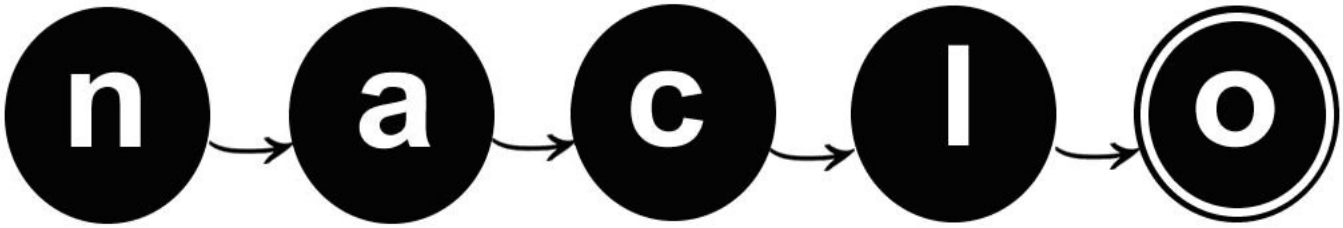


NACLO thanks the following for their generous contributions:

Carnegie Mellon University
Language Technologies Institute

NAACL

The Nineteenth Annual

North American Computational Linguistics Open Competition

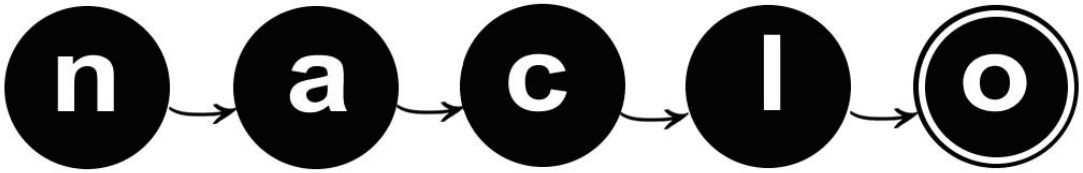
2025

www.naclo.org

Open Round
January 23, 2025

Serious language puzzles that are surprisingly fun!

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



Welcome to the nineteenth 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 three hours long and includes eight problems, labeled A to H.
2. Follow the facilitators' instructions carefully.
3. If you want clarification on any of the problems, talk to a facilitator. The facilitator will consult with the jury before answering.
4. You may not discuss the problems with anyone except as described in items 3 & 11.
5. Each problem is worth a specified number of points, with a total of 100 points. In this year's Open Round, no points will be given for explanations. Instead, make sure to fill out all the answer boxes properly.
6. All your answers should be written clearly in the Answer Sheets at the end of this booklet. ONLY THE ANSWER SHEETS WILL BE GRADED.
7. Write your name and registration number on each page of the Answer Sheets. Here is an example: Jessica Sawyer #850
8. The top 10% of participants (approximately) across the United States and Anglophone Canada in the Open Round will be invited to the Invitational 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! 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 2025 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

Logistics Committee:

Nathan Dummitt (Social Media Subchair) — Columbia Grammar and Preparatory School
Elena Filatova (Outreach and University Subchair) — City University of New York Graduate Center
Henry Gustafson — California Institute of Technology
Shuli Jones — Google
Arul Kolla — Massachusetts Institute of Technology
Lori Levin — Carnegie Mellon University
Daniel Lovsted (Canada Coordinator) — University of Ottawa
Cerulean Ozarow (High School Subchair) — Brown University
Luke Sawczak (Website Subchair) — Toronto District Christian High School
Marten van Schijndel (Grading Subchair) — Cornell University
Gloria Yip — State University of New York at Stony Brook

Logistics Committee Co-Chairs:

Shuli Jones — Google
Cerulean Ozarow — Brown University

Program Committee:

Kunaal Chandrashekar — University of Toronto
Ryan Chi — Stanford University
Brandon Cho — Princeton University
Andy Hou — Princeton University
Ken Jiang — University of Waterloo
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
Benjamin Yang — California Institute of Technology

NACLO 2025 Organizers and Credits (2/2)

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

Booklet Editor:
Daniel Lovsted

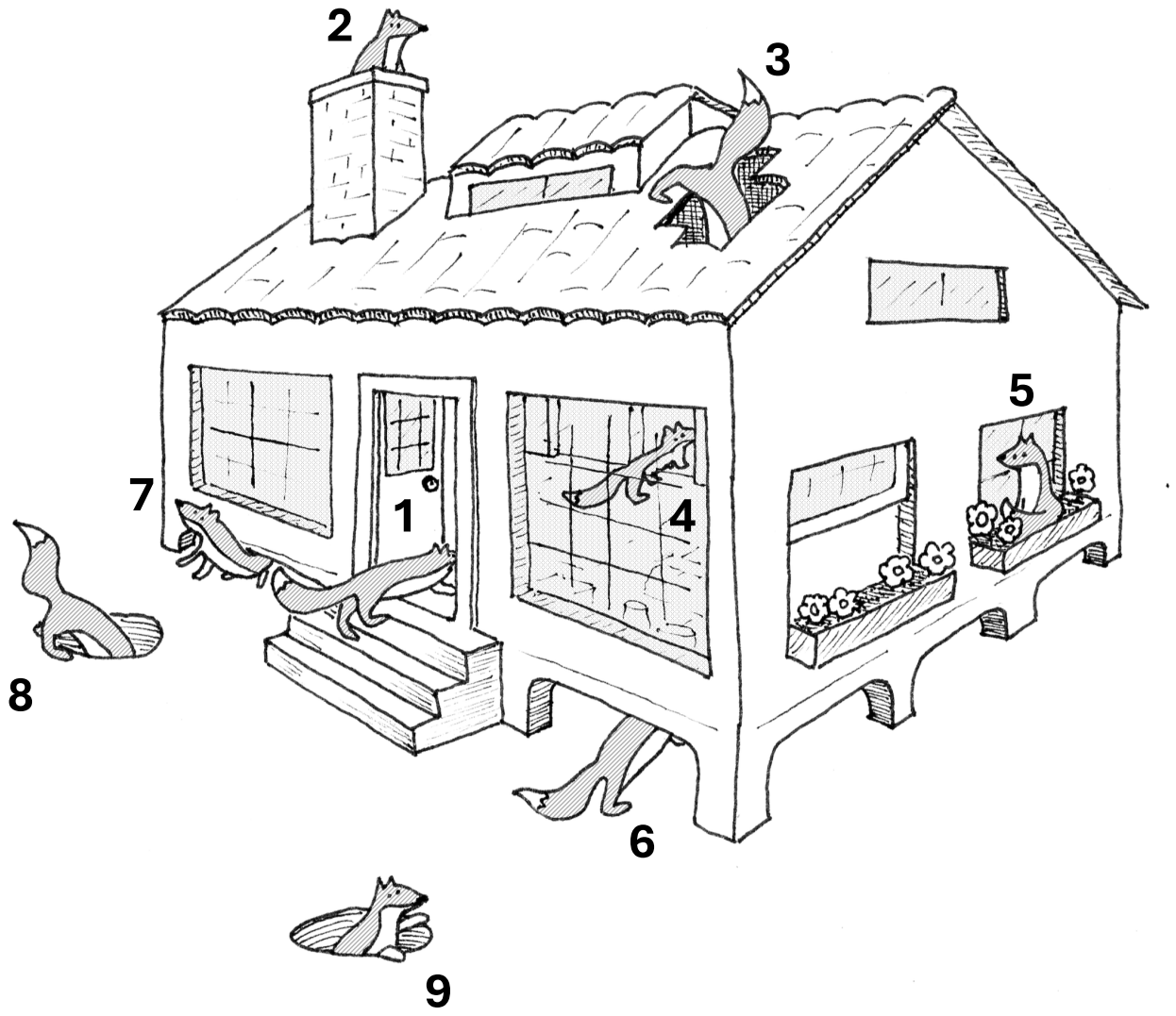
Problem Credits:
(A) Osbel Lopez-Francisco, Lori Levin, and Patrick Littell
(B) Celeste Tan
(C) Simi Hellsten and Andrew Lamont
(D) Daniel Lovsted
(E) Michael Salter
(F) Tom McCoy
(G) David Hellsten
(H) Arul Kolla

Image Credits:
The image in Problem A was created by Patrick Littell for use in this booklet. The image in Problem H was created by Pixabay user OpenClipart-Vectors and is licensed under the Pixabay Content License.

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

(A) Fox out of the Box (1/2) [10 Points]

Zorro the fox is looking for his friend Mapache the raccoon, who likes to hang out in houses where humans live. The following picture shows the fox's search through the house and its surroundings.

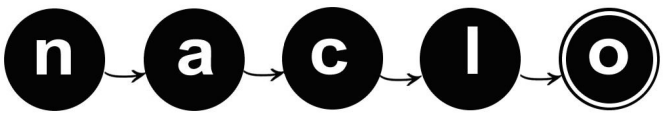


On the next page, sentences in Totonac describe the numbered phases of the search.

Although we've given English descriptions to clarify the drawings, don't assume that the Totonac sentences are direct translations of the English sentences. The sentences in both languages describe the fox's journey, but different languages can focus on different aspects of a scene.

Notes:

- In the Totonac sentences, a colon "**:**" indicates a long vowel. The word "**n**" is added to sentences based on phrasal prosody (making the sound flow well). It doesn't have a meaning (but you are still responsible for using it correctly in your answers).
- This problem features the variety of Totonac spoken in the town of Zongozotla, Mexico.



(A) Fox out of the Box (2/2)

1. **Tanu:yi kchiki wa:qni'**
The fox goes into the house through the front door.

2. **Akpu:tukuta n chiki wa:qni'**
The fox comes out of the chimney.

3. **Akpu:tanu:yi n chiki wa:qni'**
The fox enters the house through a hole in the roof.

4. **Cha:qe:taxtiyi n chiki wa:qni'**
The fox comes out the back door.

5. **Lakpa:tuju:yi n chiki wa:qni'**
The fox goes into a flower box on the side of the house.

6. **Tampi:tanu:yi n chiki wa:qni'**
The fox goes under the house.

7. **Tampi:taxtiyi n chiki wa:qni'**
The fox comes out from under the house.

8. **Tuju:yi ktaqalhwaxni wa:qni'**
The fox goes into a hole.

9. **Tukuta ktaqalhwaxni wa:qni'**
The fox comes out of a hole.

A1. On your Answer Sheets, construct Totonac sentences describing each of these events:

10. *The fox comes out the front door.*

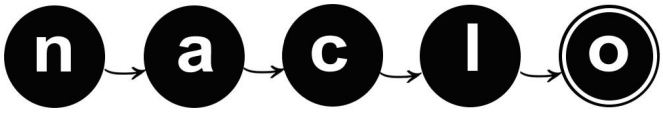
11. *The fox goes into the chimney.*

12. *The fox comes out a hole in the roof.*

13. *The fox comes in the back door.*

14. *The fox goes into a window on the side of the house.*

15. *The fox comes out of a flower box on the back of the house.*



(B) Swidden in Plain Sight (1/2) [10 Points]

Kri is a Vietic language spoken in Laos, with about 250 speakers as of 2009. Below are some words in Kri, alongside their English translations. Kri vowels come in two registers, heavy and light, but for simplicity this distinction is not marked in this problem.

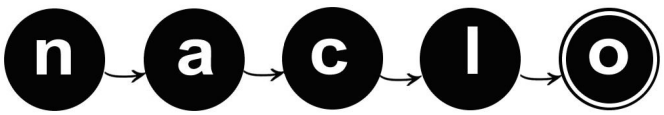
blang	*(of the eyes) to become open*[1]
balang	*to open one's eyes*
plaajh	*an arm span*[2]
palaajh	*to measure something by arm spans*
cabuuc	*to pick up or measure out purse-hand*[3]
truum	*to be on its face*
paang	*to signal ownership of a swidden*[4]
ckaang	*a hand span*[5]
basơt	*to turn off, put out*
kooq	*to live*
bsơt	*to go out (as a light or fire does)*
cakaang	*to measure something by hand spans*
cooh	*to pierce*
krnooq	*a house*
prnaang	*a sign placed in a swidden to show ownership*

B1. On your Answer Sheets, give the Kri word for each of the following:

a. *a unit of measure equal to an amount that can be picked up purse-hand*
b. *to put something on its face*
c. *a hanging thread, pierced through many objects*

(problem continues on the next page)

[1] As an example, this verb can be used to describe the point a few weeks into a puppy's life, when its eyes open for the first time.
[2] An *arm span* is the length of a person's outstretched arms, from fingertips to fingertips.
[3] *Purse-hand* means to put the thumb and other four fingers together in a pinching motion.
[4] A *swidden* is an agricultural field that has been cleared by cutting and burning vegetation.
[5] A *hand span* is the length of a person's hand.



(B) Swidden in Plain Sight (2/2)

Here are six more Kri words. Their English meanings are given below, in a random order.

A. **kaleer** B. **kleer** C. **koorq** D. **krneep** E. **krnoorq** F. **paraang**

1. *to scoop out strips of a vegetable for food preparation*
2. *to fall off (as a fruit does)*
3. *an instrument to scoop out strips of a vegetable*
4. *to take someone across (e.g., across a river)*
5. *tongs, pincers*
6. *to pick something off (e.g., to pick a berry off of a bush)*

B2. On your Answer Sheets, match the Kri words (A to F) with their English translations (1 to 6).

B3. A Kri word in this problem can be modified to translate exactly one of the English phrases below. Choose that phrase and give the Kri word that translates it.

7. *a river*
8. *to put someone into a river*
9. *to cross over*
10. *a cross*

(C) Color Me Confused (1/3) [15 Points]

The information on this page is needed for both problem (C) and problem (D). Otherwise, problems (C) and (D) are not related — you don't need to solve either one in order to solve the other.

Languages and their speakers have many ways of describing colors — sometimes by comparing to well-known materials (*copper hair*, *mustard wallpaper*); sometimes using very specific terms (*crimson ribbon*); and sometimes using very broad terms (*gray-blue sky*). English has 11 "basic color terms": *white*, *gray*, *black*, *red*, *orange*, *yellow*, *green*, *blue*, *purple*, *pink*, and *brown*. However, other languages can have as few as two or as many as 12.

Below is a table showing eight objects and the basic color word that native speakers of 15 different languages might use to describe them.

	ripe plum	*ripe strawberry*	*ripe banana*	*fresh grass*	*clear sky*	*deep ocean*	*adult crow*	*fresh snow*
Bassa	**hui**	**ziza**	**ziza**	**hui**	**hui**	**hui**	**hui**	**ziza**
Bété	**kp**	**zeli**	**zeli**	**kp**	**kp**	**kp**	**kp**	**fee**
Ejagham	**ebi**	**ebi**	**ebi**	**enyaga**	**enyaga**	**enyaga**	**enyaga**	**ebare**
Karajá	**iso**	**iso**	**tyre**	**tyre**	**tyre**	**tyre**	**ilyby**	**ura**
Matsés	**piu**	**piu**	**piu**	**umu**	**umu**	**umu**	**chëshe**	**ushu**
Apinaye	**grãgrã**	**kamrẽk**	**ràràr**	**grãgrã**	**grãgrã**	**grãgrã**	**tyk**	**aka**
Tsafiki	**luban**	**luban**	**laqueban**	**losimban**	**fiban**	**paban**	**paban**	**fiban**
Seri	**kʔe'el**	**kʔe'el**	**kwassool**	**ko'il**	**ko'il**	**ko'il**	**k'ooppool**	**ko'oxp**
Guambiano	**piguig**	**piguig**	**uscuig**	**chilga**	**pillig**	**pillig**	**yalig**	**polig**
Teribe	**dindin**	**sresren**	**shoylor**	**keson**	**dindin**	**dindin**	**sisi**	**pluplun**
Tlapanec	**minuu**	**mana**	**mojmo**	**maxa**	**maxa**	**minuu**	**skuni**	**mixa**
English	**purple**	**red**	**yellow**	**green**	**blue**	**blue**	**black**	**white**
Yakan	**taluk**	**peat**	**binaning**	**gaddung**	**bilu**	**bilu**	**ittem**	**pole**
Amuzgo	**tsjan'chi**	**wee**	**cajan**	**cachuii'**	**tsa**	**tsjo'**	**ntom**	**canchii'**
Hebrew	**sagol**	**adom**	**tsahov**	**yarok**	**t'khelet**	**kakhol**	**shakhor**	**lavan**

Information about these languages, and other languages that appear in problem (C), can be found on page (3/3) of problem (C). Note that the language information is not relevant for solving either problem (C) or (D).

(C) Color Me Confused (2/3)

In 1969, the anthropologist Brent Berlin and linguist Paul Kay proposed a set of rules about basic color terms. Five of them are given below in a simplified form, with the objects (*ripe plum*, *ripe strawberry*, etc.) replaced with the letters (A) to (H). Note that a word is described as *unique* if it is used to describe only one of the eight objects.

1. Every language has at least two basic color terms, which distinguish between (A) and (B).

2. If it has at least three, it uses different words for each of (A), (B) and (C). If it has only three, the same word is used for both (B) and (D).

3. If it has at least five, it distinguishes between (B) and (D), and also has a unique word for (E).

4. If it has at least six, it distinguishes (D) from (F).

5. A language only has a unique word for (G) if it also has one for (H). If it does not have a unique word for (H), (G) will share a word with at least one of (A), (D) or (F).

Although these rules are no longer believed to be completely true for all languages, they do still describe most languages accurately. In particular, they describe all 15 languages in the table on the previous page.

C1. Determine which object each of the letters (A) - (H) refers to. Each of the eight objects corresponds to exactly one letter.

Below is a similar table for six other languages.

	ripe plum	*ripe strawberry*	*ripe banana*	*fresh grass*	*clear sky*	*deep ocean*	*adult crow*	*fresh snow*
Kwerba	**nokonum**	**nokonum**	**kainanesenum**	**icem**	**asiram**	**icem**	**icem**	**asiram**
Lele	**wile**	**wile**	**bore**	**bore**	**bore**	**induwa**	**induwa**	**bole**
Ocaina	**tsipi**	**tsiiho**	**boora**	**moxooso**	**moxooso**	**moxooso**	**jutohfu**	**furaana**
Waorani	**opa'tawae**	**opa'tawae**	**na'amae'ta**	**we'nedae**	**we'nedae**	**we'nedae**	**wintamo**	**na'amae'ta**
Yucuna	**ipirani**	**querani**	**jehuani**	**ipureni**	**ipureni**	**ipureni**	**cameni**	**jareni**
Zapotec	**nu-bizat**	**nu-na'**	**nu-cach**	**nu-ca'**	**nu-ca'**	**nu-asul**	**nu-cas**	**nu-cacy**

C2. On your Answer Sheets, for each language, indicate whether its color words follow rules 1-5 (YES) or not (NO) by selecting the correct answer.

(problem continues on the next page)

(C) Color Me Confused (3/3)

Below is another similar table, but with one word in each row left blank.

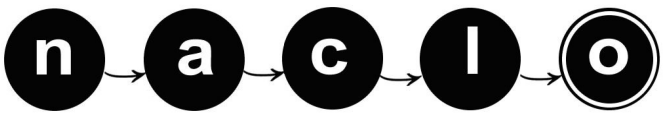
	ripe plum	*ripe strawbery*	*ripe banana*	*fresh grass*	*clear sky*	*deep ocean*	*adult crow*	*fresh snow*
Nafaanra	**nye**	**nye**	**a.**	**wɔ**	**wɔ**	**wɔ**	**wɔ**	**finge**
Tucano	**sõa'ro**	**sõa'ro**	**ewuro**	**b.**	**ya'saro**	**ya'saro**	**ñiro**	**butiro**
Kalam	**likañ**	**c.**	**walin**	**minj-kimemb**	**muk**	**muk**	**mosimb**	**tund**

C3. For each of the blanks **a.** - **c.**, choose all possible words from the list of options that the language might use to refer to that object, while still following rules 1-5. Make sure to write your answers on your Answer Sheet. In each case, the options are all the words shown in the table in that language.

- **a.** **nye**, **wɔ**, **finge**
- **b.** **sõa'ro**, **ewuro**, **ya'saro**, **ñiro**, **butiro**
- **c.** **likañ**, **walin**, **minj-kimemb**, **muk**, **mosimb**, **tund**

Information on the languages featured in problem (C) and problem (D) is below. Note that this information is not relevant for solving either problem.

Amuzgo, Tlapanec, and Zapotec are clusters of closely related Otomanguean languages spoken in Mexico by around 60,000, 150,000, and 500,000 people, respectively. Apinaye and Karajá are Macro-Jê languages spoken in Brazil by around 2000-3000 people each. Bassa and Bété are Kru languages. Bassa is spoken by around 600,000 people in Liberia, Ivory Coast and Sierra Leone. Bété is spoken by several thousand people in Ivory Coast, although the dialect used in this problem had only 50 remaining speakers in 1992. Ejagham and Nafaanra are Atlantic-Congo languages. Ejagham is part of the Bantoid subgroup, and is spoken by around 120,000 people in Nigeria and Cameroon. Nafaanra is part of the Senufo subgroup, and is spoken by around 90,000 people, mostly in Ghana. English is an Indo-European language with around 400 million native speakers, and close to 1.5 billion total speakers worldwide. Guambiano and Tsafiki are Barbacoan languages. Guambiano is spoken by around 20,000 people in Colombia. Tsafiki is spoken by around 2000 people in Ecuador. Hebrew and Lele are Afro-Asiatic languages. Hebrew is from the Semitic subgroup, and was extinct as a spoken language between the 5th and 19th century CE, although it continued to be used for religious purposes; now it has around 5 million native speakers, and around 9 million total speakers, primarily in Israel. Lele is from the East Chadic subgroup, and has around 26,000 speakers in Chad. Kalam is a Trans-New-Guinea language, spoken by around 15,000 people in Papua New Guinea. Kwerba is a Greater Kwerbic language spoken by around 2500 people in West Papua. Matsés is a Panoan language spoken by around 2000 people on the Peru-Brazil border. Ocaina is a Huitotoan language, spoken by around 50 people along the Peru-Colombia border. Seri is a linguistic isolate spoken by 700-1000 people in Mexico. Teribe is a Chibchan language spoken by around 3000 people in Panama and Costa Rica. Tucano is a Tucanoan language, spoken by around 12,000 people in Brazil and Colombia. Waorani is a linguistic isolate spoken by around 2000 people in Ecuador; more speakers may be living in uncontacted groups in Peru. Yakan is an Austronesian language, spoken by around 100,000 people in the Philippines. Yucuna is an Arawakan language, spoken by nearly 2000 people along the Miritiparaná river in Colombia.



(D) Let's Pivot! (1/1) [10 Points]

For this problem, you will need to use the information on page (1/3) of problem (C). Otherwise, problems (C) and (D) are not related — you don't need to solve either one in order to solve the other.

Imagine that you know a way to translate reliably from language *A* to language *B*, and also a way to translate from *B* to language *C*. Then you also know how to translate from *A* to *C*: just translate from *A* to *B* first, and next from *B* to *C*. When a machine translation system uses this strategy, *B* is called a *pivot language*.[1]

NACLO Labs is about to release its groundbreaking new app, *NacloTranslate*. *NacloTranslate* can translate between more language pairs than any of its competitors, and it does so by using pivot languages.

Unfortunately, *NacloTranslate* doesn't always translate totally reliably. In a study, NACLO Labs scientists tried translating **ebi** from Ejagham to Seri, using Karajá as a pivot language. They hoped to see the app give only the answers **kʔe'el** and **kwassool**, which are the results that could be produced from direct (no pivot) translation. But it was observed producing each of these three results on different test runs:

Ejagham ⇄ Seri	
ebi	**kʔe'el**

Ejagham ⇄ Seri	
ebi	**kwassool**

Ejagham ⇄ Seri	
ebi	**ko'il**

D1. Below are three more *NacloTranslate* test runs, from a different study. A pivot language was used for exactly one of the three language pairs—which one? Give a possibility for what the pivot language could be.

A.

Tlapanec ⇄ Matsés	
minuu	**umu**

B.

Yakan ⇄ Guambiano	
ittem	**chilga**

C.

Hebrew ⇄ Teribe	
kakhol	**dindin**

D2. NACLO Labs is considering its options for having *NacloTranslate* translate **paban** from Tsafiki to English.

a. What are the results of a direct translation? Select all possible answers.

b. What are the results if Karajá is used as a pivot? Select all possible answers.

c. NACLO Labs selects a different pivot language, and the results for **paban** are **green**, **blue**, and **black** (and no others). Which languages could NACLO Labs be using as pivot? Give all three possible answers.

D3. Here are six test runs of *NacloTranslate*, from a different study. From the options below, identify the pivot language used in each test run by matching options A-F to test runs 1-6. Each option was used once.

1.

Guambiano ⇄ Tlapanec	
piguig	**skuni**

2.

Tlapanec ⇄ Amuzgo	
minuu	**tsa**

3.

Seri ⇄ Yakan	
kʔe'el	**taluk**

4.

Matsés ⇄ Seri	
piu	**ko'il**

5.

Amuzgo ⇄ Matsés	
cajan	**ushu**

6.

Yakan ⇄ Guambiano	
gaddung	**piguig**

Options: A. Apinaye B. Bassa C. Bété D. English E. Teribe F. Tsafiki

[1] This strategy can be useful when training data is hard to find. Modern machine translators may need to be trained on huge amounts of *parallel text* — that is, text in one language given alongside its translation in another (much like in some NACLO problems!). Sometimes, *A*-*C* parallel text will be hard to find, but *A*-*B* and *B*-*C* parallel text will be much more common.

(E) Around the World in Armenian (1/1) [15 Points]

Armenian, the official language of Armenia, is part of the Indo-European language family, the language family that contains most European languages like English and Russian, as well as many South Asian languages like Punjabi and Hindi. Its alphabet, developed in the early 5th century CE, is also related to the Latin alphabet we use for English, since both descend from an archaic form of the Greek alphabet.

Below are the names of several of the world's countries in the Armenian script, and their English equivalents in a random order. Note that some of the names of countries are slightly different in the Armenian language.

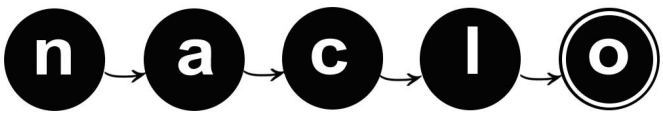
1.	Ֆրանսիա	A.	Canada
2.	Գերմանիա	B.	Australia
3.	Կանադա	C.	Israel
4.	Իտալիա	D.	America
5.	Իսրայել	E.	Peru
6.	Բուլղարիա	F.	Russia
7.	Արգենտինա	G.	England
8.	Ավստրալիա	H.	Bulgaria
9.	Ֆինլանդիա	I.	Libya
10.	Ռուսաստան	J.	France
11.	Անգլիա	K.	Germany
12.	Պերու	L.	Finland
13.	Ամերիկա	M.	Romania
14.	Լիբիա	N.	Italy
15.	Ռումինիա	O.	Argentina

E1. On your Answer Sheets, match up the English and Armenian words by writing one letter (A to O) per box.

E2. What are the names for "England" and "Russia" in Armenian? Use the Latin (English) alphabet to write how you think they are pronounced.

E3. Write the names of the following countries in the Armenian script. Make sure to follow the instructions on your Answer Sheet for writing clearly! The names of these countries are pronounced the same in Armenian as in English.

a. Iran **b.** Pakistan **c.** Cameroon **d.** Burundi



(F) Not Quite Right (1/2) [10 Points]

A chatbot is a computer program that is designed to have conversations with people. NacloCorp has recently developed a chatbot, and it is your job to test it out.

First, you ask NacloChat to decode five sentences written in a cipher, where each letter has been replaced by a different letter. The table below shows what you found, but some cells in the table are blank.

Cipher text	Correct answer	NacloChat's answer
IPVFDDN ZSJ DPQJ SFT BLQJ	**a.**	finally the time has come
ZSJN BSFVZJG ZSJ LIIPBPFD ULDPBN	**b.**	they changed the official policy
SJ TZYIIJG ZSJ ULTZJW PVTPGJ F ZYEF	**c.**	he stuffed the poster inside a tube
ZSJN TFPDJG FXFN PV F ELLZ	**d.**	**e.**
XJ QYTZ EJ APVG ZL JFBS LZZJW	**f.**	**g.**

F1. On your Answer Sheets, fill in the blanks **a.** to **g.**

Hints:

- The same cipher has been used in all five sentences. For example, if Q stands for m in the first sentence, it will stand for m in all other sentences too.
- In each of the five sentences, NacloChat got exactly one letter wrong.

Next, you ask NacloChat to form acronyms from the names of several organizations. An acronym is created by combining the first letters from all the words in a sequence of words. The table on the next page shows what you find.

F2. On your Answer Sheets, fill in the cells **h.** to **l.** that are missing from the table on the next page.

Hints:

- There is one case where NacloChat gets the right answer. In all other cases, it gets exactly one letter wrong.
- The "Organization" column is provided for fun, but it is not necessary for solving the problem.

(F) Not Quite Right (2/2)

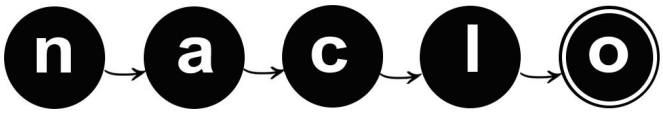
Organization	Correct answer	NacloChat's answer
Waiting Room Enthusiasts Network	WREN	WHEN
Architects Building Office Units Near Deserts	ABOUND	AROUND
Seriously Creative Hat Manufacturers Organizational League	SCHMOL	SCHOOL
Society Incentivizing Lateness — Earlybirds Not Allowed!	SILENA	SILENT
People Improving Tomorrow's Yesterday	PITY	CITY
Annoyingly Friendly Thimblemakers — Northern Region	AFTNR	**h.**
Vancouver Eggplant Relocation Bureau	VERB	**i.**
Pseudoscientists Embracing Obviously Counterfeited Evidence	PEOCE	**j.**
Filmmakers Reimagining Old Memoirs	FROM	**k.**
Western Indiana Typewriter Society	WITS	**l.**

Finally, you ask NacloChat to count how many words there are in some lists of words. For example, given the list "elephant jacket spaghetti mattress," NacloChat should answer 4. You observe that it sometimes gets the answer wrong.

F3. Below are some questions about three of the cases that NacloChat answered incorrectly. On your Answer Sheets, select the correct letter (A or B) for each one. *Note: To save some space, we are not providing the lists that were given to NacloChat — we are just saying how long each list was.*

1. When the length of the list was 91, what number did NacloChat produce? A. 90 B. 92
2. When the length of the list was 59, what number did NacloChat produce? A. 58 B. 60
3. When the length of the list was 74, what number did NacloChat produce? A. 73 B. 75

In the past few years, many chatbots — powered by a type of technology called large language models (LLMs) — have been developed that can perform impressively well on a wide range of tasks. For example, they can summarize text and write computer code. However, despite their impressive performance, there are certain types of situations where chatbots often have trouble. This problem illustrates one type of issue that many chatbots display.



(G) The Little Dog That Could Read (1/1) [15 Points]

Sakata (also known as Kisakata or Kesakata) is a language in the Bantu family spoken by several thousand people in the Democratic Republic of the Congo. It consists of several quite distinct dialects; this problem features the dialect spoken in the village of Bosobe in the 1980s.

Below are some Sakata sentences, alongside their English translations. Note that ə and ü are both vowels, and that one Sakata noun is slightly irregular.

1.	**Banka balülü bamwi bebil benini.**	*The good women saw the big fields.*
2.	**Banjir babibə badzo ibat lelülü.**	*The bad men killed the good duck.*
3.	**Keto ketita kejo ntaw ne banjir.**	*The long snake heard the men's goat.*
4.	**Mabat masinxunə ola mokan mo munjir.**	*The ducks wanted to eat the man's book.*
5.	**Mva nekikə ne banka nejao otang.**	*The women's little dog learned to read.*
6.	**Nə ala ipe le ibat lenini?**	*Who ate the big duck's wing?*
7.	**Munkar nejo nko ne bang.**	*The woman heard the children's chicken.*
8.	**Nko ibibə ila ncher i bang.**	*The bad chickens ate the children's peanuts.*
9.	**Bande bachunga ntaw ilülü.**	*The white men bought the good goats.*
10.	**Ntaw i banka ikwi o kebil ke munde.**	*The women's goats went to the white man's field.*
11.	**Nə amwi bena be bande?**	*Who saw the white men's things?*
12.	**Ngombe nekikə nemwi lecher lenini.**	*The little cow saw the big peanut.*

G1. On your Answer Sheets, translate into English:

- **a.** **Mva imwi ntaw nebibə.**
- **b.** **Munjir nekikə netang mekan melülü.**
- **c.** **Nə adzo beto be munkar?**

G2. Translate into Sakata:

- **a.** *The cows heard the big dogs.*
- **b.** *The white man learned to eat wings.*
- **c.** *The snakes of the field read the big books.*

Below are some numerical phrases in Sakata, alongside their English translations.

ledze lemo	*one finger*
ipyao lemo	*one shoulder*
nchungə ipə	*two pigs*
befwa besa	*three bones*
mete meni	*four trees*

G3. Translate into Sakata:

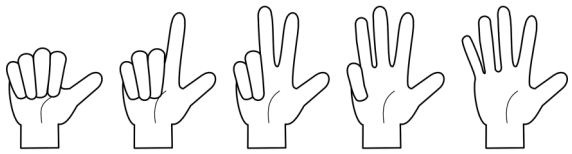
a. *one pig* **b.** *one tree* **c.** *one bone* **d.** *two shoulders* **e.** *three fingers*

(H) There Is Another Hand (1/2) [15 Points]

Most languages spoken in Europe today are Indo-European, brought to Europe by waves of migration 3000 to 8000 years ago. Before this, various other languages were spoken, such as the language of Minoan Crete, famous for its undeciphered writing system known as Linear A. Of these old European languages, only Basque survives, being spoken by around 800,000 people in northern Spain and southern France.

In the same way that French and Spanish descend from Latin, modern-day Basque descends from an older language called Proto-Basque. Although there are no speakers of Proto-Basque left, linguists can use historical methods to attempt to reconstruct what words in Proto-Basque may have looked like. For example, here is one proposal for the Proto-Basque numbers 1-9, <u>given in a random order</u>. Note that since these words are reconstructed, rather than known for certain, they are preceded by *****.

A. ***bada-eri**	B. ***berr-eri-ahur**	C. ***atz-az-berr-eri**
D. ***be-oro-atz**	E. ***berr-eri**	F. ***atz-az-bada-eri-ahur-itzi**
G. ***atz-az-berr-eri-ahur-itzi**	H. ***atz-az-eri**	I. ***eri-ahur**



Each of the Proto-Basque numbers are made of smaller word-parts, with meanings related to counting on your hand. For example, ***berr-eri** literally means *another finger.*

Below are the word-parts of Proto-Basque numbers, with their English translations <u>given in a random order</u>. Note that the meanings of ***be** and ***itzi** are not relevant (linguists have other reasons to believe that they were present).

i.	***ahur**	J.	*another*
ii.	***atz**	K.	*finger*
iii.	***az**	L.	*hand*
iv.	***bada**	M.	*(in) palm (of hand)*
v.	***berr**	N.	*there is*
vi.	***eri**	O.	*with*
vii.	***oro**	P.	*whole*

(H) There Is Another Hand (2/2)

H1. On your Answer Sheets, determine the correct correspondences between the Proto-Basque words A to I on the previous page and the numbers from 1 to 9.

H2. Determine the correct correspondences between the Proto-Basque word-parts i. to vii. and their English translations J to P on the previous page.

Below are the modern Basque numbers 5-10 in a northern dialect, given again in a random order.

Q.	**zazpi**	R.	**bortz**	S.	**zortzi**
T.	**bederatzi**	U.	**hamar**	V.	**sei**

H3. Determine which of the modern Basque numbers Q to V above correspond to each of the numbers from 5 to 10. It may be helpful to know that the Basque **z** is very similar to **s** (although pronounced with the tongue touching the teeth), and that **p** and **b** are similar consonants.

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

Answer Sheets

REGISTRATION NUMBER					

Name: _____

Contest Site: _____

Site ID: _____

City, State/Province: _____

Grade: _____

Please also make sure to **write your registration number and your name on each page of the Answer Sheets**, and **turn in all pages of the Answers Sheets** even if you have left some blank.

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

Signature: _____

YOUR NAME: REGISTRATION #

Answer Sheets (1/5)

(A) Fox out of the Box

A1. Construct a Totonac sentence to describe each of these events:

10. *The fox comes out the front door.*

11. *The fox goes into the chimney.*

12. *The fox comes out a hole in the roof.*

13. *The fox comes in the back door.*

14. *The fox goes into a window on the side of the house.*

15. *The fox comes out of a flower box on the back of the house.*

(B) Swidden in Plain Sight

B1. Give the Kri word for each of the following:

a. *a unit of measure equal to an amount that can be picked up purse-hand*

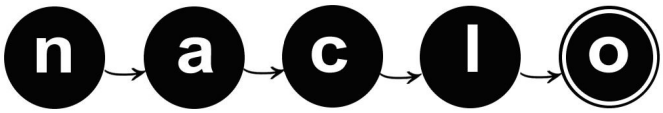
b. *to put something on its face*

c. *a hanging thread, pierced through many objects*

B2. In each box, write the letter (A-F) of the Kri word corresponding to that English phrase:

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

B3. Number of the English phrase (7-10): Kri word that translates it:



YOUR NAME: REGISTRATION #

Answer Sheets (2/5)

(C) Color Me Confused

C1. Next to each object, write the letter (A-H) that replaces it in rules 1 to 5:

ripe plum: ☐	*ripe strawberry:* ☐	*ripe banana:* ☐	*fresh grass:* ☐
clear sky: ☐	*deep ocean:* ☐	*adult crow:* ☐	*fresh snow:* ☐

C2. For each language, select YES if it follows rules 1-5, and select NO if it does not. Make your selection by filling in the bubble just to the left of the option you want to select:

Kwerba	◯ YES	◯ NO	Lele	◯ YES	◯ NO
Ocaina	◯ YES	◯ NO	Waorani	◯ YES	◯ NO
Yucuna	◯ YES	◯ NO	Zapotec	◯ YES	◯ NO

C3. For each blank, select all words from the given list that could fill it, assuming that each language follows rules 1-5. Make your selections by filling in the box just to the left of the word:

a. ☐ **nye** ☐ **wɔ** ☐ **finge**

b. ☐ **sõa'ro** ☐ **ewuro** ☐ **ya'saro** ☐ **ñiro** ☐ **butiro**

c. ☐ **likañ** ☐ **walin** ☐ **minj-kimemb** ☐ **muk** ☐ **mosimb** ☐ **tund**

(D) Let's Pivot!

D1. Which pair used a pivot? Write A, B, or C: ☐

Give one possible pivot language, from the 15 options on page (1/3) of problem (C): ☐

D2. **a.** Select all possible answers. Make your selections by filling in the box just to the left of the word:

☐ **purple** ☐ **red** ☐ **yellow** ☐ **green** ☐ **blue** ☐ **black** ☐ **white**

b. Select all possible answers. Make your selections by filling in the bubble just to the left of the word:

☐ **purple** ☐ **red** ☐ **yellow** ☐ **green** ☐ **blue** ☐ **black** ☐ **white**

c. Write all three language names, in any order, choosing from the 15 options on page (1/3) of (C):

☐

Problem (D) continues on the next page.

YOUR NAME: REGISTRATION #

Answer Sheets (3/5)

(D) Let's Pivot! (continued)

D3. Write one letter (A-F) per box to match each language option to the test run in which it was the pivot:

1. ☐ 2. ☐ 3. ☐ 4. ☐ 5. ☐ 6. ☐

(E) Around the World in Armenian

E1. Write one letter (A-O) per box to match the English country name to the Armenian country name:

1. ☐ 2. ☐ 3. ☐ 4. ☐ 5. ☐
6. ☐ 7. ☐ 8. ☐ 9. ☐ 10. ☐
11. ☐ 12. ☐ 13. ☐ 14. ☐ 15. ☐

E2. Use the English alphabet to write how these country names are pronounced in Armenian:

England: ☐ Russia: ☐

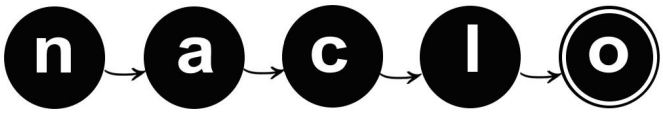
E3. Write the names of the following countries in the Armenian script. **Please write clearly!** Use the dotted lines provided as a guide: the tops and/or bottoms of some Armenian symbols should touch the dotted line, and other symbols should have parts that extend above or below the dotted line. As an example, here is one of the country names from the problem, as it would be written within the dotted lines:

Տրանսիա

Your writing does not need to be picture-perfect, but it does need to clearly distinguish the symbol you are writing from other Armenian symbols. The dotted lines should help you do this.

a. Iran: ☐ **b.** Pakistan: ☐

c. Cameroon: ☐ **d.** Burundi: ☐



YOUR NAME: REGISTRATION #

Answer Sheets (4/5)

(F) Not Quite Right

F1. Fill in the blanks **a.** to **g.** in the table below:

Cipher text	Correct answer	NacloChat's answer
IPVFDDN ZSJ DPQJ SFT BLQJ	**a.**	finally the time has come
ZSJN BSFVZJG ZSJ LIIPBPFD ULDPBN	**b.**	they changed the official policy
SJ TZYIIJG ZSJ ULTZJW PVTPGJ F ZYEF	**c.**	he stuffed the poster inside a tube
ZSJN TFPDJG FXFN PV F ELLZ	**d.**	**e.**
XJ QYTZ EJ APVG ZL JFBS LZZJW	**f.**	**g.**

F2. For each correct answer below, write NacloChat's answer:

AFTNR **h.** VERB **i.** PEOCE **j.**

FROM **k.** WITS **l.**

F3. Select A or B to answer each question by filling in the bubble just to the left of your choice:

1. When the length of the list was 91, what number did NacloChat produce? ◯ A. 90 ◯ B. 92
2. When the length of the list was 59, what number did NacloChat produce? ◯ A. 58 ◯ B. 60
3. When the length of the list was 74, what number did NacloChat produce? ◯ A. 73 ◯ B. 75

(G) The Little Dog That Could Read

G1. Translate into English:

a. **Mva imwi ntaw nebibə.**

b. **Munjir nekikə netang mekan melülü.**

c. **Nə adzo beto be munkar?**

Problem (G) continues on the next page.

YOUR NAME: REGISTRATION #

Answer Sheets (5/5)

(G) The Little Dog That Could Read (continued)

G2. Translate into Sakata:

a. *The cows heard the big dogs.*

b. *The white man learned to eat wings.*

c. *The snakes of the field read the big books.*

G3. Translate into Sakata:

a. *one pig*

b. *one tree*

c. *one bone*

d. *two shoulders*

e. *three fingers*

(H) There Is Another Hand

H1. Write one <u>number</u> (1-9) per box to translate each Proto-Basque word:

A. B. C. D. E.

F. G. H. I.

H2. Write one letter (J-P) per box to give the meaning of each word-part:

i. ii. iii. iv. v. vi. vii.

H3. Write one <u>number</u> (5-10) per box to translate each modern Basque word:

Q. R. S. T. U. V.