slots in the table so that the intended secret message is still communicated. ***Hint:*** Every image is used at least once.

Remember: we are assuming that participants behave according to Hypothesis 5. When you write the sentences, you should include asterisks, but you don't need to include bolding. You can still get partial credit for correct sentences and trajectories even if you don't figure out the secret message. For the sentences, there are many correct answers.

	Sentence	Trajectory
i	I bought my friend a birthday ***campfire***	(a)
ii	(b)	(c)
iii	(d)	F27, F23, F19, F14, F10
iv	My neighbor rides to work every morning in my ***bow tie***	(e)
v	He signed his name with a dark blue ***pencil***	(f)
vi	My neighbor rides to work every morning in my ***traffic cone***	(g)
vii	Outside the construction site there was a pile of ***strawberries***	(h)

D6. What is the secret message mentioned in D5?

Make sure you record your answers in your Answer Sheets!

n → a → c → l → o
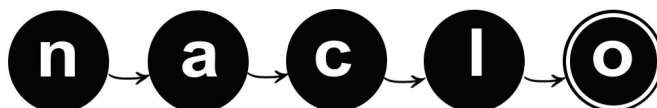
(E) The Script of Your Dreams (1/2) [15 Points]

Mandombe is a script, or writing system, created in 1978 by Wabeladio Payi. Today, it is used to write the Democratic Republic of the Congo's four national languages — Kongo, Lingala, Tshiluba, and Swahili. According to legend, the script was revealed to Payi by Congolese prophet Simon Kimbangu in a dream.[1]

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

[1] That's not all for the backstory — at one point, the script landed Payi in trouble with the authorities. The problem was related to the fact that in traditional Kongo culture, mirrors are powerful, magical devices. Evidently, something about the script was too mirror-like.

(E) The Script of Your Dreams (2/2)

On the previous page, in an arbitrary order, are 22 words of the Kongo language[2] written in the Mandombe script. Below are the same Kongo words, written in the Roman alphabet, each one accompanied by its English translation.

A.	wonso	*"at all"*		L.	mongo	*"mountain"*
B.	tewa	*"bang"*		M.	mpuku	*"mouse"*
C.	tuti	*"cloud"*		N.	mwisi	*"smoke"*
D.	mfumvu	*"cord"*		O.	fula	*"to blow"*
E.	ngombe	*"cow"*		P.	zenga	*"to cut"*
F.	fumbuka	*"dead"*		Q.	zuba	*"to hit"*
G.	kutu	*"ear"*		R.	simba	*"to hold"*
H.	tiya	*"fire"*		S.	yimbila	*"to sing"*
I.	mfinda	*"forest"*		T.	venza	*"to strike"*
J.	mpunda	*"horse"*		U.	maza	*"water"*
K.	lumonso	*"left"*		V.	nani	*"who"*

E1. Match the words written in Mandombe to their Romanized equivalents.

E2. Given the numerals 1, 2, 3, and 4, how would you write 5 in Mandombe? Choose from W, X, Y, and Z:

1	2	3	4
ꓘ	ꓔꓲ	Ꞙ	Ꟛ

W	X	Y	Z
ꓜ	ꓙ	ꓱ	ꓖ

Make sure you record your answers in your Answer Sheets!

---

[2] Kongo, also called Kikongo, is a language of the Bantu family spoken natively by about 7 million people in central Africa. It is also widely learned and used as a second language in the Democratic Republic of the Congo, and has served as a basis of several creole languages in Africa and the Americas.

n → **a** → **c** → **l** → **o**

(F) The Wiles of Reptiles (1/2) [15 Points]

Here are eight sentences in Dime.[1] The first seven are given along with their English translations.

1. *guurafis kénim šuiftin*   The crocodiles deceived them.

2. *ʔaté goštú gúdúmubim ʔóllóχ kúbit*   I carried a tall man slowly.

3. *wúdúr čək'k'ind s'áámisim yefin*   A small girl saw the eagle.

4. *guuru bášmubis ʔeftafim gáʔin*   The fearful crocodile bit birds.

5. *zimaf daχidis zití s'anubim ʔólóχ maddin*   The strong chiefs helped a black ox quickly.

6. *kété ʔámzi kitimindim kóχin*   They loved a selfish woman.

7. *goštáf koχsimid s'ááme koχsimubisim šiʔin*   Pleasant men washed the pleasant eagle.

8. *ʔámzaf gúdúmidis zime bášmubim kúbin*

Notes:

In this problem, an accent above a vowel (e.g., in "kété") indicates that the vowel has high tone. Unmarked vowels have low tone. A consonant or vowel written twice in a row (e.g., in "guurafis") is pronounced for a longer time by the speaker. An apostrophe following a consonant (e.g., in "s'anubim") indicates that the consonant is pronounced ejectively (i.e., with air pressure produced in the throat instead of the lungs).

F1. Translate the eighth sentence into English.

The six sentences below were uttered by a linguist who had been trying to learn Dime. Unfortunately, only three of them were fully understood by native speakers; the other three were deemed incorrect, meaning they contained mistakes or were not understood at all.

F2. For each of the following six sentences, choose from the options provided to convert it into a grammatically correct Dime sentence. (Remember that three of the sentences are already grammatically correct — for these three sentences, you should choose option W.) For every sentence, only one option will successfully produce a grammatically correct sentence.

(a) *ʔaté s'ááme s'anubim kóχin*
   W.   The sentence is correct as is (change nothing).
   X.   Change "*kóχin*" to "*kóχit*".
   Y.   Change "*s'anubim*" to "*s'anudim*".
   Z.   Change "*s'ááme*" to "*s'áámaf*".

---

[1] Dime is an Afroasiatic language spoken in Ethiopia. Estimates of the number of speakers of Dime vary from a few hundred, on the low end, up to a few thousand; all studies agree, however, that the number of speakers is decreasing, and the language is threatened with extinction. Dime speakers live in a series of farming communities located along a mountain range and have relatively little contact with outside groups. The speakers are skilled metal producers, with techniques for extracting and smelting ores forming an important part of their traditional cultural knowledge.

(F) The Wiles of Reptiles (2/2)

(b) *wúdúraf gúdúmid zime kitimubisim ʔóllóχ šuiftin*

      W.   The sentence is correct as is (change nothing).

      X.   Change *"wúdúraf gúdúmid"* to *"wúdúrid gúdúmaf"*.

      Y.   Change *"šuiftin"* to *"šuiftiť"*.

      Z.   Change *"ʔóllóχ"* to *"ʔámzi"*.

(c) *goštú zitáfisim koχsim yefin*

      W.   The sentence is correct as is (change nothing).

      X.   Change *"goštú zitáfisim"* to *"goštáfis zitáfid"*.

      Y.   Change *"zitáfisim"* to *"zitidim"*.

      Z.   Change *"zitáfisim koχsim"* to *"zitáf koχsimidisim"*.

(d) *wúdúris guuru s'anisubim maddin*

      W.   The sentence is correct as is (change nothing).

      X.   Change *"s'anisubim"* to *"s'anindisim"*.

      Y.   Change *"s'anisubim"* to *"s'anubisim"*.

      Z.   Change *"s'anisubim"* to *"s'anisim"*.

(e) *goštú čək'k'ubis kénim šiʔin*

      W.   The sentence is correct as is (change nothing).

      X.   Change *"čək'k'ubis kénim"* to *"čək'k'im kété"*.

      Y.   Change *"goštú čək'k'ubis"* to *"goštaf čək'k'idis"*.

      Z.   Change *"šiʔin"* to *"šiʔiť"*.

(f) *kété s'ááme bášmubisim ʔólóχ gáʔin*

      W.   The sentence is correct as is (change nothing).

      X.   Change *"ʔólóχ"* to *"ʔámzi"*.

      Y.   Change *"s'ááme"* to *"s'áámim"*.

      Z.   Change *"bášmubisim"* to *"bášmidisim"*.

F3. Translate the corrected versions (or the original versions, for the sentences that were already correct) of the six sentences above into English.

Make sure you record your answers in your Answer Sheets!

(G) The Skates of Wrath (1/1) [15 Points]

Here are some words in Sauk[1] and their unmatched English translations in alphabetical order.

1.	âchimowa	A. computer
2.	anemôha	B. dog
3.	anemôhêha	C. Ferris wheel
4.	chîtapikâneki	D. foal[2]
5.	chîtapiwa	E. laundromat
6.	kôkenikâneki	F. library
7.	mehikowîthenikâneki	G. living room
8.	meshitêtêpithoni	H. metal
9.	meshotêneki	I. Mexican restaurant
10.	methenahikanani âchimôni	J. newspaper
11.	methenahikani	K. oats / hay
12.	methenahikanikâneki	L. Oklahoma City[2]
13.	pîwâpehkwi	M. paper
14.	pîwâpehkwi âchimôni	N. puppy
15.	pîwâpehkwi wînêtepi	O. roller skates
16.	têtêpithonani mahkathêhani	P. telephone
17.	tôskashi wîtheniweni	Q. to sit
18.	tôskashêhâ	R. to tell

G1. Determine the correct correspondences.

G2. Translate the following into English: tôskashâ, kôkenêwa

G3. Translate the following into Sauk: *brain, shoe store, town*

Make sure you record your answers in your Answer Sheets!

---

[1] Sauk is an Algonquian language spoken in Iowa, Kansas, and Oklahoma. Like many Indigenous languages of North America, Sauk was negatively impacted by European colonization and is now at risk of extinction. Today, Sauk is spoken by several hundred people; of these people, only a handful speak Sauk monolingually (i.e., as their only language). However, recent revitalization efforts, including courses on Sauk offered in schools, may help slow or reverse its decline.
[2] A foal is a young horse. Oklahoma City is Oklahoma's capital and largest city.

(H) Sequitur (1/2) [5 Points]

Space is valuable, so it's often desirable to compress data — that is, to use less space to convey the same information. One common data compression strategy is to identify repeated patterns within the data and somehow consolidate these repetitions.[1] First, let's look at Sequitur, a fast compression algorithm that uses the repeated pattern strategy. The table below shows Sequitur running on the input "abcdbcabcd", with its output at the very bottom:

Step Number	Column 1	Column 2
1)	a	S = a
2)	ab	S = ab
3)	abc	S = abc
4)	**(a)**	S = abcd
5)	abcdb	**(b)**
6)	abcdbc	S = abcdbc
	aXdX	S = aXdX X = bc
7)	**(c)**	**(d)** **(e)**
8)	aXdXab	**(f)** X = bc
9)	aXdXabc	S = aXdXabc X = bc
	aXdXaX	S = aXdXaX **(g)**
	(h)	**(i)** X = bc Y = aX
10)	**(j)**	**(k)** X = bc Y = aX
	ZXZ	S = ZXZ X = bc Y = aX Z = Yd
S = ZXZ; X = bc; Y = aX; Z = Yd		

---

[1] It just so happens that identifying repeated patterns in a human language can tell us a lot about how that language works. If you want to know more about how linguists (and especially computational linguists) use compression in language technologies, try problem (I), Non Sequitur, in this booklet. Be aware, however, that solving problem (I) will not give you any advantage in solving this problem.

(H) Sequitur (2/2)

H1. Fill in the blanks (marked with bolded letters) in the table on the previous page.

Next, take a look at Byte Pair Encoding (BPE), a widely used compression algorithm that also uses the repeated pattern strategy. Like in the previous example, BPE is running with an input of "abcdbcabcd", and its output is shown at the bottom of the table:

Step Number	Column 1	Column 2
1)	abcdbcabcd	S = abcdbcabcd
2)	**(a)**	**(b)** X = bc
3)	YdXYd	S = YdXYd X = bc **(c)**
4)	ZXZ	S = ZXZ X = bc Y = aX Z = Yd
S = ZXZ; X = bc; Y = aX; Z = Yd		

H2. Fill in the blanks (marked with bolded letters) in the table above.

As you can see, for the input "abcdbcabcd", Sequitur and BPE produce the same output! But this isn't always the case.

H3. For each of the following inputs, say whether Sequitur and BPE give the same or different outputs:

    a)   abcabdbcbc
    b)   abbcaddca
    c)   bacbcbabacbcba
    d)   ccdbccdbccacc
    e)   ccdbccdbccaccacca

__Make sure you record your answers in your Answer Sheets!__

(I) Non Sequitur (1/3) [10 points]

At this very moment, scientists at NACLO Labs are hard at work on *Def21*, a definition-generating machine for English words. If they can just fix a few small problems, *Def21* is sure to be a huge popular success.

Def21 has access to a word list, L1, that includes some common English words and abbreviations, along with their definitions:

> L1
>
> | | |
> |---|---|
> | al: *American League* | rest: *stay still* |
> | cent: *a penny* | st: *a street* |
> | grate: *a grid of metal bars* | stat: *a quantitative fact* |
> | in: *contained by* | sting: *sharply injure* |
> | ion: *a charged particle* | sure: *certain* |
> | rate: *give a score* | union: *a worker's organization* |
> | real: *truly existing* | |

The NACLO Labs team also compiled a second list, L2, for *Def21* to use, by scanning many English words with their compressor machine.[1] The compressor identified patterns of letters that showed up frequently in the words it scanned, and it turned out that many of these word pieces had meanings of their own! For instance, when the compressor scanned the words "redo" (meaning "do again"), "regrow" (meaning "grow again"), and "replay" (meaning "play again"), it picked out "re" as a frequent repetition. Even better, with a little clever programming, the scientists were able to find out the meaning of "re" automatically (that is, based only on the meanings of the scanned words).

Here are the items in L2, and for each one, a sample of three words that it was found in:

> L2:
>
> ation *(adaptation, consideration, installation)*
> ing *(jumping, knowing, wandering)*
> ize *(equalize, publicize, randomize)*
> pre *(predawn, prehistoric, premodern)*
> re *(redo, regrow, replay)*
> un *(uncertainty, uncommon, untie)*

Finally, *Def21* has a sophisticated way of combining definitions from L1 with the meanings of its L2 items to guess the definition of a word it does not recognize. So, when *Def21* is given the word "unsure", which it does not recognize, *Def21* should recognize the pieces "un" from L2 and "sure" from L1 and produce the definition "the opposite of certain". In this case, *Def21* is confident about its result; when *Def21* is unsure about its guess, it will flag the result with one or more question marks.

---

[1] If you want to know more about how NACLO Labs' compressor machine works, try problem (H), Sequitur, in this booklet. The NACLO Labs machine operates on the same basic principles as the algorithms shown there. Be aware, however, that solving problem (H) will not give you any advantage in solving this problem.

What *Def21* still struggles with is dividing unrecognized words into pieces. The great minds at NACLO Labs are trying out four different strategies — *1, 2, 3,* and *4* — for this task. Here are some words that *Def21* did not recognize, and the results of each strategy:

ingrate
> *1:* contained by give a score (?)
> *2:* contained by a grid of metal bars
> *3:* continually give a score
> *4:* contained by a grid of metal bars

resting
> *1:* contained by a street again (?)
> *2:* sharply injure again
> *3:* continually stay still
> *4:* sharply injure again

unionize
> *1:* make into the opposite of a charged particle
> *2:* make into a workers' organization
> *3:* make into a workers' organization
> *4:* make into a workers' organization

predation
> *1:* a charged particle again (????)
> *2:* not yet a charged particle (???)
> *3:* not yet the result of (?)
> *4:* not yet the result of (?)

realize
> *1:* make into American League again
> *2:* make into truly existing
> *3:* make into truly existing
> *4:* make into truly existing

station
> *1:* a street a charged particle (??)
> *2:* a quantitative fact a charged particle
> *3:* the result of a street
> *4:* the result of a street

(I) Non Sequitur (3/3)

Here are brief descriptions the designers wrote up of the four strategies, with a few key words removed:

*1: Choose the **(a)** piece from **(b)** at each step.*
*2: At the first step, choose the **(c)** piece from **(d)**. At each step after this, choose the **(e)** piece from **(f)**.*
*3: At the first step, choose the **(g)** piece from **(h)**. At each step after this, choose the **(i)** piece from **(j)**.*
*4: Choose the **(k)** piece from **(l)** at each step.*

I1. From the options *longest, shortest, L1, L2,* and *L1 or L2*, fill in the blanks **(a)** to **(l)**.

I2. Give the results of each strategy *1-4* on each of the inputs below. In a case where there is not enough information to choose between multiple possible results, you may enter any result that is consistent with the examples given.

  (a) reunion

  (b) unrest

  (c) presto

I3. In a recent test, a Canadian scientist asked *Def21* for the definition of "centre" (which Americans typically spell "center"). This time, all four strategies *1-4* agreed on a result. What's more, *Def21* noticed that this result was the same as that given by all four strategies for a *different* unrecognized English word. What is the other word, and what definition was given to both words by all strategies?

Make sure you record your answers in your Answer Sheets!
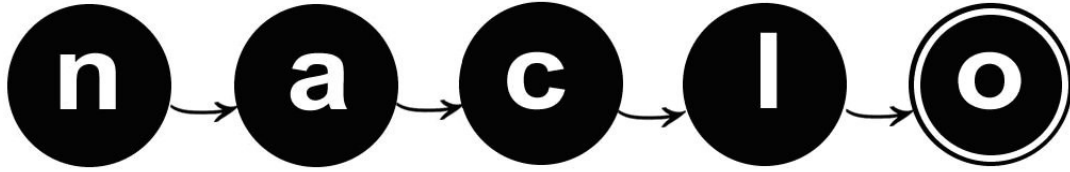
n → a → c → l → o

naclo

Contest Booklet

REGISTRATION NUMBER				

Name: _____

Contest Site: _____

Site ID: _____

City, State/Province/Territory: _____

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: _____
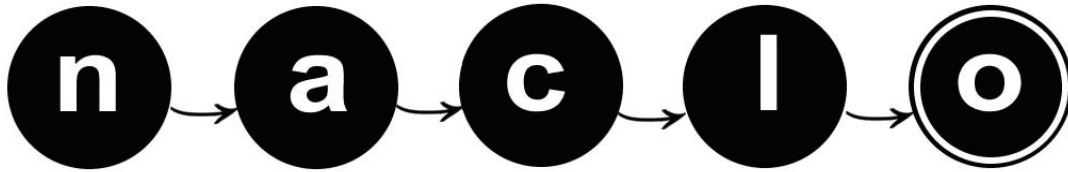
n a c l o

Demographics

Gender: ☐Female ☐Male ☐Other: _____ ☐Prefer not to respond

If you are competing for the US, please fill out this box:

Do you consider yourself Hispanic/Latinx?

☐ Yes
☐ No
☐ Prefer not to respond

Which of the following categories describe you? Select all that apply.

☐ American Indian or Alaskan Native
☐ Asian
☐ Black or African American
☐ Native Hawaiian or Pacific Islander
☐ White
☐ Other: _____
☐ Prefer not to respond

If you are competing for Canada, please fill out this box:

Which of the following categories describe you? Select all that apply.

☐ Non-Indigenous
☐ Métis
☐ First Nations
☐ Inuit
☐ Other: _____
☐ Prefer not to respond

Which of the following categories describe you? Select all that apply.

☐ Arab
☐ Black
☐ Chinese
☐ Filipino
☐ Japanese
☐ Korean
☐ Latin American
☐ South Asian
☐ Southeast Asian
☐ West Asian (e.g. Iranian, Afghan, etc.)
☐ White
☐ Other: _____
☐ Prefer not to respond

Answer Sheets (1/7)

(A) The Tortoise and the Rabbit

1. Translate the following sentences into the other language:

    (a) The tortoise meets the dog.

    | |
    | |

    (b) Nitsiky ny sokatra mainty.

    | |
    | |

2. **(a)** This small dog speaks.

    | |
    | |

    (b) Niaro ny saka ity bitro lehibe ity.

    | |
    | |

    (c) This green bird greeted that small frog.

    | |
    | |

(B) Junk Mail: Letters We Don't Need

1.   Translate the following words into "Duw Ulbubud":

    words: | | *the:* | |

    fifth: | | *squeaky:* | |

    jazz: | | *bagpipes:* | |

    vertex: | | *calming:* | |

2.   Title in Duw Ulbubud: | |

        Misinterpretation: | |
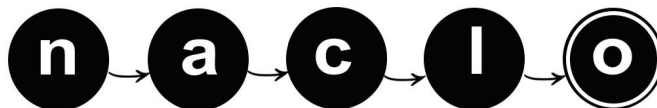
Answer Sheets (2/7)

(B) Junk Mail: Letters We Don't Need (continued)

3.   Fruit:

      Vegetable:

4.   Translation of new name:

(C) Versatile Verb

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

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

2.   The verb is

3.   Translate the following English sentences into Waama:

      (a)   The children gathered under the house.

      (b)   I sold my car.

      (c)   Her friend played in the rain.

4.   Translate the following Waama sentences into English:

      (a)   *Ba kɔɔsi ti kɔɔsu.*

Answer Sheets (3/7)

(C) Versatile Verb (continued)

(b)   *N susu kpi.*

(c)   *Maari daaso fa faa.*

(D) Made You Look!

1.   Fill in the slots from the table:

(a) [   ]   **(b)** [   ]   **(c)** [   ]   **(d)** [   ]

(e) [   ]   **(f)** [   ]   **(g)** [   ]   **(h)** [   ]

(i) [   ]   **(j)** [   ]   **(k)** [   ]   **(l)** [   ]

2.   What trajectory will the participant's eyes follow if they view Image HH as a rabbit?   [   ]
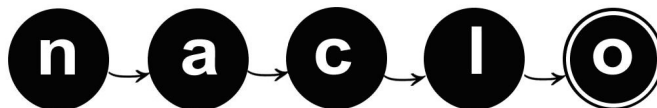
What trajectory will the participant's eyes follow if they view Image HH as a duck?   [   ]

3.   Circle which one of the following is the participant's word for a carbonated beverage:

pop   *soda*   *coke*   *fizzy drink*

4.   What is image FF? Your answer should be one word.   [   ]

5.   Fill in the slots from the table:

(a) [   ]

(b) [   ]

Answer Sheets (4/7)

(D) Made You Look! (continued)

(c) [   ]

(d) [   ]

(e) [   ]

(f) [   ]

(g) [   ]

(h) [   ]

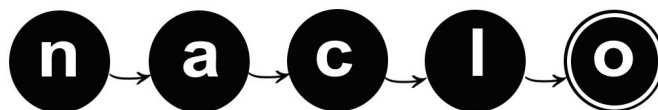6.   What is the secret message? You should enter one letter per square.

[ ][ ][ ]   [ ][ ][ ][ ]

(E) The Script of Your Dreams

1.   For each Mandombe word, write the letter (from A to V) of its Romanized equivalent:

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

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

17. [ ]   18. [ ]   19. [ ]   20. [ ]   21. [ ]   22. [ ]

2.   The number 5 is written as (write W, X, Y, or Z):   [ ]

(F) The Wiles of Reptiles

1.   Translate the eighth sentence into English:

[   ]

nac lo

Answer Sheets (5/7)

(F) The Wiles of Reptiles (continued)

2. Write the letter of the correct option for each of the six sentences:

(a) ☐   **(b)** ☐   **(c)** ☐   **(d)** ☐   **(e)** ☐   **(f)** ☐

3. Write the English translation of the correct version of each of the six sentences:

(a) ☐

(b) ☐

(c) ☐

(d) ☐

(e) ☐

(f) ☐

(G) The Skates of Wrath

1.   For each Sauk word, write the letter (from A to R) of its English translation:

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

7. ☐   8. ☐   9. ☐   10. ☐   11. ☐   12. ☐

13. ☐   14. ☐   15. ☐   16. ☐   17. ☐   18. ☐

2.   Translate the following Sauk words into English:

tôskashâ   ☐

kôkenêwa   ☐

3.   Translate the following English words into Sauk:

brain   ☐

Answer Sheets (6/7)

(G) The Skates of Wrath (continued)

shoe store

town

(H) Sequitur

1.   **(a)** ____   **(b)** ____   **(c)** ____

(d) ____   **(e)** ____   **(f)** ____

(g) ____   **(h)** ____   **(i)** ____

(j) ____   **(k)** ____

2.   **(a)** ____   **(b)** ____   **(c)** ____

3.   Circle the correct answer:
(a)   abcabdbcbc   Same   Different
(b)   abbcaddca   Same   Different
(c)   bacbcbabacbcba   Same   Different
(d)   ccdbccdbccacc   Same   Different
(e)   ccdbccdbccaccacca   Same   Different

(I) Non Sequitur

1.   a) ____   b) ____   c) ____

d) ____   e) ____   f) ____

g) ____   h) ____   i) ____

j) ____   k) ____   l) ____

n → **a** → **c** → **l** → **o**
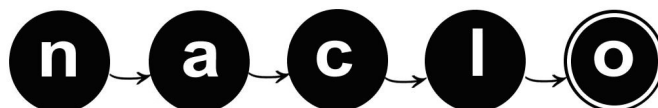
Answer Sheets (7/7)

(I) Non Sequitur (continued)

2.   Write the definitions of the following words according to each of the strategies:

(a) reunion   1:

2:

3:

4:

(b) unrest   1:

2:

3:

4:

(c) presto   1:

2:

3:

4:

3.   Other word:

Definition: