

NACLO thanks the following for their generous contributions:



***The Seventeenth
Annual***

**North American
Computational
Linguistics
Open
Competition**



2023

www.nacloweb.org



Carnegie Mellon University
Language Technologies Institute



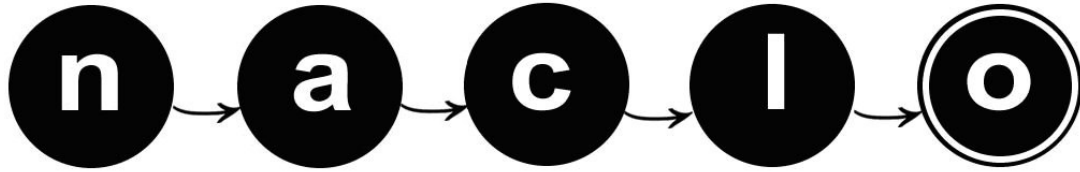
Yale University

Open Round

January 26, 2023

Serious language puzzles that are surprisingly fun!

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



Welcome to the seventeenth 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. Note that this year's contest has one fewer problem than in some previous years.
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 Anglo-phone 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 2023 Organizers

NACLO Co-Chairs:

Aleka Blackwell — Middle Tennessee State University
Lori Levin — Carnegie Mellon University

Anglophone Canada NACLO Coordinator:

Daniel Lovsted — University of Ottawa

Organizing Committee:

Aleka Blackwell — Middle Tennessee State University
Robin Hammer — Carnegie Mellon University
Shuli Jones — Massachusetts Institute of Technology
Riley Kong — Massachusetts Institute of Technology
Lori Levin — Carnegie Mellon University
Yixin Liu — Yale University
Daniel Lovsted — University of Ottawa
Margarita Misirpashayeva — Duolingo
Cerulean Ozarow — Brown University
Dragomir Radev — Yale University

Program Committee Chair:

Tom McCoy — Princeton University

Program Committee Area Chairs:

Ethan Chi — Hudson River Trading
Jane Li — Johns Hopkins University
Daniel Lovsted — University of Ottawa

Program Committee:

Aleka Blackwell — Middle Tennessee State University
Timothy Blackwell — Encore Capital Group
Ellie Bultena — Massachusetts Institute of Technology
Ethan Chi — Hudson River Trading
Ryan Chi — Stanford University
Riley Kong — Massachusetts Institute of Technology
Pranav Krishna — Massachusetts Institute of Technology
Ben LaFond — Harvard University
Lori Levin — Carnegie Mellon University



NACLO 2023 Organizers (cont'd)

Program Committee (continued):

Jane Li — Johns Hopkins University

Daniel Lovsted — McGill University

Benjamin McAvoy-Bickford — University of Pennsylvania

Tom McCoy — Princeton University

Cerulean Ozarow — Brown University

Dragomir Radev — Yale University

Tom Roberts — University of Amsterdam

USA University Site Coordinators:

Brigham Young University — Deryle Lonsdale

California State University Dominguez Hills — Iara Mantenuto

California State University Long Beach — Michael Ahland

Carnegie Mellon University — John Friday, Robin Hammer, Lori Levin

College of William and Mary — Anya Hogoboom

Columbia University — Daniel Bauer, Kathleen McKeown

Cornell University — Sam Tilsen, Marten van Schijndel

Emory University — Jinho Choi

Florida Atlantic University — Viktor Kharlamov

Fort Hays State University — Destiny Gu, Sherri Matlock

Georgetown University — Janet Liu

Georgia Institute of Technology — Hongchen Wu

Indiana University — Sandra Kuebler

Johns Hopkins University — Liz Salesky

Massachusetts Institute of Technology — Pranav Krishna

Middle Tennessee State University — Aleka Blackwell

Minnesota State University Mankato — Becky Bates

Montclair State University — Lauren Covey, Anna Feldman, Jonathan Howell

Northeastern Illinois University — Lewis Gebhardt, Jill Hallett

Northwestern University Chicago — Suhuai Chen

Ohio State University — Micha Elsner, Marie de Marneffe

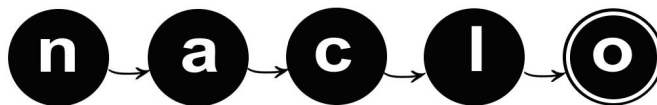
Princeton University — Christiane Fellbaum, Misha Khodak, Tom McCoy, Oliver Weizel

Rollins College — Margarita Azbel

Stanford University — Kelly Battcher, Ryan Chi, Matthew Early, Emily Lake, Daisy Leigh,

Stanford University (continued) — Eli Strauss-Reis, George Wang

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



NACLO 2023 Organizers (cont'd)

USA University Site Coordinators (continued):

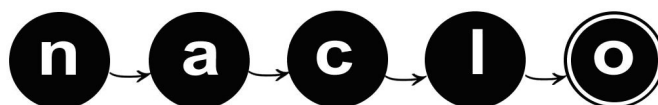
Union College — Nick Webb
University at Buffalo — Jeff Good, Cassandra Jacobs, Leslie Ying
University of Arizona — Sandiway Fong
University of California, Berkeley — Pranav Jayachand, Andres Sanchez, Cor Zanda
University of California, Los Angeles — Meghana Gorti, Noel Guzman, Jennifer Miyaki
University of California, San Diego — Akil Iyer, Eric Meinhardt, Will Styler
University of Florida — Sarah Moeller
University of Maryland — Alex Krauska, Jan Michalowski, Sigwan Thivierge
University of Michigan — Marcus Berger, Ezra Keshet, Varun Ponnusamy, Sally Thomason
University of North Carolina at Charlotte — Seethalakshmi Gopalakrishnan, Hossein Hematialam,
University of North Carolina at Charlotte (continued) — Kodzo Wegba, Wlodek Zadrozny
University of Notre Dame — David Chiang
University of Pennsylvania — Chris Callison-Burch, Anne Cocos, Cheryl Hickey, Mitch Marcus,
University of Pennsylvania (continued) — Oliver Sayeed, Derry Wijaya
University of South Carolina — Mila Tasseva-Kurktchieva
University of South Florida — Jenny Qin
University of Southern Maine — Claire Holman, Dana McDaniel
University of Texas at Dallas — Jing Lu, Vincent Ng, Gerardo Ocampo Diaz
University of Utah — Aniello De Santo, Karen Marsh Schaeffer
University of Washington — Jim Hoard, Joyce Parvi
University of Wisconsin, Milwaukee — Joyce Boyland, Gabriella Pinter, Anne Pycha
Western Washington University — Kristin Denham
Wichita State University — Jill Fisher, Mythili Menon
Yale University — Raffaella Zanuttini

Canada University Site Coordinators:

McGill University — Lola Bradford, Lisa Travis, Michael Wagner
Trinity Western University / Canada Institute of Linguistics — Anita Lebold, Isaac Munnalall
Tyndale University / Canada Institute of Linguistics — Tom Scott
University of British Columbia — Kaili Vesik
University of Ottawa — Andrés Pablo Salanova
University of Saskatchewan — Zhi Li, Bettina Spreng
University of Toronto — Ellina Zhang

Special thanks to:

The hosts of the 100+ High School Sites



NACLO 2023 Organizers and Credits (cont'd)

Booklet Editor:

Tom McCoy — Princeton University

Problem Credits:

(A) Cynthia Allen

(B) Ellie Bultena

(C) Tom McCoy

(D) Pavel Iosad

(E) Malcolm Kennedy

(F) Ryan Chi

(G) Ben LaFond

Consultants: Jonathan Amith and Ben Leeming

(H) Michael Salter

Contest graders:

Ugonna Ahumibe, Meesum Alam, Aleka Blackwell, Timothy Blackwell, Matthew Fort, Sarah Hicks, Umida Hikmatilla, Aidan Gabriel Iacobucci, Ken Jiang, Shuli Jones, Lily Kawaoto, Snigdha Khanna, Riley Kong, Sandra Kuebler, Ben LaFond, Heidi Lei, Jane Li, Daniel Lovsted, Koosha Maleknia, Tom McCoy, Margarita Misirpashayeva, Margarita Gabrielle Morrison, Larry Moss, Cerulean Ozarow, William Pan, Rishab Parthasarathy, Robert Pugh, Dragomir Radev, Tom Roberts, Noah Sauer, Em Singleton, Corinne Soucy, Lily Stella, Daniel Swanson, Evelyn Sun, Rhodes Talaga, Ariel Wang, Yi-Chung Wang, Sydney Weber, Audrey Yang, Ellen Zhang

We are grateful to our problem authors and consultants for their expertise. Any errors remain our own.

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

All material in this booklet © 2023, North American Computational Linguistics Open Competition and the authors of the individual problems. Please do not copy or distribute without permission.



(A) The Way We Were (1/4) [15 Points]

Languages are constantly changing. For example, when you talk to your friends, you probably use some words that did not exist ten years ago. Even the basic grammar of a language can change.

Over time, these changes accumulate until new languages are born. In this way, the English that is spoken today—known as Present Day English—is descended from a language called Old English, which was spoken approximately a thousand years ago. Since that time, the language has transformed so much that, if you wanted to understand Old English, you would need to study it as a foreign language. This problem focuses on some of the differences between Old English and Present Day English.

Study the table below, which contains some sentences in Old English along with their translations in Present Day English.¹ Then answer the questions on the next three pages.

	Old English	Present Day English
1.	Æpelwine and Wulfstan gesawon þas menn. Hi hi gretton.	Athelwine and Wulfstan saw these men. They greeted them.
2.	Nereþ hit us?	Does it save us?
3.	Se cyning and se bisceop and ic gesawon þone arcebisceop. We hine gretton.	The king and the bishop and I saw the archbishop. We greeted him.
4.	Hwæt cweþe we?	What do we say?
5.	Seo lar eow nereþ.	The teaching saves you.
6.	We cumað.	We come.
7.	Ne here we þas wifmen.	We do not hear these women.
8.	þas weras hereað þas wifmen. Ne gesawon hi hi.	These men hear these women. They did not see them.
9.	þu hierdest þas menn.	You heard these men.
10.	þa herde we þe.	Then we heard you.
11.	þu þe hierdest, and we us hierdon.	You heard yourself, and we heard ourselves.
12.	Alfred cyning and se bisceop ridon into þære byrig. þa gesawe we hi.	King Alfred and the bishop rode into the town. Then we saw them.
13.	Eadweard hine geseah. Gesawe ge us?	Edward saw himself. Did you see us?
14.	Hwæt fremest þu?	What do you perform?
15.	Fremede we hit?	Did we perform it?

1. The pronunciation of the letters is not important for this problem, but see the end of the problem for pronunciation notes.



(A) The Way We Were (2/4)

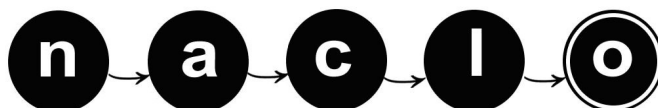
A1. On your Answer Sheets, fill in the blank cells from the table below:

	Old English	Present Day English
16.	(a)	archbishop
17.	(b)	King Edward
18.	menn	(c)
19.	weras	(d)
20.	Ne hierde ge hi.	(e)
21.	Ne gesawon hi þe.	(f)
22.	Hi hit fremedon.	(g)
23.	Ne nerede we þe.	(h)

A2. On your Answer Sheets, fill in the blanks from these Old English sentences. A blank can be filled in with one word or a sequence of multiple words. For some blanks, there may be more than one solution; you should only give one.

	Old English	Present Day English
24.	Ne grette ge ____ (a) ____ .	You did not greet them.
25.	Seo lar hi ____ (b) ____ .	The teaching saves them.
26.	þu ____ (c) ____ .	You heard it.
27.	þa ____ (d) ____ .	Then they heard us.

Make sure you record your answers in your Answer Sheets!



(A) The Way We Were (3/4)

Here are some more Old English sentences:

	Old English	Present Day English
28.	Wit hieraþ þæt word and wit hit cwepaþ.	We hear the word and we say it.
29.	Se cyning and ic gesawon þas boceras, ac ne grette wit hi. þas boceras unc gretton.	The king and I saw these scribes but we did not approach them. These scribes greeted us.
30.	þonne cume wit to his rice.	Then we come to his kingdom.
31.	Æpelbald and ic eow gesawon, ac ne hierde wit eow. Wit eow gretton.	Athelbald and I saw you, but we didn't hear you. We greeted you.
32.	Edmund and Swiþun, ne hiere git þas wifmen, ac wit hi hieraþ. þas wifmen inc hieraþ.	Edmund and Swithun, you do not hear these women, but we hear them. These women hear you.

A3. Below on the left are six Old English sentences which can all be translated into Present Day English by the single sentence *We saw you*. On your Answer Sheets, match each Old English sentence with the context (on the right) in which it might have been said.

	Old English
i.	Wit eow gesawon.
ii.	Wit inc gesawon.
iii.	We eow gesawon.
iv.	We inc gesawon.
v.	We þe gesawon.
vi.	Wit þe gesawon.

	Context
A.	I said " We saw you " to tell your people that my people and I had seen them.
B.	Mother said " We saw you " to tell me that she and my father had seen me.
C.	One bishop said " We saw you " to tell the archbishop that all the bishops had seen him.
D.	I said " We saw you " to tell my sister and brother that my mother and I had seen them.
E.	Athelbald said " We saw you " to tell Swithun and Edmund that he and his men had seen them.
F.	The king said " We saw you " to tell the people that he and his queen had seen them all.

Make sure you record your answers in your Answer Sheets!



(A) The Way We Were (4/4)

A4. Six distinct words are used to translate the Present Day English word *you* into Old English. On your Answer Sheets, for each of the following sentences, write the correct *you* word that is missing.

	Old English	Present Day English
33.	Ne hierdest ___(a)___ hi.	(Speaking to Paul) You didn't hear them.
34.	Ne hierde ___(b)___ hi.	(Speaking to Alfred & Edmund) You didn't hear them.
35.	Æpelwine and Wulfstan, gesawon hi ___(c)___ ?	Athelwine and Wulfstan, did they see you?
36.	Hierde ___(d)___ hine?	(Addressing the crowd) Did you hear him?
37.	Paulus, gretton hi ___(e)___ ?	Paul, did they greet you?
38.	Gretton hi ___(f)___ ?	(Speaking to Alfred, Athelwine, and Paul) Did they greet you?

Some more information about Old English: If someone referred to Old English, would you think of the language of Shakespeare, who was active at the end of the sixteenth century and the beginning of the seventeenth, or possibly of Geoffrey Chaucer's *Canterbury Tales*, written circa 1385? In the standard historical division of English, Shakespeare's language is actually Early Modern English, while Chaucer was writing late Middle English. English was brought to England from Germany and northern Denmark in the early fifth century, and before the conversion of the English to Christianity, was occasionally written in inscriptions using the Germanic runic alphabet. A small number of English writings in the Latin alphabet appear circa 700, but it is only in the ninth century that we find a substantial body of texts. The Old English (also known as Anglo-Saxon) period extended to circa 1100, when enough changes to the language had happened that it would be unintelligible to the people who brought the language to England. If Old English would have been incomprehensible to the speakers of the Middle English period, this is even more true for Present Day English speakers.

Additional notes:

- For the most part, Latin letters were used to represent similar sounds in Old English. The letter <æ> (which is capitalized as <Æ>) represented a vowel similar to that in Present Day English *hat*. The symbol <þ>, called *thorn*, comes from the runic alphabet and was used to represent sounds not found in Latin, the two sounds of <th> in Present Day English *this* and *thigh*.
- Word order in Old English was flexible. In your answers for this problem, you should only use the word orders illustrated in the example sentences.



(B) Don't Stress (1/1) [10 Points]

Ho-Chunk is a language in the Siouan family (an Indigenous language family of North America) spoken by 250 people in Iowa, Wisconsin, and Nebraska, USA. This problem is about the stress system of Ho-Chunk.

Stress systems give emphasis to some syllables over others. In the English word *desert* (the place where a cactus grows), the first syllable is stressed. In the English word *dessert* (a sweet treat after dinner), the second syllable is stressed. There are different levels of stress that a syllable can have. In this problem, we indicate levels of stress by writing letters under syllables: The most strongly-stressed syllable in a word is indicated with P (short for *primary stress*), other stressed syllables are indicated with S (short for *secondary stress*), and syllables with no stress are indicated with N. A word can only have one primary stress, but it can have multiple secondary stresses or unstressed syllables.

For example, the English word *kangaroo* would be written as shown on the right: Its strongest stress is on the final syllable, and it has a secondary stress on the first syllable. kangaroo
 $\begin{array}{c} \text{S} \quad \text{N} \quad \text{P} \\ \text{---} \end{array}$

Below are some words in Ho-Chunk annotated for stress along with their English translations. The English translations are not necessary for solving this problem. Some Ho-Chunk syllables contain two vowels. If the two vowels are the same (such as **aa**), the syllable has a long vowel sound. If the vowels are different (such as **ai**), then the syllable combines two distinct vowel sounds. Ho-Chunk uses a few symbols not present in English: **j** and **q** are vowels, and **g** is a consonant.

$\begin{array}{c} \text{hokewe} \\ \text{N} \quad \text{N} \quad \text{P} \\ \text{---} \end{array}$	“enter”	$\begin{array}{c} \text{waakit'e} \\ \text{N} \quad \text{P} \quad \text{N} \\ \text{---} \end{array}$	“speak to”
$\begin{array}{c} \text{wanjigra} \\ \text{N} \quad \text{N} \quad \text{P} \quad \text{N} \\ \text{---} \end{array}$	“the bird”	$\begin{array}{c} \text{haahere} \\ \text{N} \quad \text{P} \quad \text{N} \\ \text{---} \end{array}$	“last night”
$\begin{array}{c} \text{hokiwaroke} \\ \text{N} \quad \text{N} \quad \text{P} \quad \text{N} \quad \text{S} \\ \text{---} \end{array}$	“swing”	$\begin{array}{c} \text{bookagaja} \\ \text{N} \quad \text{P} \quad \text{N} \quad \text{S} \\ \text{---} \end{array}$	“knock down obviously”
$\begin{array}{c} \text{hakirujikgaja} \\ \text{N} \quad \text{N} \quad \text{P} \quad \text{N} \quad \text{S} \quad \text{N} \\ \text{---} \end{array}$	“after he pulls taut”		

B1. On your Answer Sheet, indicate the stress for the Ho-Chunk words below. There are nine stresses in total (counting both primary and secondary stress).

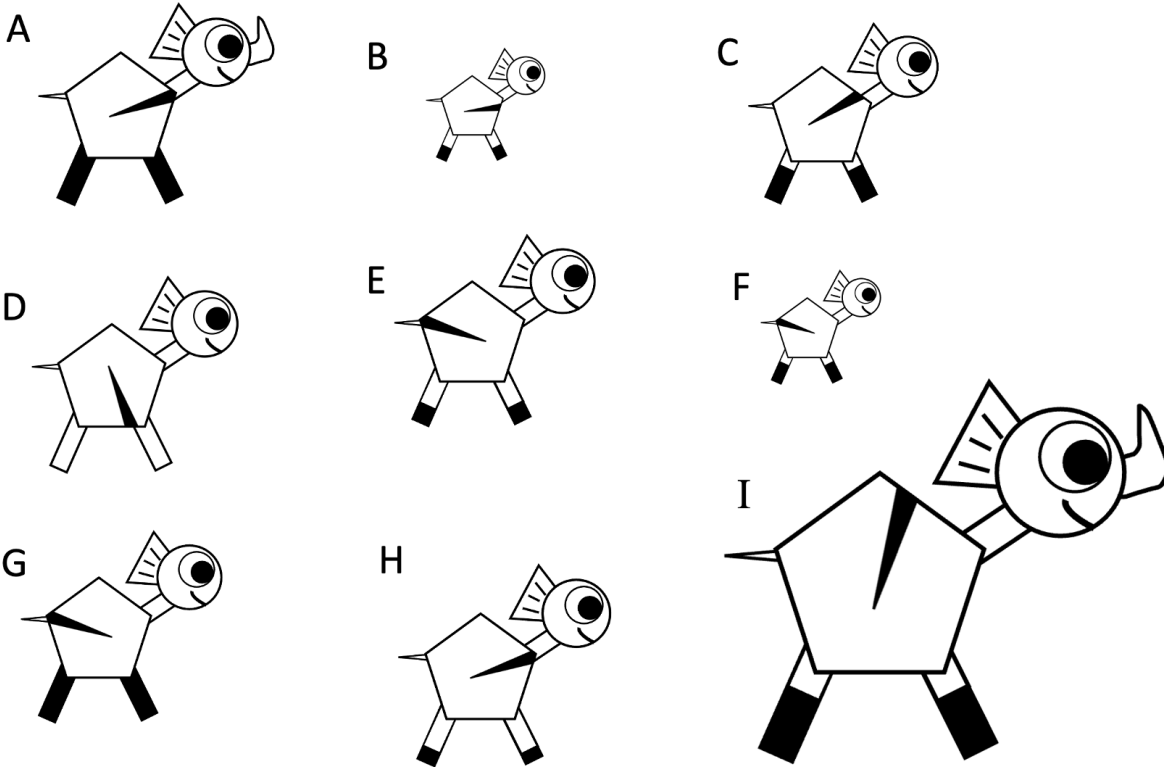
$\begin{array}{c} \text{hipirak} \\ \text{---} \end{array}$	“belt”	$\begin{array}{c} \text{waaporoporo} \\ \text{---} \end{array}$	“snowball”
$\begin{array}{c} \text{xorojike} \\ \text{---} \end{array}$	“hollow”	$\begin{array}{c} \text{hirakorohonira} \\ \text{---} \end{array}$	“the fact that you did not get ready”
$\begin{array}{c} \text{waiperesga} \\ \text{---} \end{array}$	“linen”		

Make sure you record your answers in your Answer Sheets!



(C) Vloxes (1/2) [15 Points]

Exciting news: You've gotten clearance to visit Mars! Your job there is to take care of some Martian animals called vloxes. Here are the nine vloxes that you are in charge of:



Your manager asks you to identify which vloxes fit the following descriptions:

- The two mellivu vloxes
- The two mellovu vloxes
- The two kibb vloxes
- The two azeltorf vloxes
- The qlikli vlox
- The imxo vlox
- The three ravv vloxes
- The kibbish-zuff vlox
- The zuffish-kibb vlox

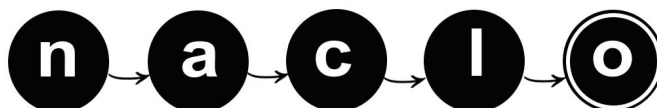
Unfortunately, many of these words are unfamiliar to you, and you don't have a dictionary! Luckily, you do have an encyclopedia article (on the next page) that uses these words. Even though the article doesn't define the words directly, you can figure out a lot about a word based on how it is used in context. In fact, this basic strategy is an important approach in computational linguistics: When computers need to learn the meanings of words, they do so by analyzing the contexts that those words appear in.

C1. Based on what you can infer from the encyclopedia article, fill in the letter or letters of the vloxes that fit each description from your manager. You will need to use some letters more than once.

C2. Fill in the blanks to create a description of the youngest vlox in the image. Each blank should be filled in with one of the following words: *kibb*, *womm*, *zuff*, *tezz*, *ravv*. You should not use the same word twice.

___(a)___ ish-___(b)___

Make sure you record your answers in your Answer Sheets!



(C) Vloxes (2/2)

Here is the encyclopedia article about vloxes that you should use to help you understand the unfamiliar words in your manager's descriptions.

Vlox

From Marsipedia, the Martian encyclopedia

The vlox (*Vloxis vloxia*) is the most imxo species of zoobleporf. Due to its playful disposition, it is a popular pet throughout Mars.

Contents

- 1 Appearance
- 2 In popular culture

Appearance

Vloxes occur in two breeds: the polar vlox and the valley vlox. Polar vloxes are usually more mellivu than valley vloxes.^[1] Regardless of breed, all vloxes are born completely qliqli. As they mature, they become less and less qliqli. It is believed that all prehistoric vloxes were zuffish-kibb or tezzish-kibb, but modern vloxes display great variety in zlorvitude.

In popular culture

- In the "Olympus Mons Explorers" comics, Captain Dfdgb's spaceship is called *The Mellovu Vlox*.
- The first Martian prime minister, Jzirla Jzirlsdaughter, was often seen with her kibbish-zuff vlox named Klgrdm.
- In the sitcom "Phobos and Deimos," the main character Phobos has three imxo vloxes: a kibb vlox named Blorblim, a zuffish-ravv vlox named Qqqq, and a ravvish-womm vlox named Robert.
- The most popular song by the Bellona Quartet is "Lookin' Like a Nearly Azeltorf Vlox," which spent seven weeks at the #1 position on the Mercury Top 60 chart.

Notes

¹ In a previous edition of this encyclopedia, we said that polar vloxes are more mellovu than valley vloxes. This was a typo - obviously, it makes no sense to say "more mellovu"!



(D) For the Sake of Their House (1/1) [10 Points]

The Permyak language is spoken by around 60,000 people on the western slope of the Ural Mountains in Russia. It belongs to the Finno-Ugric language family, along with Finnish, Estonian, Hungarian, and other languages that are mostly spoken in Russia.

Examine the following words in Permyak and their translations in English:

	Permyak	English
1.	cerkulaŋ	towards the house
2.	pizannezitulən	of your (sg.) desks
3.	ponit	your (sg.) dog
4.	purtnis	their knife
5.	käinnezis	his wolves
6.	vərälən	of my forest
7.	purtəla	for the sake of my knife
8.	tiezitkət	with your (sg.) lakes
9.	cerkuezliç	from the houses
10.	juççezə	my swans
11.	kokiskət	with his foot
12.	ciitlaŋ	towards your (sg.) hand

Notes: ï and ə are special vowel sounds of Permyak; c, ç, and ŋ are pronounced similarly to k, s, and n, respectively. The abbreviations *sg.* and *pl.* stand for *singular* and *plural*. The word *their* has a plural meaning rather than a gender-neutral singular meaning.

D1. Translate the following Permyak words into English.

- pizaniçliç
- vərrezlən
- ponnit
- juçla

D2. Translate the following English phrases into Permyak.

- his hands
- my dogs
- of your (pl.) feet
- for the sake of their house

Make sure you record your answers in your Answer Sheets!

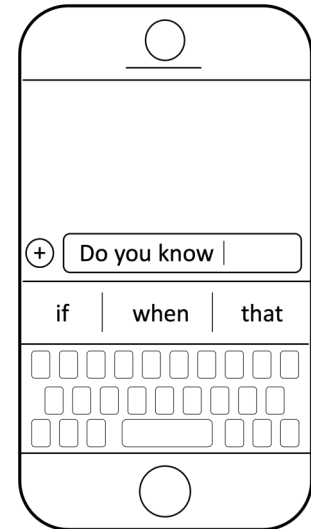


(E) Bengalese Finch Song (1/3) [5 Points]

One important feature of human language is unpredictability. If you knew what people were going to say beforehand, they wouldn't need to say it!

But language isn't completely random either. If you've ever used the auto-complete feature on a phone, you'll know that it can often predict your next word based on the words you've already typed. For instance, after the word *know*, your phone might suggest *if*, or *when*, or *that*, since those are the words that are most frequent after *know*.

This type of prediction is not just relevant for words: it can also be done with letters. For example, in English, after the letter *q*, we can predict with nearly 100% certainty that the next letter will be (you guessed it!) *u*. But for most letters in English, the next letter is less certain. After *t*, the most common next letter is *h*, which follows it around 40% of the time (mostly because of a few common words like *the*, *this*, *that*, and *then*) — but it is also frequently followed by the vowels *o*, *e*, and *i*.



For a computer to predict what letter will come next, the computer needs to store transitional probabilities, which define how likely each letter is to appear after each other letter. For example, we might say:

$$p(t \rightarrow h) = 0.40$$

This means “the probability that *t* will be followed by *h* is 0.40.” We can calculate transitional probabilities by using a piece of text. For instance, $p(t \rightarrow h)$ would be estimated as the number of times that *th* occurs in the text, divided by the number of times that *t* occurs.

Below is some text in Hawaiian, which uses 13 letters in its writing system. Spaces have been removed, and the symbol ` is a letter like any other, not punctuation — it's pronounced like the sound in the middle of the English exclamation *uh-oh*. At the top of the next page is a partially-filled table showing the steps to computing Hawaiian transitional probabilities (*L1* and *L2* stand for *letter 1* and *letter 2*).

ikinohihanakeakuaikalaniamekahonuahe`ano`ole
kahonuaua`oloheloheamalunanookahohonukapouli
ho`opunanaiholaka`uhaneokeakuamalunaokawai

E1. On your Answer Sheets, fill in the blanks from the table on the next page.

E2. Suppose a Hawaiian speaker is typing, and the last letter they have typed is *k*. Based on the text above, what is the next letter most likely to be? Answer on your Answer Sheets.



(E) Bengalese Finch Song (2/3)

Table described on the previous page:

L1	L2	Count of L1 followed by L2	Count of L1	P(L1 → L2)
i	h	3	8	3/8
k	u	2	12	1/6
k	e	2	12	1/6
k	i	a.	b.	c.
k	a	d.	e.	f.

Below are four sequences of letters. Two of the sequences are Hawaiian written in code (each letter has been replaced with a different letter). Don't worry about trying to figure out which letter in the code stands for which Hawaiian letter — that is not necessary for solving the problem.

The other two sequences are from the songs of two Bengalese finches, a species of songbird originally from Southeast Asia. In the case of the Bengalese finch song, the letters represent “syllables,” which are the basic units that biologists have identified in their songs.

Bengalese finch song is interesting because, unlike many other types of birdsong, it is not entirely predictable — but it's still more predictable than human language. In other words, given that you hear one syllable, you can be more certain — on average — about which syllable you'll hear next, than you could be when reading human language.

E3. Which two sequences are Bengalese finch song? Note that the code used might be different between two texts of the same type—for example, *a* might stand for a different syllable in the first sequence of finch song than in the second.

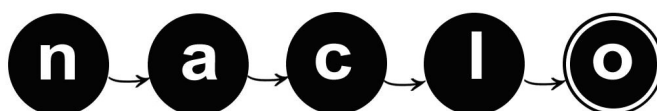
Sequence A: abcbaefbdgdabhiijgdbcbhgdabkieidgdahbjaficblbfefkbcf

Sequence B: abacdefghahbhicdefghahbhgicdefgbjklcdefgammlcdefgajkl

Sequence C: abcdefbfcdefbfcdeaghijkbcdefafcdebcdeaghiffcdefafcdef

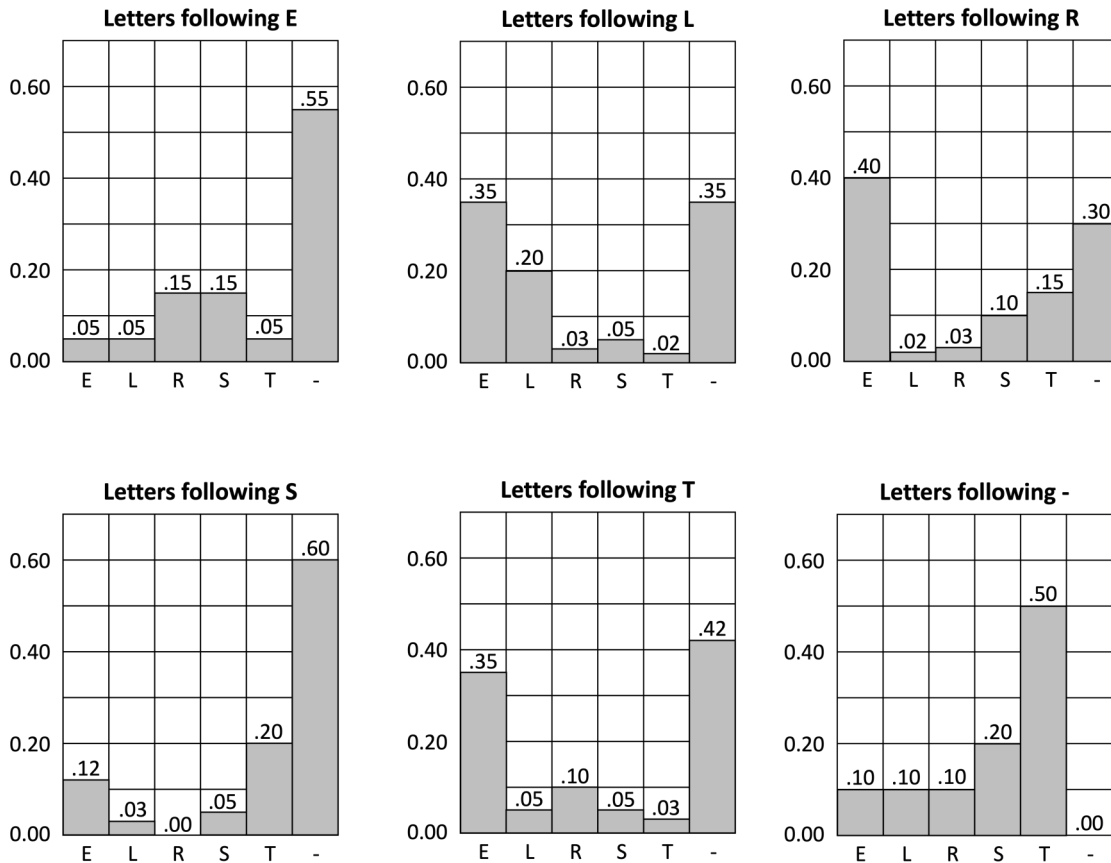
Sequence D: abcdedfgfdgfbahibjdkbghcbfcbffjdcbgbidgdldbfgdibjdgba

Make sure you record your answers in your Answer Sheets!



(E) Bengalese Finch Song (3/3)

Below are some plots showing transitional probabilities for a few English letters, estimated based on Wikipedia. For example, $p(R \rightarrow E) = 0.40$. The dash symbol - is used to indicate a space.



Given a piece of text, we can label each letter with its transitional probability based on the previous letter. For instance, TREES-TREES-TREES would be labeled as shown here (with no label on the first letter, since it has no previous letter to use for determining the transitional probability):

	.10	.40	.05	.15	.60	.50	.10	.40	.05	.15	.60	.50	.10	.40	.05	.15
T	R	E	E	S	-	T	R	E	E	S	-	T	R	E	E	S

E4. Below is a phrase labeled in this way — but all of the letters are missing, leaving just the transitional probabilities! What is the phrase? In your answer, capitalization does not matter, and you can use either a space “ ” or a dash “-” to indicate a space.

	.40	.15	.20	.05	.35	.15	.05	.60	.10	.35	.05	.03	.35	.15	.10
?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?

Make sure you record your answers in your Answer Sheets!



(F) A Calabash Calculation (1/1) [15 Points]

While touring the *sii jangidē* ("school"), you attempt to learn a few words of Nizaa, with the help of a few trusty flashcards. Unfortunately, someone failed to print the flashcards double-sided! Your job is to match the Nizaa (left) and English (right) back together.

- | | |
|--------------------------|------------------------------------|
| 1. <i>beè</i> | A. <i>Anthocleista vogelii</i> |
| 2. <i>beè nim</i> | B. <i>calabash</i> |
| 3. <i>bòw</i> | C. <i>carpenter</i> |
| 4. <i>bòw-ηwiij</i> | D. <i>cup (for drinking water)</i> |
| 5. <i>bɯɯ</i> | E. <i>dog</i> |
| 6. <i>cam fwèèη</i> | F. <i>duck</i> |
| 7. <i>cam-ηwiij</i> | G. <i>finger nail</i> |
| 8. <i>cún fwèèη</i> | H. <i>hairdresser</i> |
| 9. <i>cún nim</i> | I. <i>head</i> |
| 10. <i>ḏóɔη</i> | J. <i>hunter</i> |
| 11. <i>ḏóɔη nim</i> | K. <i>little finger</i> |
| 12. <i>hodē ḏóɔη</i> | L. <i>puppy</i> |
| 13. <i>nii cún</i> | M. <i>roof</i> |
| 14. <i>nii gírcí bɯɯ</i> | N. <i>sauce</i> |
| 15. <i>nii kwéεη</i> | O. <i>sauce spoon</i> |
| 16. <i>nim</i> | P. <i>soup</i> |
| 17. <i>nim bɯɯ</i> | Q. <i>source of water</i> |
| 18. <i>sii bɯɯ</i> | R. <i>teardrop</i> |
| 19. <i>sijw kwéεη</i> | S. <i>tree bark</i> |
| 20. <i>sijw nim</i> | T. <i>water</i> |
| 21. <i>yír nim</i> | U. <i>wild bird</i> |

Notes: Nizaa is spoken by about 10,000 people in Cameroon. *Anthocleista vogelii* is a species of tree that thrives in the rainforest. A *calabash* is a gourd (a type of hard-shelled fruit, similar to a pumpkin) that is often hollowed out and dried for use as a container. Pronunciation is not relevant for solving this problem, but *b*, *η*, and *ḏ* are consonants; *ɯ*, *ε*, and *ɔ* are vowels; and marks above vowels indicate tone—the pitch used when pronouncing a vowel.

F1. On your Answer Sheets, indicate which English translations go with which Nizaa words and phrases.

F2. Translate into Nizaa: a. *bush* b. *hut* c. *to read* d. *to shave*

Make sure you record your answers on your Answer Sheets!



(G) Feathers of the Roseate Spoonbill (1/2) [15 Points]

Classical Nahuatl was the language of the Aztec Empire. Although Classical Nahuatl is no longer actively spoken, it is related to modern Nahuatl, which has approximately 1.5 million speakers, mainly in Mexico.

Over the centuries, scholars have proposed several different writing systems for representing Classical Nahuatl. Some early approaches from the 16th century were based on the system used for Spanish at the time. However, this approach faced challenges in how to represent sounds and distinctions that were not present in Spanish.

One solution is to use the International Phonetic Alphabet (IPA). The IPA is designed to represent all sounds of human speech unambiguously, and so it provides a reliable transcription system that can be read by linguists across the world. However, the IPA is often unwieldy and does not reflect the needs of particular language communities. Therefore, the IPA is rarely used by non-linguists.

Some more recent proposals are therefore designed specifically with Nahuatl in mind, in order to ensure both ease of use (which the IPA lacks) and faithfulness to the language (which the 16th-century systems lacked). To illustrate some of the writing systems that have been proposed over the years, the name of the Nahuatl language can be written as *nahuatlātōlli* in a 16th-century writing system, /nawatʎaʔto:lli/ in the IPA, or *nahuatlàtōlli* in a 20th-century writing system.

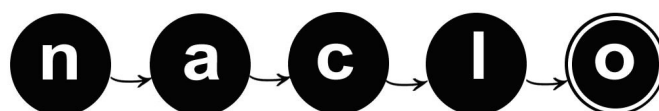
Below is the first stanza of a Nahuatl poem, “The Song of Temilotzin.” On the left is the poem as it appears in a manuscript of the 16th century (modified slightly for clarity). On the right, the text has been rewritten in a system proposed more recently, in the 20th century.

16th-century writing system

Ye nihualla, antocnihuan in:
noconcozcazoya,
nictzinitzcamana,
nictlahquecholihumolohua,
nicteocuitlaicuiya,
nicquetzalhuixtoilpiz
in icnihyotl.
Niccuicailacatzoa cohuayotl.
In teucpan nicquixtiz,
an ya tonmochin,
quin icuac tonmochin in otiyaque
ye Mictlan.
In iuh ca zan tictlanehuico.

20th-century writing system

Ye nihuāllā, antocnihuān in:
noconcōzcazōya,
nictzinitzcamana,
nictlāuhquechōlihumolohua,
nicteōcuitlaicuiya,
nicquetzalhuixtoilpīz
in icnihyōtl.
Niccuīcailacatzoa cōhuayōtl.
In tēucpan nicquixtīz,
ān ya tonmochīn,
quin ìcuāc tonmochīn in ōtiyàquè
ye Mictlān.
In iuh ca zan tictlānehuīcō.



(G) Feathers of the Roseate Spoonbill (2/2)

Here are some words and phrases from “The Song of Temilotzin” in the 16th-century writing system, along with their equivalents in the IPA, as well as English translations:

nictlahuquecholihuimolohua

/niktʰa:wketʃo:liwimolowa/

‘I adorn it with feathers of the roseate spoonbill (a kind of bird)’

nictecuitlaicuiya

/nikteo:k^witʰaiʔk^wija/

‘I wrap it with coils of gold’

nicquetzalhuixtoilpiz in icniuhyotl

/nikketsalwiftoilpi:s in ikni:wjo:tʰ/

‘I have bound up our friendship like a captive with feathers of the quetzal (another kind of bird)’

in teucpan nicquixtiz

/in te:k^wpan nikki:ʃti:s/

‘I will make it emerge in the palace’

otiyaque ye Mictlan

/o:tijaʔkeʔ je miktʰa:n/

‘we will have gone to Mictlan (the land of the dead)’

G1. Convert the following phrases from the 16th-century writing system into the IPA:

a. *ye nihualla* ‘I have already come’

b. *antocnihuan in* ‘my friends’

c. *quin icuac* ‘because’

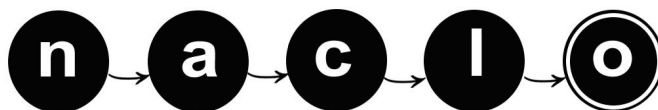
G2. Convert the following words from the IPA into the 20th-century writing system:

a. /we:weʔ/ ‘old man’

b. /tʃokola:tʰ/ ‘chocolate’

c. /miktʰa:nte:k^wtʰi/ ‘Lord of the dead’

G3. In set theory (a branch of mathematics), it is said that there is a *bijection* between two sets if it is possible to convert from one set to the other and back without losing any information. This concept can be extended to describe writing systems. For example, there is a bijection between the set of numbers written in numerals (e.g., 43 or 11) and the set of numbers written in letters (e.g., *forty-three* or *eleven*). If you ignore capitalization, there is a bijection between two of the writing systems discussed here. Which two writing systems are they?



(H) My Grandmother and I (1/2) [15 points]

Pronouns are words like *us* or *she* that refer to a person or group relevant to the conversation. This problem is about the system of pronouns in the language Lardil, which is spoken on Mornington Island, off the north coast of Australia. Lardil's pronoun system may seem complex at first, but it is organized around a systematic and logical set of principles.

Below are thirteen groups of people, followed by pronouns that could refer to them in Lardil. Note that the Lardil words are not translations of the English—they are pronouns that could refer to the group described in English, similar to how the English pronoun *us* is not defined as “Alex and me” but might refer to “Alex and me” in a particular context. The English and Lardil are in the same order as each other, not scrambled.

English	Lardil
my grandmother and I	nyarri
he, his brother, and his sister	bili
you, your sister, and your grandson	kili
my great-grandfather and I	nyaanki
you and your brother	kirri
my grandparents and I	nyali
you and your aunt	nyiinki
you, your father, and your mother	kilmu
my son, my mother, and I	nyalmu
she and her grandparents	bili
my son and I	nyaanki
my sister, my grandson, and I	nyali
she and her great-granddaughter	niinki

H1. On your Answer Sheets, match up the following groups and pronouns:

- | | |
|--|-----------|
| a. you and your parents | 1. nyalmu |
| b. my grandfather and I | 2. bili |
| c. you and your grandchildren | 3. kilmu |
| d. she, her grandmother, and her brother | 4. kili |
| e. he and his sister | 5. nyarri |
| f. my daughter, my mother, and I | 6. birri |

Make sure you record your answers in your Answer Sheets!



(H) My Grandmother and I (2/2)

H2. What would be the correct Lardil pronoun for:

- a. my father and I
- b. you and your granddaughter
- c. my uncle, my aunt, and I
- d. he and his parents
- e. she and her mother
- f. you and your cousin

H3. On your Answer Sheets, circle all of the following groups that would be designated as *birri*. The number of answers that should be circled could be 0, 1, 2, or 3.

my uncle
and I

you, your brother,
and your grandfather

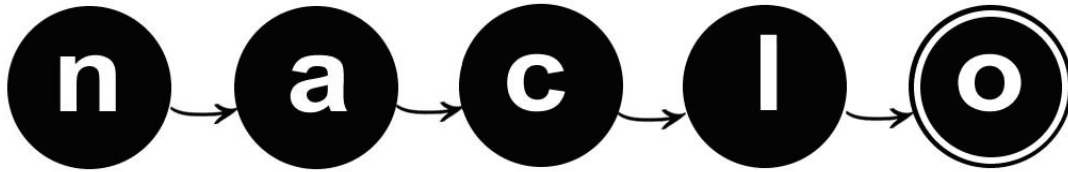
she and
her sister

H4. There is another pair of pronouns in Lardil, *ngakuli* and *ngakulmu*, which are used to refer to groups made of “you, I, and one other person.” Assuming that “I” and “you” are brother and sister, which word would you use to refer to:

- a. you, I, and our grandfather
- b. you, I, and our aunt

H5. What Lardil word would refer to “you and I”, where “I” refers to the speaker and “you” refers to the speaker’s sister?





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: _____

Answer Sheets (1/6)

(A) The Way We Were

A1. Fill in the empty cells in the table.

	Old English	Present Day English
16.	a.	archbishop
17.	b.	King Edward
18.	menn	c.
19.	weras	d.
20.	Ne hierde ge hi.	e.
21.	Ne gesawon hi þe.	f.
22.	Hi hit fremedon.	g.
23.	Ne neredede we þe.	h.

A2. Complete the partial Old English sentences by writing in the blanks.

	Old English	Present Day English
24.	a. Ne grette ge _____ .	You did not greet them.
25.	b. Seo lar hi _____ .	The teaching saves them.
26.	c. Ðu _____ .	You heard it.
27.	d. Ða _____ .	Then they heard us.



Answer Sheets (2/6)

(A) The Way We Were (continued)

A3. Write the letter of the context that corresponds to each Old English sentence.

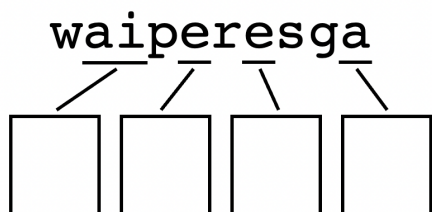
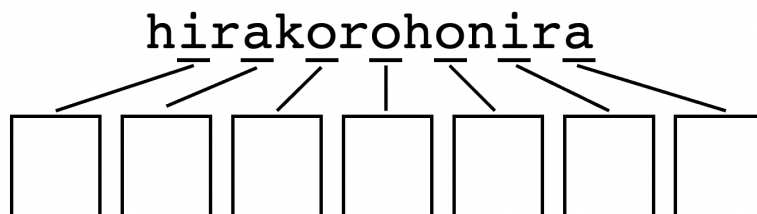
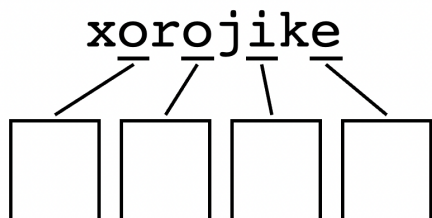
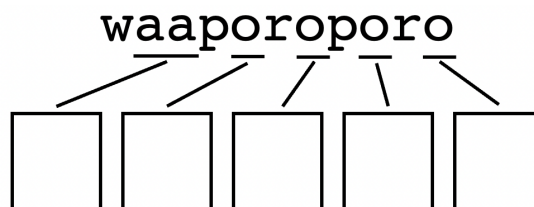
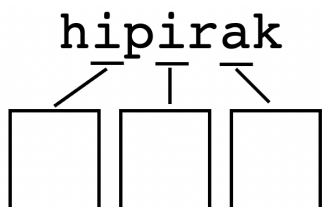
- i. ii. iii. iv. v. vi.

A4. Write the missing Old English word that goes in each blank.

- a. b. c.
 d. e. f.

(B) Don't Stress

B1. For each Ho-Chunk word, indicate the stress by writing one letter (N, P, or S) in each blank.



YOUR NAME:

REGISTRATION #

Answer Sheets (3/6)

(C) Vloxes

C1. Write the letter or letters of the vloxes that fit each description (one letter per vlox). If an answer contains multiple letters, the order of the letters does not matter.

A. the two mellivu vloxes	<input type="text"/>	<input type="text"/>	f. the imxo vlox	<input type="text"/>		
B. the two mellovu vloxes	<input type="text"/>	<input type="text"/>	g. the three ravv vloxes	<input type="text"/>	<input type="text"/>	<input type="text"/>
C. the two kibb vloxes	<input type="text"/>	<input type="text"/>	h. the kibbish-zuff vlox	<input type="text"/>		
D. the two azeltorf vloxes	<input type="text"/>	<input type="text"/>	i. the zuffish-kibb vlox	<input type="text"/>		
E. the qliqli vlox	<input type="text"/>					

C2. Fill in the blanks. ish-

(D) For the Sake of Their House

D1. Write the English translation for each Permyak word.

a. pizanişliç

b. vərrezlən

c. ponniŧ

d. juçla

D2. Write the Permyak translation for each English phrase.

a. his hands

b. my dogs

c. of your (pl.) feet

d. for the sake of their house



YOUR NAME:

REGISTRATION #

Answer Sheets (4/6)

(E) Bengalese Finch Song

E1. Fill in the missing cells in the table.

L1	L2	Count of L1 followed by L2	Count of L1	P(L1 → L2)
i	h	3	8	3/8
k	u	2	12	1/6
k	e	2	12	1/6
k	i	a.	b.	c.
k	a	d.	e.	f.

E2. What is the next letter most likely to be? Write your answer in the box to the right.

E3. Circle the two sequences that are Bengalese finch song.

Sequence A

Sequence B

Sequence C

Sequence D

E4. Fill in the blanks in the bottom row to complete the missing phrase. (The cell in the top left should remain blank.)

	.40	.15	.20	.05	.35	.15	.05	.60	.10	.35	.05	.03	.35	.15	.10



YOUR NAME:

REGISTRATION #

Answer Sheets (5/6)

(F) A Calabash Calculation

F1. Write the letter of the English translation that corresponds to each Nizaa word or phrase.

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

8. 9. 10. 11. 12. 13. 14.

15. 16. 17. 18. 19. 20. 21.

F2. Write the Nizaa translation for each English word or phrase.

a. bush

b. hut

c. to read

d. to shave

(G) Feathers of the Roseate Spoonbill

G1. Convert each phrase from the 16th-century writing system to the IPA.

a. ye nihualla

b. antocnihuan in

c. quin icuac

G2. Convert each word from the IPA to the 20th-century writing system.

a. /we:we?/

b. /tʃokola:tʃ/

c. /miktʰa:nte:kʷtʰi/

G3. Circle the pair of writing systems between which there is a bijection.

16th-century and 20th-century

16th-century and IPA

20th-century and IPA



YOUR NAME:

REGISTRATION #

Answer Sheets (6/6)

(H) My Grandmother and I

H1. Write the number of the Lardil pronoun that corresponds to each group.

- a. b. c. d. e. f.

H2. Write the Lardil pronoun for each group.

a. my father and I

b. you and your granddaughter

c. my uncle, my aunt, and I

d. he and his parents

e. she and her mother

f. you and your cousin

H3. Circle each group that would be designated as *birri*.

my uncle
and I

you, your brother,
and your grandfather

she and
her sister

H4. Write the Lardil pronoun for each group.

a. you, I, and our grandfather

b. you, I, and our aunt

H5. Write the Lardil word that would refer to “you and I”, where “I” refers to the speaker and “you” refers to the speaker’s sister.

