n a c l o

NACLO thanks the following for their generous contributions:

U.S. National Science Foundation

boomi virtusa

Carnegie Mellon University
Language Technologies Institute

NAACL

...and many individual donors.

THE TWENTIETH ANNUAL

North American Computational Linguistics Open Competition

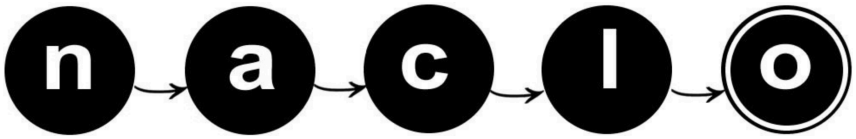
2026

Invitational Round: March 19, 2026

www.naclo.org

> *"Serious language puzzles that are surprisingly fun!"*
>
> —Will Shortz, *crossword editor of The New York Times and Puzzlemaster for NPR*

2026



Welcome to the twentieth annual North American Computational Linguistics Open Competition! We (the NACLO organizers) are 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 four hours long and includes 8 problems, labeled from J to Q.
2. Follow the facilitators' instructions carefully.
3. No electronic devices of any kind may be used or accessed during the contest. All such devices must be turned off and put away as instructed by the facilitators. No aids are permitted other than blank paper and writing utensils.
4. If you want clarification on any of the problems, talk to a facilitator. The facilitator will consult with the jury before answering.
5. You may not discuss the problems with anyone except as described in items 4 & 11.
6. Each problem is worth a specified number of points, with a total of 100 points. In the Invitational Round, some questions require explanations. **Make sure to fill out all the answer boxes properly in accordance with the guidelines on the cover of the Answer Sheets.**
7. All your answers should be written clearly in the Answer Sheets at the end of this booklet. **ONLY THE ANSWER SHEETS WILL BE GRADED.**
8. Write your name and NACLO ID on each page of the Answer Sheets.
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! This is a *contest*, and not a *test*—there is no "passing" or "failing" score. If we have done our job well, very few people will solve all these problems completely in the time allotted.
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 2026 Organizers and Credits (1/2)

NACLO Chair:
Lori Levin — Carnegie Mellon University

Executive Committee:
Lori Levin — Carnegie Mellon University
Patrick Littell — National Research Council Canada
Daniel Lovsted — University of Ottawa
Tom McCoy — Yale University

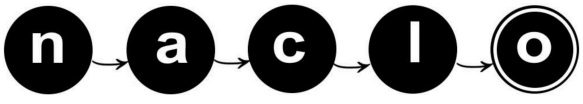
Organizing Committee:
Elena Filatova — City University of New York Graduate Center
Arul Kolla — Massachusetts Institute of Technology
Lori Levin — Carnegie Mellon University
Jane Li — Johns Hopkins University
Daniel Lovsted — University of Ottawa
Cerulean Ozarow — Brown University
Luke Sawczak — Toronto District Christian High School

Organizing Committee Co-Chairs:
Elena Filatova — City University of New York Graduate Center
Jane Li — Johns Hopkins University
Cerulean Ozarow — Brown University

Program Committee:
Aaron Chai — Massachusetts Institute of Technology
Brandon Cho — Princeton University
Angela Hou — Princeton University
Ken Jiang — University of Waterloo
Devin Joe — Massachusetts Institute of Technology
Arul Kolla — Massachusetts Institute of Technology
Lori Levin — Carnegie Mellon University
Patrick Littell — National Research Council Canada
Daniel Lovsted — University of Ottawa
Tom McCoy — Yale University
Varin Sikka — Stanford University
Nina Stadermann — University of California, Berkeley
Aidan Wang — University of Toronto
Kyle Zhang — Massachusetts Institute of Technology

Program Committee Chair:
Arul Kolla — Massachusetts Institute of Technology

Academic Committee:
Rebecca Knowles – National Research Council Canada
Patrick Littell — National Research Council Canada
Tom McCoy — Yale University

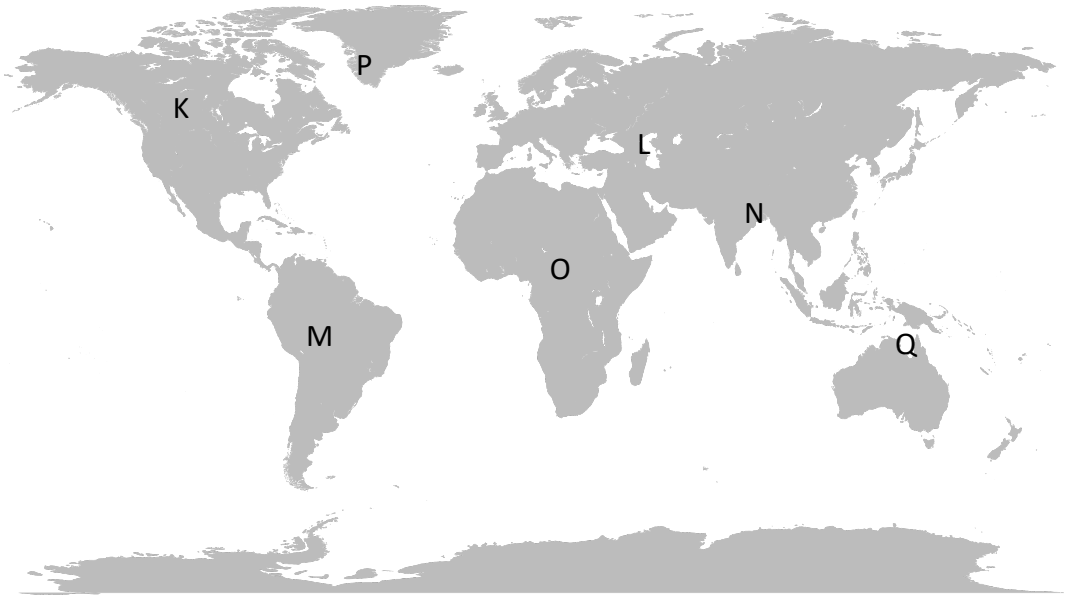


NACLO 2026 Organizers and Credits (2/2)

Special thanks to:
200+ High School and University Site hosts in Canada and the USA

Booklet Editor:
Arul Kolla

Problem Credits:
(J) Kyle Zhang
(K) Arul Kolla
(L) Angela Hou
(M) Arul Kolla
(N) Varin Sikka and Angela Hou
(O) Arul Kolla
(P) Nina Stadermann
(Q) Devin Joe



Locations in the map are approximate.

We are grateful for the support of many institutional and individual donors who make this contest possible.

(J) Strawbery Shortcake (1/2) [5 Points]

Real-world AI language models sometimes confidently state things that are completely made up; this is often called a **hallucination**. Recently, NACLOLabs has hired you to help understand why its new chatbot (an artificial intelligence system designed to have conversations), ParrotBot, has been hallucinating. Instead of understanding facts or meaning, ParrotBot only cares about patterns. Here is ParrotBot's single rule for generating text:

> *Look at the last word I just wrote, look through my training data to see what word follows it most frequently, and output that word.*

Here, *training data* means the collection of example text that ParrotBot was trained on. For example, consider ParrotBot trained on the training data 1a–1c shown on the right. Given an initial input `poison` and asking it to output 5 more words after, ParrotBot's response is generated using its method as follows:

Training data 1:
1a. poison can be bad for you
1b. vitamins can be good for you
1c. exercise can be good for people

	Initial input	→	`poison`
1.	After `poison`, only `can` appears (1x).	→	`can`
2.	After `can`, only `be` appears (3x).	→	`be`
3.	After `be`, `good` (2x) beats `bad` (1x).	→	`good`
4.	After `good`, only `for` appears (2x).	→	`for`
5.	After `for`, `you` (2x) beats `people` (1x).	→	`you`

Therefore, ParrotBot's full message is "`poison can be good for you`," which wasn't in the training data at all! This shows how our simple language model can create factually untrue sentences even when only provided true ones.

J1. NACLOLabs creates a new chatbot called ParrotBot$_2$, which is trained on *only* the data 2a–2e below. Given an initial input `small` and asking it to output 4 more words after, write down ParrotBot$_2$'s full message in your Answer Sheets.

Training data 2:
2a. small elephants eat small strawberries
2b. small elephants are not spies
2c. some spies are alleged spies
2d. alleged spies are very sneaky
2e. no elephants are alleged blueberries

Modern AI language models are more sophisticated than ParrotBot, but they still sometimes make mistakes. One of the most widely known hallucinations in modern AI language models involves letter-based reasoning. For example, when asked how many times the letter *r* appears in the word *strawberry*, language models often give incorrect answers such as *2*.

The mistake most likely comes from how the model represents text. Instead of working letter-by-letter, it usually breaks text into pieces called **tokens** and then works with those tokens. A token is a piece of text the model processes as one unit—sometimes a single letter, sometimes a fragment of a word, and sometimes a whole word. As such, it can struggle with tasks that require careful reasoning at the individual-letter level.

Make sure you record your answers in your Answer Sheets!

(J) Strawbery Shortcake (2/2)

Because tokens are built from frequent patterns in text, they often look like meaningful pieces of words, but they do not always line up with meaning. For example, one possible tokenization of the sentence *She read the bread is heavy* is [*S*][*he*] [*read*] [*t*][*he*] [*b*][*read*] [*is*] [*he*][*a*][*vy*]. Notably, the sequence *read* inside *bread* is just part of the word, not a meaningful piece on its own, and the sequence *he* inside *the* does not mean the pronoun *he*.

Below is a message from an AI language model, responding to a question asking it to count how many letters are common between certain pairs of words. However, instead of reasoning at the level of individual letters, it used its internal token representation; that is, it counted how many *tokens* are exactly the same, and then reported that number as though it were the number of letters in common.

The six words given to the AI have been replaced with numbered stickers [1] – [6]; the same word is always covered by the same sticker.

> [1] and [2] have two letters in common.
> [1] and [3] have two letters in common.
> [4] and [6] have two letters in common.
> [1] and [4] have one letter in common.
> [2] and [5] have one letter in common.
> [4] and [5] have one letter in common.
> [1] and [6] have no letters in common.
> [2] and [4] have no letters in common.
> [3] and [5] have no letters in common.
> None of these words have any repeating letters, so I didn't need to worry about how to count duplicates. And interestingly enough, all the words are four letters long!

Here are the words [1] – [6], in arbitrary order:

A. *changing* B. *coaches* C. *excessive* D. *proceeding* E. *processes* F. *started*

J2. In your Answer Sheets, match [1] – [6] to their appropriate words A–F.
Hint: It is known that the model tokenizes the word *coaches* as [*co*][*a*][*ch*][*es*].

J3. How does the model tokenize the word *started*?

Make sure you record your answers in your Answer Sheets!

(K) Speaking Your Truth (1/1) [15 Points]

Dene Sųłiné is a Na-Dene language spoken by around 12,000 people in and around Saskatchewan and Alberta, Canada. Below are some phrases in Dene Sųłiné and their English translations. It may help to know that the mark ´ above a vowel denotes high tone.

jíetthoghé	— *corn*	**erihtł'ístué**	— *ink*	**sjchéné**	— *my arm*
yak'é	— *window*	**dene tthí dhéth**	— *scalp*	**tu nédheli**	— *soup*
tł'izi	— *fly (insect)*	**chízechogh**	— *lion*	**tsąba dhéth**	— *wallet*
tuchogh	— *ocean*	**dzółkǫę**	— *billiards hall*	**dene sųłiné**	— *true person*

Below are some more phrases in Dene Sųłiné and their English translations, in arbitrary order.

1. **chizaze**
2. **dechenkálé**
3. **dene ké**
4. **dene láké**
5. **dene tthí nédheli**
6. **denetthíaze**
7. **dzółaze**
8. **erihtł'ís chené**
9. **jíaze**
10. **jíegaié**
11. **jíegaié dhéth**
12. **ké sųłiné**
13. **kéchogh**
14. **sa tł'ulé**
15. **setthíghá**
16. **sjla**
17. **tł'izi tthoghéchogh**
18. **tł'uk'etį**
19. **tł'uk'etįchogh**
20. **tł'ulekálé**
21. **tsąba delgai**
22. **tsąba deltthoghi**
23. **tsąbakǫ́ǫę**
24. **tuk'é**

A. *bank (business)*
B. *beans*
C. *boot*
D. *cat*
E. *fever*
F. *fingerprint*
G. *gold (metal)*
H. *guitar*
I. *hornet*
J. *marble* →
K. *moccasin*
L. *my hair*
M. *my hand*
N. *pen* →
O. *person's shoe*
P. *pod (plant casing)*
Q. *postage stamp*
R. *raisin*
S. *silver (metal)*
T. *strap*
U. *sunbeam*
V. *violin*
W. *water well*
X. *wooden board*

Notes: A *billiards hall* (or *pool hall*) is a place where people play games using long sticks to hit balls across special tables. A *moccasin* is a type of soft leather shoe traditionally worn by the Indigenous peoples of North America.

K1. In your Answer Sheets, match each Dene Sųłiné phrase to its English translation.

K2. A linguist asked for the English translation of **sa**, but was surprised to get the response *clock*. What did she expect, and what explains the response?

K3. Translate into English: **dechentué**, **denechogh**, **tł'izi tthoghé**.

K4. Below are the names of some animals in Dene Sųłiné and their English translations, in arbitrary order.

25. **dlíe**
26. **łue**
27. **łuechogh**
28. **sas delgai**
29. **yagolas**

AA. *blue whale*
BB. *butterfly*
CC. *fish*
DD. *polar bear*
EE. *squirrel*

Match each Dene Sųłiné animal to its English translation.

Make sure you record your answers in your Answer Sheets!

(L) Hide and Seek in Hunzib (1/2) [10 Points]

Hunzib is a Northeast Caucasian language spoken by approximately 3,500 people around Dagestan, Russia. Below are some Hunzib verbs in their singular and plural forms along with their English translations.

The plural form of a verb generally indicates that the verb's subject or object is plural, or that the action is performed frequently or repeatedly. Below are two sets of Hunzib verbs (under the tables we specify the differences between Set A and Set B). Some relevant pronunciation notes are listed on the next page. Throughout this problem, the abbreviation '*s.*' stands for someone, somewhere, or something.

Set A		
Singular	Plural	Translation
ãqáa	**ãqabáa**	*be thirsty*
áq'u	**ȧáq'u**	*bring s.*
čóqe	**čowáqe**	*get known to s.*
ék'e	**eyák'e**	*burn*
éže	**eyáže**	*take s. away*
əq'ə	**əwáq'e**	*lead s.*
gíč'	**giyáč'**	*slip into water*
gúše	**guwáše**	*touch s.*
háhel	**hȧhál**	*feed with milk*
íčox	**ĩčáx**	*stay*
ĩzáa	**ĩzabáa**	*wash s.*
k'ék'el	**k'ek'ál**	*get frightened*
λíše	**λiyáše**	*be sticky*
λɬ̃q'ə	**λɬ̃wáq'e**	*dress up*

Set B		
Singular	Plural	Translation
áhuk'	**ȧhák'**	*make s. take s. off*
ékle	**ekále**	*drop s.*
ézel	**ežál**	*chase s.*
əc'k'ə	**əc'ák'e**	*fill s.*
gúk'le	**guk'ále**	*put s. on to s.*
hálduk'	**haldák'**	*make s. white*
háldul	**haldál**	*be white*
ɬ̃c'lə	**ɬ̃c'ále**	*be renewed*
ɬ̃qəl	**ɬ̃qál**	*hide s.*
ɬq'k'ə	**ɬq'ák'e**	*grow s.*
ɬq'lə	**ɬq'ále**	*grow up*
ɬxk'ə	**ɬxák'e**	*warm s.*
λíšel	**λišál**	*glue s.*
rélek'	**relák'**	*straighten*

The difference between Set A and Set B involves two endings called the verbalizer and the causative: the verbs in Set A do not include them, while each verb in Set B includes exactly one of them. Here's how these two endings are defined:

- The verbalizer turns an adjective into a verb or changes the meaning of an existing verb. It has the forms **-l**, **-lə**, and **-le**.
- The causative is similar, but it adds another participant to the verb, usually meaning that one person or thing causes another to be or do something. It has the forms **-k'**, **-k'ə**, and **-k'e**.

L1. Give the plural forms for the Hunzib verbs **íλ̃'e**, **koxáa**, **λ̃'əq'ə**, **óč'ok'**, **ůq'e**, and **ůxel** (their meanings are not relevant for this task).

L2. Give the Hunzib plural forms of *dress s. up*, *put s. on*, *put the fire on*, *renew s.*, *take s. off*, and *warm up*.

L3. Explain your observations about how the verbal plural, causative, and verbalizer are formed in Hunzib.

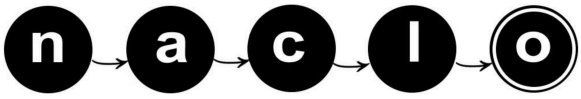
L4. Here are two more Hunzib verbs in the singular and plural:

Singular	Plural	Translation
úhle	**uwáhle**	*spoil s.*

Singular	Plural	Translation
gíč'k'e	**giyáč'k'e**	*pour s.*

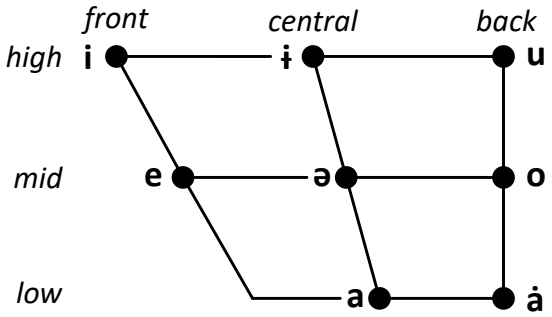
They do not follow the rules established previously. What plural forms would you have expected, and how might these irregular forms have arisen?

Make sure you record your answers in your Answer Sheets!



(L) Hide and Seek in Hunzib (2/2)

Pronunciation notes: the vowels used in this problem are illustrated in the diagram below. The words *high*, *mid*, *low*, *front*, *central*, and *back* refer to the position of the tongue in the mouth. Their exact meanings are not relevant for solving the problem.



c = *ch* in *chocolate*; **x** = *gh* in *ugh*. **q** is a **k**-like sound made further back in the throat. **š**, **ž**, and **č** are pronounced like **s**, **z**, and **c**, but further back in the mouth, with the tongue against the hard palate. **ƛ** and **ƛ̌** are pronounced like **s** and **c** but with air flowing along the side of the tongue.

The symbol **ʼ** indicates an ejective consonant (pronounced with a sharp popping sound). The symbol **˜** indicates a nasalized vowel (pronounced through the nose). The symbol ´ indicates a stressed vowel.

(M) Can You Speak Aikanã? (1/1) [15 Points]

Aikanã is a language isolate spoken by around 150 people in western Brazil. Below are some sentences in Aikanã and their English translations. It may be helpful to know that **y** = *y* in *yes*; **ü** ≈ **i**, but with the lips rounded; and the mark **˜** denotes a nasalized vowel (pronounced through the nose).

nãrüaepü damaẽ	—	*I recognize you and I feel jealous.*
apameana yoaneaẽ	—	*You find me and I smell good.*
ãwĩwãhena enumemeẽ	—	*He sleeps and you are angry with him.*
apaepü aweriaheẽ	—	*He sees you unexpectedly and he lies.*
apaena hapaẽ	—	*He sees you unexpectedly and you see him unexpectedly.*
ãrüakeana denumeẽ	—	*He remembers me and I hate him.*
amamepü haweriaaẽ	—	*You ban him and you trick me.*
enuena enumekeaẽ	—	*You are thirsty and he is angry with me.*
herekaapü nũpaneẽ	—	*I am warm and I dream weirdly.*
ãrüakaena apakeaẽ	—	*I remember you and he finds me.*

M1. In your Answer Sheets, translate into English:

1. **haweriana enuaẽ**
2. **yoanena henumeẽ**
3. **dapapü ũpanekaẽ**

M2. In your Answer Sheets, translate into Aikanã:

4. *I oversleep and I trick you.*
5. *He bans me and you dream.*
6. *You remember him and you are warm.*

M3. Explain your observations about how sentences are formed in Aikanã.

Make sure you record your answers in your Answer Sheets!

(N) A Token of Your Attention (1/2) [15 Points]

When computers process human language, the fundamental unit of language they use is usually the **token**, which can be anywhere from a single letter to a whole word. **In this problem, we will assume that tokens correspond to *morphemes*, or the smallest units of meaning in each word.** For example, we can represent the word *chasing* with a token for the verb *chase* and another for its suffix *ing*.

Modern approaches for processing language with computers use a mechanism called **attention**[1] to decide which words matter most when processing a sequence. NACLOLabs has trained a multilingual system that uses attention, and your task is to figure out how its attention works!

The cat eats the rat on the mat.

	the	*cat*	*eat*	*s*	*the*	*rat*	*on*	*the*	*mat*
the	.60	.05	.05	.05	.05	.05	.05	.05	.05
cat	.05	.35	.30	.05	.05	.05	.05	.05	.05
eat	.05	.15	.25	.20	.05	**.10**	.05	.05	.10
s	.05	.05	.40	.25	.05	.05	.05	.05	.05
the	.05	.05	.05	.05	.60	.05	.05	.05	.05
rat	.05	.05	**.20**	.05	.05	.45	.05	.05	.05
on	.05	.05	.05	.05	.05	.05	.60	.05	.05
the	.05	.05	.05	.05	.05	.05	.05	.60	.05
mat	.05	.05	.20	.05	.05	.05	.05	.05	.45

In this problem, attention is visualized as a matrix of decimals[2]. Each cell shows how much the word at the left of that row pays attention to the word at the top of that column. Attention has the following constraint placed on it: the values in each row must add up to 1. For example, in the sentence *The cat eats the rat on the mat*, shown in the matrix on the right, the word *eat* gives .10 (i.e. 10%) of its attention to *rat*, while *rat* gives .20 of its attention to *eat*. Because each row must sum to 1, attention has to be selective: it cannot give full weight to every word, so it must split its attention across the sequence.

Below is another attention matrix for the sentence *The dog was chasing the fast cat around the yard*. However, only the first five rows are visible!

The dog was chasing the fast cat around the yard.

	the	*dog*	*was*	*chase*	*ing*	*the*	*fast*	*cat*	*around*	*the*	*yard*
the	.60	.04	.04	.04	.04	.04	.04	.04	.04	.04	.04
dog	.04	.40	.04	.24	.04	.04	.04	.04	.04	.04	.04
was	.04	.04	.32	.32	.04	.04	.04	.04	.04	.04	.04
chase	.04	.12	.16	.20	.16	.04	.04	.08	.04	.04	.08
ing	.04	.04	.04	.32	.32	.04	.04	.04	.04	.04	.04

N1. In your Answer Sheets, write the attention values (as decimals) for the word pairs **(a)**–**(d)** from the sentence *The dog was chasing the fast cat around the yard*. Here, A → B stands for the amount of attention A gives to B; for example, *chase* → *dog* = .12 .

(a) *fast* → *fast* **(b)** *fast* → *dog* **(c)** *cat* → *cat* **(d)** *yard* → *chase*

N2. In your Answer Sheets, fill in the gaps **(e)**–**(i)** in the attention matrix on the right. You don't have to fill in the shaded cells.

The cat meowed.

	the	*cat*	*meow*	*ed*
the				
cat		.40	**(e)**	
meow		**(f)**	.40	**(g)**
ed			**(h)**	**(i)**

[1]The mechanism in this problem is more specifically referred to as **encoder self-attention**, where a sequence is compared to itself rather than to another sequence.

[2]This problem uses a simplified but consistent procedure for calculating attention values. All decimal values are rounded down.

Make sure you record your answers in your Answer Sheets!

(N) A Token of Your Attention (2/2)

Santali is an Austroasiatic language spoken by approximately 7.6 million people around Bengal, with a range that extends toward the foothills of the Himalayas. Santali is primarily written in Ol Chiki, a writing system developed by Raghunath Murmu in 1925.

NACLOLabs' attention mechanism works broadly similarly across Santali and English, placing attention on roughly analogous parts of the sequence. Below are six Santali sentences and their attention matrices, in arbitrary order.

However, two of the sentences have the exact same attention values, so they correspond to the same matrix! Fortunately, you are told that Santali sentence 4 has the same meaning as one of the prior three English sentences with an attention matrix.

1. ᱜᱤᱫᱤᱠᱮᱱ ᱢᱮᱱᱮᱞᱟᱵᱷ
2. ᱡᱟᱰᱤ ᱢᱟᱵᱷ ᱱᱛᱪᱵᱮᱡᱮᱟᱭ
3. ᱪᱛᱱᱟ ᱠᱮᱚᱟ ᱥᱛᱰᱮᱞᱟᱮ
4. ᱡᱤᱠᱨ ᱜᱤᱫᱤ ᱡᱟᱚᱨᱥᱮ ᱱᱛᱪᱮᱟᱮ
5. ᱠᱮᱚᱟ ᱡᱤᱠᱨᱠᱟᱩ ᱢᱮᱱᱷᱜᱟᱮ
6. ᱠᱮᱚᱟᱠᱟᱩ ᱥᱛᱰᱮᱞᱟᱭ

A.

						(j)	
	.352	.058	.058	.058	.352	.058	.058
	.058	.470	.058	.058	.235	.058	.058
	.058	.058	.470	.058	.235	.058	.058
	.058	.058	.058	.647	.058	.058	.058
	.176	.117	.117	.058	.235	.117	.176
(j)	.058	.058	.058	.058	.235	.470	.058
	.058	.058	.058	.058	.352	.058	.352

B.

						(k)
	.444	.055	.055	.333	.055	.055
	.055	.555	.055	.222	.055	.055
	.055	.055	.722	.055	.055	.055
	.166	.111	.055	.277	.222	.166
	.055	.055	.055	.444	.333	.055
(k)	.055	.055	.055	.333	.055	.444

C.

		(l)			
	.533	.066	.266	.066	.066
(l)	.066	.733	.066	.066	.066
	.133	.066	.333	.266	.200
	.066	.066	.533	.266	.066
	.066	.066	.400	.066	.400

D.

			(m)			
	.705	.058	.058	.058	.058	.058
	.058	.529	.235	.058	.058	.058
(m)	.058	.117	.294	.235	.117	.176
	.058	.058	.470	.294	.058	.058
	.058	.058	.235	.058	.529	.058
	.058	.058	.352	.058	.058	.411

E.

				(n)	
	.666	.083	.083	.083	.083
	.083	.417	.333	.083	.083
	.083	.166	.333	.250	.166
(n)	.083	.083	.500	.250	.083
	.083	.083	.333	.083	.417

N3. In your Answer Sheets, match each Santali sentence 1–6 with its corresponding attention matrix A–E. Remember that two sentences correspond to the same attention matrix!

N4. Identify the English sentence that has the same meaning as sentence 4.

N5. Fill in gaps **(j)** through **(n)** in the attention matrices with the correct Ol Chiki tokens in the corresponding Santali sentences. For the matrix that corresponds to two sentences, provide both tokens. You don't have to fill in the shaded cells.

N6. Explain your observations about how NACLOLabs calculates attention values between tokens in a sequence, regardless of which language.

Make sure you record your answers in your Answer Sheets!

(O) Sum Times in Na (1/1) [10 Points]

Na is a Nilo-Saharan language spoken by around 35,000 people in southern Chad. Below are some additions and multiplications in Na. Note that some Na words are written with dashes to indicate word parts.

$$\textbf{ko} + \textbf{ko} = \textbf{ko-jo} \quad (1)$$

$$\textbf{maka} \times \textbf{maka} = \textbf{ko bi-jɔ maka} \quad (2)$$

$$\textbf{jɔkɛ bi-jɔ mitikijo} \times \textbf{jo} = \textbf{ko-bi-jɔkɛ bi-jɔ sɔ} \quad (3)$$

$$\textbf{ko bi-jɔ mitikijo} + \textbf{ko-bi-jɔkɛ bi-jɔ muta} = \textbf{ko-muta} \quad (4)$$

$$\textbf{jɔkɛ bi-jɔ jo} \times \textbf{muta} = \textbf{ko-jo bi-jɔ maka} \quad (5)$$

$$\textbf{jɔkɛ-jo bi-jɔ maka} + \textbf{jɔkɛ-jo bi-jɔ mitikijo} = \textbf{ko-muta bi-jɔ muta} \quad (6)$$

O1. Write the equations (1)–(6) in digits.

O2. Write in digits: **ko-jo-bi-jɔkɛ bi-jɔ muta**.

O3. Translate into Na: 44, 117.

O4. Explain your observations about how numbers are formed in Na.

(P) Wear There's a Will (1/1) [15 Points]

Kalaallisut is an Inuit language spoken by around 50,000 people in western Greenland. Below are some sentences in Kalaallisut and their English translations. It may be helpful to know that **ʁ** denotes an **r** sound pronounced farther back in the mouth; **q** denotes a **k**-like sound made in the same place as **ʁ**; and **ɑ** and **ɛ** are vowels.

Flax is a crop used for both its nutritious seeds and its strong fibers which are used to make linen fabric. A *fiancé* is a man who is engaged to be married (as opposed to *fiancée*, who is a woman who is engaged to be married). *Raw oil* is oil in its natural state, while *processed oil* is oil that has been treated so it can be used more easily for things like fuel, cooking, or manufacturing.

nuliɑq annɔʁaamik iluɑʁsaassissɑɑq.	—	*The wife will mend a coat.*
nuliɑʁa pisussɑɑq.	—	*My wife will walk.*
uissɑq ɔʁsɔʁa katissavaa.	—	*The fiancé will collect my processed oil.*
palasiga sinissɑɑq.	—	*My priest will sleep.*
kunngiga qalissiamik tigumiɑʁsissɑɑq.	—	*My king will wear linen.*
palasissɑq viinnissamik katɛʁsissɑɑq.	—	*The religion student will collect a grape.*
palasima annɔʁɑɑʁa iluɑʁsaassavaa.	—	*My priest will mend my coat.*
uima kunngi isussavaa.	—	*My husband will criticize the king.*

P1. In your Answer Sheets, translate into English:

1. **nuliassɑq viinniga katissavaa.**
2. **palasissɑʁa annɔʁaassamik isɔʁsissɑɑq.**
3. **palasi ɔʁsussɑq tigumiassavaa.**

P2. In your Answer Sheets, translate into Kalaallisut:

4. *My husband will collect the raw oil.*
5. *My priest will collect flax.*
6. *My fiancé will criticize a crown prince.*

P3. Explain your observations about how sentences are formed in Kalaallisut.

Make sure you record your answers in your Answer Sheets!

(Q) Sea Saw (1/1) [15 Points]

Anindilyakwa is an East Arnhem language spoken by around 1,500 people in northern Australia. Below are some verb forms in Anindilyakwa and their English translations.

nangarringka	— *he saw her*
nanirringka	— *he saw him*
nginirringka	— *he saw me; he saw you*
ninginirringka	— *I saw him*
ningarrarringka	— *I saw them*
yingirringkatjingwa	— *she saw herself*
yinginirringka	— *she saw him*
yingimwarringka mwangwiyiwanga	— *she saw the shark*
narrangarringka	— *she saw them*
nanamwirringka mwangwiyiwanga	— *the shark saw him*
narramwirringka mwangwiyiwanga	— *the shark saw them*
yingakwirringka apwirtha	— *the whale saw her*
narrakwirringka apwirtha	— *the whale saw them*
narringarringka	— *they saw her*
ngipwirringka	— *they saw me; they saw you*
ningkwingarringka	— *you saw her*
ningkwirringka apwirtha	— *you saw the whale*
ningkwarrarringka	— *you saw them*

Q1. In your Answer Sheets, translate into English:

1. **narranirringka**
2. **narrirrarringka**
3. **ngikwirringka apwirtha**
4. **yingamwirringka mwangwiyiwanga**

If you think multiple translations are possible, write all of them.

Q2. In your Answer Sheets, translate into Anindilyakwa:

5. *I saw her*
6. *the shark saw you*
7. *the whale saw him*
8. *you saw yourself*

Q3. Explain your observations about how verbs are formed in Anindilyakwa.

Make sure you record your answers in your Answer Sheets!

The North American Computational Linguistics Open Competition
www.naclo.org

Answer Sheets

NACLO ID					

Name: _____

Contest Site: _____

Site ID: | | | | |

Grade: | |

Instructions for Filling Out the Answer Sheets

1. Write all answers clearly and in **dark, legible** writing. Because your responses will be scanned (and some may be graded automatically using optical character recognition software), faint or illegible responses cannot be graded. If your answers cannot be read, they may not be scored properly!
2. To change a response, **erase it cleanly**. If you cannot, clearly cross out the incorrect answer completely and write the new answer next to it.
3. For questions involving selecting one answer out of a list, fill in the circle for the **singular correct option** completely: ● Correct option ○ Incorrect option

   Examples of **INVALID** marks: ⊘ ⊗ ⊖ ⊙ ◶
4. For questions of the form "select all that apply", fill in the appropriate boxes for **all correct options** completely: ■ Correct option 1 □ Incorrect option ■ Correct option 2

   Examples of **INVALID** marks: ☑ ☒ ⊟ ⊡ ◪
5. For matching questions, write the **CAPITAL LETTER** of each correct match clearly beside its corresponding number: 1. A 2. C 3. B

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.

Signature: _____

YOUR NAME: NACLO ID: | | | | | | |

Answer Sheets (1/7)

(J) Strawbery Shortcake

J1. Starting from `small`, write the next 4 words of ParrotBot$_2$'s message on training data 2. (You shouldn't write the word `small`, which is already written for you.)

`small` | | | | |

J2. Write one capital letter (A–F) per box to match the stickers ① – ⑥ to their appropriate words:

① . | ② . | ③ . | ④ . | ⑤ . | ⑥ .

J3. Write the tokenization of *started* in the same format as in the problem. For example, *coaches* would be written as [*co*][*a*][*ch*][*es*].

started |

(K) Speaking Your Truth

K1. Write one capital letter (A–X) per box to match each Dene Sųłiné word or phrase to its English translation:

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

K2. What translation did the linguist expect for **sa**? |

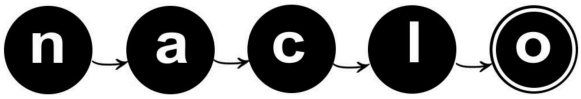
Explain why the speaker responded *clock* in 1–2 sentences:

K3. Translate each Dene Sųłiné word into English:

dechentué	**denechogh**	**tł'izi tthoghé**

K4. Write <u>two</u> capital letters (AA–EE) per box to match each Dene Sųłiné animal to its English translation:

25. | 26. | 27. | 28. | 29.



YOUR NAME: NACLO ID: | | | | | | |

Answer Sheets (2/7)

(L) Hide and Seek in Hunzib

L1. Give the plural forms for the following verbs:

íƛ'e		**koxáa**		**ƛ'áq'ə**	
óč'ok'		**ű́q'e**		**ű́xel**	

L2. Give the Hunzib plural forms of the following verbs:

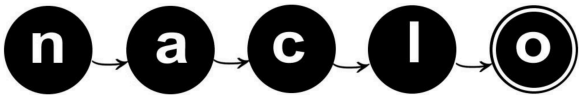
dress s. up		*put s. on*		*put the fire on*	
renew s.		*take s. off*		*warm up*	

L3. Explain your observations about how the verbal plural, causative and verbalizer are formed in Hunzib:

L4. Write the verbal plurals you would expect:

úhle **gíč'k'e**

Explain how the irregular forms might have arisen:



YOUR NAME: NACLO ID: | | | | | | |

Answer Sheets (3/7)

(M) Can You Speak Aikanã?

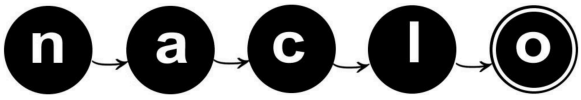
M1. Translate to English:

1. **haweriana enuaẽ**
2. **yoanena henumeẽ**
3. **dapapü ũpanekaẽ**

M2. Translate to Aikanã:

1. *I oversleep and I trick you.*
2. *He bans me and you dream.*
3. *You remember him and you are warm.*

M3. Explain your observations about how sentences are formed in Aikanã:



YOUR NAME: NACLO ID: | | | | | | |

Answer Sheets (4/7)

(N) A Token of Your Attention

N1. Write the attention values for the following word pairs **as decimals**, with two digits after the decimal point:

(a) *fast → fast* [ ] **(b)** *fast → dog* [ ] **(c)** *cat → cat* [ ] **(d)** *yard → chase* [ ]

N2. Fill in the gaps in the attention matrix for *The cat meowed*:

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

N3. Write one capital letter (A–E) per box to match each Santali sentence to its corresponding attention matrix:

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

N4. Fill in the circle corresponding to the English translation of sentence 4:

○ *The cat eats the rat on the mat.* ○ *The dog was chasing the fast cat around the yard.* ○ *The cat meowed.*

N5. Fill in the gaps in the attention matrices with the correct Ol Chiki tokens. **Please write clearly!** For the matrix that corresponds to two sentences, write both tokens:

(j)

(k)

(l)

(m)

(n)

N6. Explain your observations about how NACLOLabs calculates attention weights between tokens in a sequence, regardless of which language:

YOUR NAME: NACLO ID: | | | | | | |

Answer Sheets (5/7)

(O) Sum Times in Na

O1. Write the equalities (1)–(6) in digits:

(1) ☐ + ☐ = ☐

(2) ☐ × ☐ = ☐

(3) ☐ × ☐ = ☐

(4) ☐ + ☐ = ☐

(5) ☐ × ☐ = ☐

(6) ☐ + ☐ = ☐

O2. Write in digits: **ko-jo-bi-jɔkɛ bi-jɔ muta** = ☐

O3. Write in Na:

44 = ☐

117 = ☐

O4. Explain your observations about how numbers are formed in Na:

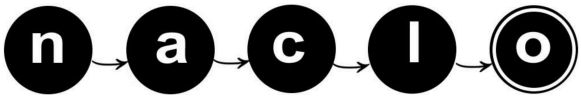
(P) Wear There's a Will

P1. Translate to English:

1. **nuliassɑq viinniga katissavaa.**

2. **palasissɑʁa annɔʁaassamik isɔʁsissɑɑq.**

3. **palasi ɔʁsussɑq tigumiassavaa.**



YOUR NAME: NACLO ID: | | | | | | |

Answer Sheets (6/7)

(P) Wear There's a Will (continued)

P2. Translate to Kalaallisut:

4. *My husband will collect the raw oil.*

5. *My priest will collect flax.*

6. *My fiancé will criticize a crown prince.*

P3. Explain your observations about how sentences are formed in Kalaallisut:

YOUR NAME: NACLO ID: | | | | | | |

Answer Sheets (7/7)

(Q) Sea Saw

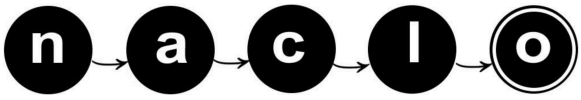
Q1. Translate into English in all possible ways:

1. **narranirringka**
2. **narrirrarringka**
3. **ngikwirringka apwirtha**
4. **yingamwirringka mwangwiyiwanga**

Q2. Translate into Anindilyakwa:

5. *I saw her*
6. *the shark saw you*
7. *the whale saw him*
8. *you saw yourself*

Q3. Explain your observations about how verbs are formed in Anindilyakwa:



YOUR NAME: NACLO ID: | | | | | | |

Additional Answer Space (1/1)

If you use this additional space, please do both of the following:

1. *On this sheet, clearly indicate which question(s) you are answering. E.g., write "Problem Q4."*
2. *In the regular answer space(s) for the question(s) you are answering, note that you are using additional answer space. E.g., at the end of the answer space for problem Q4 in the regular Answer Sheets, add a note saying "See additional answer space."*