NACLO thanks the following for their generous contributions:

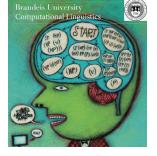




Carnegie Mellon University Language Technologies Institute













North American Computational Linguistics Open Competition

2022

www.nacloweb.org

Open Round January 27, 2022

Serious language puzzles that are surprisingly fun!

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

Welcome to the sixteenth 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 nine problems, labeled A to I. Note that this year's contest has one more 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 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 2022 Organizers

Organizing Committee:

Adam Hesterberg — Massachusetts Institute of Technology Aleka Blackwell — Middle Tennessee State University Ali Sharman — University of Michigan Andrew Tockman — Massachusetts Institute of Technology Annie Zhu — Harvard University Ben LaFond — Harvard University Brian Xiao — Massachusetts Institute of Technology Cerulean Ozarow — Brown University Daniel Lovsted — McGill University David Mortensen — Carnegie Mellon University Dragomir Radev — Yale University Duligur Ibeling — Stanford University Ellina Zhang- University of Toronto Ethan Chi — Stanford University Heidi Lei — Massachusetts Institute of Technology James Hyett — University of Toronto Jakin Ng— Massachusetts Institute of Technology James Pustejovsky — Brandeis University Ji Hun Wang — Stanford University Ken Jiang — University of Waterloo Kevin Liang — University of Pennsylvania Lori Levin — Carnegie Mellon University Margarita Misirpashayeva — Massachusetts Institute of Technology Matt Gardner — Allen Institute for AI Mihir Singhal — Massachusetts Institute of Technology Nathan Kim — Stanford University Nathaniel Satriya — University of California, San Diego Patrick Littell — National Research Council Canada Pranav Krishna — Massachusetts Institute of Technology Shuli Jones — Massachusetts Institute of Technology Skyelar Raiti — University of Michigan Stella Lau — Massachusetts Institute of Technology Rui Zhang — Pennsylvania State University Ryan Chi — Stanford University Ryan Guan — Stanford University



NACLO 2022 Organizers (cont'd)

Organizing Committee (cont'd)

Tom McCoy — Johns Hopkins University Tom Roberts — University of California, Santa Cruz Yilu Zhu — Fordham University

Organizing Committee Co-Chairs:

Kevin Liang — University of Pennsylvania Shuli Jones — Massachusetts Institute of Technology

Program Committee:

Adam Hesterberg — Massachusetts Institute of Technology Aleka Blackwell — Middle Tennessee State University Ali Sharman — University of Michigan Cerulean Ozarow — Brown University Daniel Lovsted — McGill University Dragomir Radev — Yale University Ethan Chi — Stanford University Evan Hochstein — Yale University Jonathan Huang — Massachusetts Institute of Technology Lori Levin — Carnegie Mellon University Patrick Littell — University of British Columbia Pranav Krishna — Massachusetts Institute of Technology Ryan Chi — Stanford University Skyelar Raiti — University of Michigan Tom McCoy — Johns Hopkins University

Program Committee Co-Chairs:

Daniel Lovsted — McGill University Tom McCoy — Johns Hopkins University

Reviewers and Problem Testers:

Ben LaFond, Jalen Chrysos, Ken Jiang, Matt Gardner, Nathaniel Satriya, William Pan, and others



NACLO 2022 Organizers (cont'd)

USA University Site Coordinators:

Boston Area NACLO Site — Shuli Jones California State University, Dominguez Hills — Iara Mantenuto Carnegie Mellon University — John Friday, Lori Levin College of William and Mary — Dan Parker Columbia University — Brianne Cortese, Daniel Bauer, Kathy McKeown, Smaranda Muresan Fort Hays State University — Destiny Gu, Jodi Hill, Sherri Matlock Georgia Tech — Hongchen Wu Middle Tennessee State University — Aleka Blackwell Minnesota State University, Mankato — Dean Kelley, Louise Chan, Rebecca Bates Montclair State University — Anna Feldman, Jonathan Howell, Lauren Covey Northeastern Illinois University — Ariana Bancu, Lewis Gebhardt Ohio State University — Marie de Marneffe, Micha Elsner, Michael White Planet Word Museum (Washington, D.C. Area NACLO Site) — Emily Gref, Rebecca Roberts Princeton University — Christiane Fellbaum, Misha Khodak, Oliver Weizel Rollins College — Margarita Azbel San Diego State University — Rob Malouf San Francisco Bay Area NACLO Site — Ethan Chi, Ryan Chi Stony Brook University — Jeffrey Heinz, Lori Repetti, Sarena Romano Union College — Kristina Striegnitz, Nick Webb University at Buffalo — Jeff Good, Leslie Ying, Cassandra Jacobs University of California, Irvine — Kristen Salsbury, Sameer Singh, Zhengli Zhao University of Maryland — Jan Michalowski, Polina Pleshak, Sigwan Thivierge University of North Carolina at Charlotte — Hossein Hematialam, Kodzo Wegba, Seethalakshmi Gopalakrishnan, Wlodek Zadrozny University of Notre Dame — David Chiang University of Pennsylvania — Anne Cocos, Cheryl Hickey, Chris Callison-Burch, Derry Wijaya, Oliver Sayeed, Mitch Marcus University of Southern California, ISI campus — Jon May University of Southern Maine — Claire Holman, Dana McDaniel University of Utah — Aniello De Santo, Justin Nistler, Karen Marsh Schaeffer University of Washington — Jim Hoard, Joyce Parvi University of Wisconsin, Milwaukee — Anne Pycha, Gabriella Pinter, Joyce Boyland Wichita State University — Jill Fisher, Mythili Menon Yale University — Raffaella Zanuttini



NACLO 2022 Organizers (cont'd)

Canada University Site Coordinators:

McGill University — Lisa Travis, Michael Wagner Opus Academy — Janette Lim, Lydia Cheng University of British Columbia — Jozina Vander Klok, Yadong Liu University of Ottawa — Andrés Pablo Salanova University of Toronto — Ellina Zhang

Special thanks to:

The hosts of the 130+ High School Sites

Booklet Editor: Daniel Lovsted — McGill University

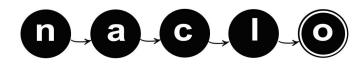
US Team Leaders:

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

> Canadian Team Leader: Daniel Lovsted — McGill University

NACLO Co-Chairs:

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



NACLO 2022 Credits

Problem Credits:

(A) Michael Salter

(B) Tom McCoy Consultant: Asli Celikyilmaz

(C) Ryan Chi

(D) Daniel Lovsted *Consultants:* Edna MacLean, Kirk Miller, Myles Creed

(E) Aleka Blackwell

(F) Catherine Scanlon Story Author: Jubli Sowati, collected by Kate Lindsey Consultants: Jack Dipa, Mathias Dugal, Tonny Warama, Warama Kurupel

(G) Tom McCoy

(H) Simi Hellsten Consultant: Gilles Polian

(I) Babette Verhoeven

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

Image Credits:

(A), map: Adapted from maps by Wikipedia users Cruickshanks and Tintazul (CC-BY SA: https://creativecommons.org/licenses/by-sa/3.0/deed.en)
(E), map: By Monika Feinen, cropped from the original (CC-BY: https://creativecommons.org/licenses/by/2.0/)
(F), ant: by Pixabay user DreamDigitalArtist (https://pixabay.com/vectors/ant-insect-animal-silhouette-6585114/)
(F), fish: by Pixabay user GDJ (https://pixabay.com/vectors/fish-animal-line-art-scales-5965349/)

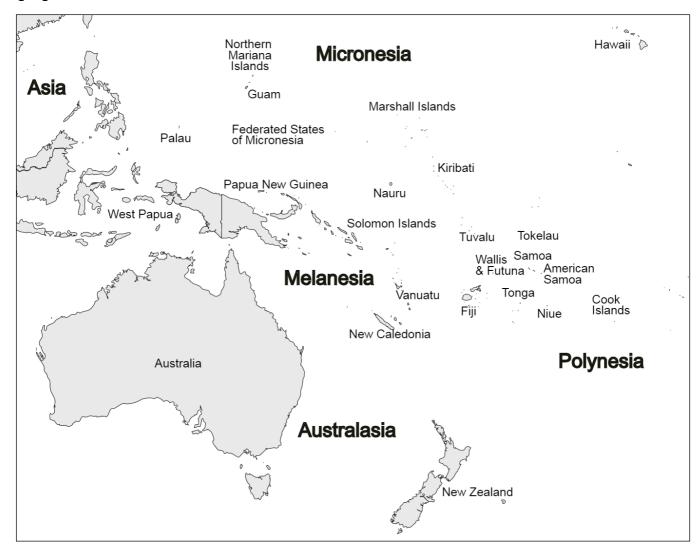
We are grateful for the support of many institutional and individual donors who make this contest possible.

All material in this booklet © 2022, North American Computational Linguistics Open Competition and the authors of the individual problems. Please do not copy or distribute without permission.



(A) Lines in the Sand (1/2) [10 Points]

The Avoiuli writing system is used on Pentecost Island in Vanuatu, a Pacific island country (near the center of the map below). With a design inspired by traditional sand drawings, Avoiuli was developed over fourteen years by Chief Viraleo Boborenvanua, as part of a movement he leads to revitalize Vanuatuan culture and language.



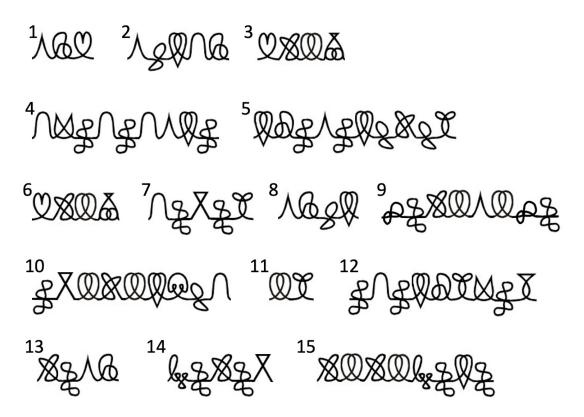
One of the languages that Avoiuli is used for is Bislama, one of the official languages of Vanuatu. As you will see, Bislama is a creole language¹ – many of its words come from English, some come from French, and others from the indigenous languages of Vanuatu.

^{1.} When multiple groups of people who speak different languages need to communicate, they often develop a simplified system called a *pidgin*, which mixes words from their various languages but lacks the grammatical complexity of a language. When a pidgin is taught to a new generation of children, they typically develop it into a fully complex language called a *creole*.



(A) Lines in the Sand (2/2)

Below is a sign written in Bislama, using the Avoiuli script, on the front of a college on Pentecost Island. For the purposes of this problem, each of the words in the script below has been allotted a number:



A1. Give the number of a word (any one, if there are multiple) which corresponds to these Bislama words:

a. filosofi b. institiut c. blong

A2. Give the number of a word (any one, if there are multiple) which means:

a. village b. center c. teach(ing) d. technology e. humanity

A3. Which geographical area of the Pacific is mentioned in the text on the sign?



(B) Who Saw the Bear? (1/1) [10 Points]

Turkish is the national language of Turkey, spoken by approximately 80 million people in Turkey and elsewhere.

B1. Below are some sentences in Turkish, along with their English translations. On your Answer Sheet, fill in the entries that are missing from the table.

Turkish	English	
Balık yüzdü.	The fish swam.	
Zürafa atladı.	The giraffe jumped.	
Fil koştu.	The elephant ran.	
Kaplan kaplumbağa için atladı.	The tiger jumped for the turtle. ¹	
Balık kaplan gibi yüzdü.	The fish swam like the tiger.	
Fil ayı gibi yüzdü.	The elephant swam like the bear.	
Fil atladı.	a.	
b.	The turtle ran for the fish.	

B2. Here are some more Turkish sentences with English translations. On your Answer Sheet, fill in the entries that are missing from the table.

Kurbağa zürafayı gördü.	The frog saw the giraffe.	
Zürafa martıyı gördü.	The giraffe saw the seagull.	
Arı kurbağayı gördü.	The bee saw the frog.	
Kim martıyı gördü?	Who saw the seagull?	
Kim arıyı gördü?	Who saw the bee?	
Arı kimi gördü?	Whom did the bee see?	
Balina kimi gördü?	Whom did the whale see?	
Martı kaplumbağayı gördü.	a.	
b.	Who saw the bear?	
с.	Whom did the bear see?	

1. Whenever "for" appears in a translation in this problem, it means "for the benefit of," not "toward." (This information is not relevant for solving the problem.)



(C) To Make a Long Story Short (1/5) [15 Points]

Every day, the Internet brings a flood of new information: scientific articles, blog posts, Tweets, news stories, Wikipedia pages, and more. One way we can take advantage of this information without being overwhelmed is through **automatic summarization**, where a computer condenses a long document into a summary that is much shorter. With this summary, you can get the main points of the document without reading the whole thing. You can also read the summaries of many documents to help decide which of the documents are worth reading in their entirety.

Below is a news article that we wish to summarize (written out as a list of numbered sentences):

The Final Word on Language

The NACLO News

Are Wickelphones making a comeback? Story on G2.

A STICKY SITUATION

SS1: On January 22, Ronald McBubble, president of the Bigger Bubble corporation, was found chewing a stick of gum from rival company Made to Stick.

SS2: The incident set off an uproar among Bigger Bubble faithfuls, with over 10,000 gum enthusiasts in just one week signing a petition for McBubble to step down.

SS3: According to USA Today, the nine-year-long "great gum war" may soon be over, as Bigger Bubble's stock price has fallen 50 percent since that fateful day.

SS4: McBubble maintains that the gum was "planted" on him and that the whole incident was "100 percent a set-up."

SS5: Why did he continue chewing after reading the label?

SS6: "Because the gum was too sticky to remove."

SS7: One must hope that McBubble has since managed to render the gum less adhesive, although the fiasco does make for some . . . *sticky* headlines.

How would a computer summarize this story? One approach is to have the computer select a few sentences from the document to serve as the summary. In order to choose which sentences to use, the computer first assigns values to **features** of every sentence. In this context, features are things that you can observe or count that help determine whether the sentence will be useful in a summary.

C1. At the top of the next page, on the left, are definitions of the features that we will use. Further down is a table showing the feature values that would be assigned to each sentence. Unfortunately, some words are missing from the feature definitions! In addition, in the table, the sentences are in scrambled order, and some of the sentence numbers are missing.

i. Fill in the blanks a. through g. by choosing words or numbers from the word bank at the top of the next page, on the right. You will not use all of the entries in the word bank; you might use some entries more than once.

ii. Determine which sentences correspond to h. through k. You should answer using the sentence numbers that start with SS.



(C) To Make a Long Story Short (2/5)

<u>F1</u> : Assign 1 point if the sentence contains at least		W	ORD BAN	<u>IK</u>	
one <u>a.</u> ; otherwise, assign 0 points.	italicized	d word	capital	ized word	ł
<u>F2</u> : Assign 1 point for every proper noun in the	underlin	ned word	hyphei	nated wo	rd
sentence.	commor	n noun	past-te	ense verb	
<u>F3</u> : Take the number of words in the sentence	adjectiv	е	verb		
minus <u>b.</u> . If this value is negative, give the	number		plural	noun	
sentence that many points. Otherwise, give the	abbrevia	ation			
sentence 0 points.	word the	at appeai	s in the t	itle	
<u>F4</u> : Assign 1 point for every <u>c.</u> in the	noun re	ferring to	candy or	gum	
sentence.	colon	comma	•	period	ł
<u>F5</u> : Assign 1 point for every <u>d.</u> or <u>e.</u>	first	second		third	
that is repeated in at least one other sentence in	last	second	-to-last	third-1	to-last
the story.	1	2	3	4	5
<u>F6</u> : Assign 1 point if this sentence is the <u>f.</u>	6	7	8	9	10
or <u>g.</u> one in the story; otherwise, assign 0	11	12	13	14	15
points.	16	17	18	19	20

Notes: A proper noun is a word or phrase, usually capitalized, that describes a specific entity such as a person, place, or organization. Examples of proper nouns are Brazil and Alice. A common noun is a noun that is not a proper noun. For some terms in the feature definitions, there might be multiple ways to count (e.g., multiple ways to count proper nouns for F2). It is up to you to figure out exactly how the computer is counting each category.

Sentence		Features					
	F1	F2	F3	F4	F5	F6	Scores
h.	0	1	0	1	2	0	4
SS1	0	4	0	1	1	1	7
i.	0	0	-1	0	0	0	-1
SS3	1	2	0	2	1	0	6
j.	1	1	0	0	2	1	5
k.	0	0	-2	0	2	0	0
SS2	0	2	0	2	2	0	6



(C) To Make a Long Story Short (3/5)

The overall score for a sentence is determined by adding together all of its feature values. The sentences with the highest scores are the ones chosen to serve as the summary (in this problem, we will be using the 3 highest-scoring sentences). For this particular article, since sentences SS1, SS2, and SS3 have the highest summed scores (7, 6, and 6), they would comprise the summary. In the summary, the sentences are presented in the same order as they appear in the original story - in this case, SS1 followed by SS2 followed by SS3.

C2. Below are two more news articles and their accompanying feature scores. (The features, F1 through F6, are the same as those used for the previous story.) Unfortunately, due to some serious data corruption, the rows in the tables have been scrambled, and various feature scores have been erased! Your job is to fill in all the blanks.

The Final Word on Language

The RACLO News

REPORT: The elephant is sleeping. More on E1.

IN HOT WATER

HW1: Specialty soup restaurant Hot Water Village came under scrutiny last week, after an exposé was published in *The Wall Street Journal*.

HW2: No fewer than three sources (including *WikiLeaks*) accused the restaurant of re-serving unfinished soup, salvaged from the bowls of previous customers.

HW3: One witness further claimed, "the bowls are never washed."

HW4: "They're just refilled where the last customers left off, one bowl after another."

HW5: Hot Water Village firmly denies any such wrongdoing, with one spokesperson noting that "these blatantly false accusations degrade our bowls, restaurant, and very soup" and "cause the hot water to flow from my eyes."

HW6: If the accusations are to be believed, Hot Water Village, it seems, has succeeded in implementing not only a farm-to-table but also a *table-to-kitchen* approach.

Sentence		Features					Sum of
	F1	F2	F3	F4	F5	F6	Scores
a.	0	1	b.	1	4	0	6
HW4	с.	0	0	1	2	d.	e.
f.	0	0	g.	h.	1	0	1
HW6	1	i.	0	0	1	1	j.
k.	1	2	0	0	3	I.	m.
n.	1	1	0	0.	4	0	р.



(C) To Make a Long Story Short (4/5)

The Final Word on Language

The NACLO Rews

Feature: The latest in self-reference. Turn to C4.

THE STICKY SAGA CONTINUES

SSC1: After suffering a PR disaster and a sorbitol shortage, Ronald McBubble faces tough times ahead. **SSC2:** "We really gummed up the works," he notes miserably.

SSC3: Meanwhile, his rival, Chicle "Colonel" Sanders, is experiencing problems of his own, including a disappointing quarterly report.

SSC4: Perhaps most seriously, he is being investigated by the government's Bubble Safety Bureau on two counts of alleged wrapper forgery.

SSC5: One thing is for certain: it's a difficult time to be in the bubblegum business.

Sentence		Features						
	F1	F2	F3	F4	F5	F6	Scores	
q.	0	0	0	1	0	1	2	
r.	0	1	0	0	0	0	1	
s.	0	0	-1	0	0	0	-1	
t.	0	1	0	0	0	1	2	
u.	0	1	0	1	0	0	2	

C3. Which three sentences make up the summary for "In Hot Water"? You should answer using the sentence numbers that start with HW.

C4. Which three sentences make up the summary for "The Sticky Saga Continues"? You should answer using the sentence numbers that start with SSC.

The approach illustrated above (selecting a few sentences from the document to serve as the summary) is called **extractive summarization**. An alternative approach is called **abstractive summarization**. Instead of selecting sentences from the story, an abstractive summarization system generates its own summary, without being restricted to the sentences in the original story. This approach is closer to how humans summarize, but it is also more challenging because abstractive summarization systems can often introduce factual errors. Below is a one-sentence abstractive summary for the story "A Sticky Situation."

Abstractive summary: Made to Stick corporation is facing serious problems after its vice president was found chewing a stick of gum from rival company Bigger Bubble on February 22.



(C) To Make a Long Story Short (5/5)

C5. This summary contains several types of errors that are common in standard abstractive summary systems. Specifically:

- a. The abstractive summary includes one word that should be deleted. Which word?
- b. In the abstractive summary, there are two phrases that should be swapped (each of these phrases is two or three words long). What are these phrases?
- c. In the abstractive summary, there is a single word that should be replaced with a different word. What word should be replaced, and what word should be its replacement?

C6. As it turns out, even extractive summary systems can sometimes create factual errors! Consider the extractive summaries that were generated for "A Sticky Situation," "In Hot Water" and "The Sticky Saga Continues." (For "A Sticky Situation," we told you what the summary was: SS1, SS2, and SS3. For the other stories, the summaries are the ones that you specified in your answers to C3 and C4.) In one of those three summaries, there is a single word that has an incorrect meaning in the summary.

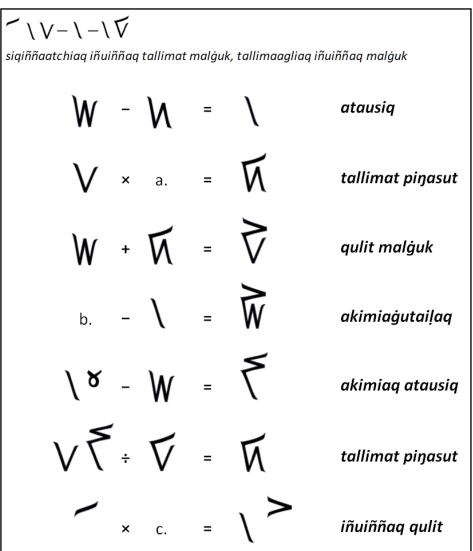
- a. Which sentence does this word appear in (you should answer using the sentence numbers that start with SS, HW or SSC)?
- b. Which word is it?



(D) Real Numbers (1/2) [10 Points]

In 1994, the students of a middle-school math class in Kaktovik, Alaska noticed that Arabic numerals (that is, the digits 0 to 9) were badly suited to counting in their language, Iñupiaq.¹ They decided to come up with a new numeral system, one that would reflect the Iñupiaq number system.

The Kaktovik Iñupiaq numerals (as they have come to be known) made it much easier for students to do math in Iñupiaq, and are easy to use and remember, since each symbol has an intuitive relationship with the number it represents. Within a few years, schools and colleges across northern Alaska were using the numerals, and they were recommended for use in Canada. If you walk into a math classroom in northern Alaska today, you might well see a blackboard that looks like the one below, which features Kaktovik Iñupiaq numerals, Iñupiaq words, and a few places intentionally left blank:



1. Iñupiaq, meaning "real or genuine person" (iñuk = person, -piaq = real, genuine), is the language of the Iñupiaq people (or Iñupiat). There are about 3000 speakers of the language, most of them living in northern Alaska. Iñupiaq is closely related to several other Inuit languages spoken in Canada and Greenland.



(D) Real Numbers (2/2)

D1. Fill in the blanks a., b., and c. from the blackboard with Kaktovik Iñupiaq numerals. Remember to record your answers on your Answer Sheet.

D2. Give the Iñupiaq word or words for:

a. three b. eleven c. twenty-two

D3. Write in Arabic numerals (i.e., using digits 0-9):

a. atausiq b. tallimat c. iñuiññaġutaiļaq

D4. In the top left corner of the blackboard, the same thing is written twice: once in Kaktovik Iñupiaq numerals, and once slightly differently in Iñupiaq words. Write in Arabic numerals (i.e., using digits 0-9):

a. the Kaktovik Iñupiaq numerals in the top left corner of the blackboard

b. quliagliaq



(E) Sleeping in the Shade (1/2) [10 Points]

The data in this problem are drawn from Ik, the native language of the Ik people who live on a narrow swath of land in the northeastern corner of Uganda, East Africa. (A map is provided on the next page.) The people call their language Icétôd, which means 'Ik-speech' or 'Ik-talk' and is pronounced ee-CHAY-TOad or [ītʃétôd] in phonetic symbols. Approximately 7,500 people are native speakers of Ik.

The Ik phrases and sentences in this problem are written in the commonly used orthography. Keep in mind the following as you examine the data:

- The digraph *ts* and the trigraph *ts'* are used to represent two different consonant sounds in Ik.
- A diacritic borrowed from the International Phonetic Alphabet has been added to the data to reflect a pronunciation feature relevant to the problem. The symbol is a small ring that appears below a letter. This diacritic reflects that the sound represented orthographically is pronounced without the vocal folds vibrating or, in less technical terms, is whispered by the native speakers of the language. Something similar happens in English when native speakers pronounce words like <u>potato</u> [p^hə't^heɪroʊ] and <u>peculiar</u> [p^hə'kju:liə] where the vowel in the first syllable becomes voiceless. To appreciate how this phonetic process works, try saying the word <u>peculiar</u> or <u>potato</u> with a pause between the first and second syllable of the word, and notice that your vocal folds do not begin vibrating until you pronounce the second syllable of the word. In the case of English, the vowel in the first syllable of words with this phonological structure is *devoiced*, which means it is pronounced without vocal fold vibration.

In the left column below appear sentences and phrases in Ik. In the right column, their English translations appear in a scrambled order.

1.	Kaa bee abaŋą.
2.	Atsę.
3.	Epa ŋoka na bets 'ą.
4.	Minia pecayą.
5.	Ats'a ŋoka ɔkaką.
6.	Kae zekwę.
7.	Atsia hoǫ.
8.	Epa ŋoka ɲcie kuruǫ.
9.	Atsima bee Isopiao.
10.	Mina cekia ntsię.
11.	hoika dii pcie lebetse ni gaaną
12.	Kaa oŋorǫ.
13.	Epa ŋoka kuruo na daą.
14.	hoiką
15.	Epa bee ŋoka hoo.
16.	Atsa abaŋa Isopia <i>q</i> .
17.	Epa oŋorą.
18.	Mina nakawaką.
19.	hoika dii

Α.	My dog is sleeping in the shade.
В.	these two huts of mine that are bad
C.	We came from Ethiopia.
D.	The dog is sleeping in the nice shade.
E.	The dog slept in the hut.
F.	Come.
G.	He loves coffee.
Н.	Father is coming from Ethiopia.
١.	these huts
J.	The dog that is white is sleeping.
К.	I love tea.
L.	Go sit.
M.	The elephant is sleeping.
N.	He loves his wife.
0.	the huts
Ρ.	Father went.
Q.	I am coming from the hut.
R.	The dog is chewing the bone.
S.	He is going with the elephant.



(E) Sleeping in the Shade (2/2)

E1. Match the English sentences and phrases to their Ik translations.

E2. You matched the two Ik sentences below to their well-formed English translations above in E1. Your new task is to translate them word-forword into English in a way that reveals the meaning of each Ik word, as we have started doing for S1. Remember to record your answers in your Answer Sheets!

S1. Epa ŋoka kuruo na daą. Sleeps _____ in-shade _

S2. Epa ŋoka na bets 'a.

E3. Translate into Ik. If there are two possible translations into Ik, please provide them both.

- a. these
- b. my huts
- c. I love my wife.
- d. Father is coming from the nice hut.

E4. Translate into English. If you can think of two structurally different translations into English, please provide them both.

- a. Zekwata oŋorika kuruo.
- b. Mina ŋoka ɔkaka ntsię.
- c. Minima oŋorika ni epą.
- d. noka na ncię

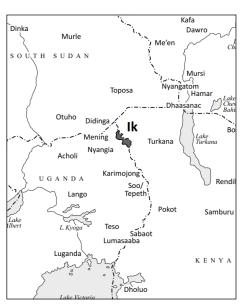
E5. Select the best option regarding the Ik phrases P1 and P2 below.

- P1. hoika dii ncie lebetsęP2. hoika dii ncie ni lebetsę
- (a) Only P1 is grammatical.
- (b) Only P2 is grammatical.
- (c) Both P1 and P2 are grammatical.
- (d) Neither P1 nor P2 is grammatical.

E6. State a rule that describes when speakers of Ik devoice vowels in their utterances.

E7. Some argue that Ik does not have adjectives as a grammatical category of words in the same way that English does. Give the numbers (1-19) of two sentences or phrases that would make people think this.





(F) Splash to Save (1/2) [10 Points]

Ende is a Pahoturi River language spoken by 600 to 1,000 people in Western Province, Papua New Guinea. Ende speakers are often multilingual due to marriages with speakers of other languages (e.g., Taeme and Kawam). However, in this problem we focus only on Ende. The first part of an Ende story is given below, with the English translation of each sentence directly below the Ende.¹

- Ankom a tärko ubi eragwaeya ddobae ai abal nag dagwaeya.
 Ant and Small Fish, they were really very good friends.
- (2) Ankom obo ma me daeya llo toko me.Ant was in his house on top of a tree.
- (3) Tärko walle me daeya ddob kollba nagnag oba peyang.Small Fish was in the river with some fish friends.
- (4) Ttongo ag me d\u00e4be tt\u00e4ng\u00e4m a s\u00e4res\u00e4res\u00e4rem ang gogon.One morning that place was about to rain.
- (5) Yäbäd de ddapall käkan da dakonewän.Clouds covered the sun.
- (6) Malla wätali gogon, wel ulle da gongkamän.It wasn't long before a big wind began.

F1. The English translation for the rest of the story is presented on the next page. The Ende sentences follow, in scrambled order. On your Answer Sheet, match the sentences with their translations by writing the letter of the Ende sentence in the blank by the number for the corresponding English translation.

1. Pronunciation notes:

-*ll* is similar to English r (as in read).







⁻*ä* is a vowel similar to the first syllable of the English word *about*.

⁻*tt* is similar to *ch* in English (as in *church*), except that *tt* is retroflex, meaning that it is pronounced with the tip of the tongue facing the back of the mouth.

⁻dd is similar to the first and last sounds of the English word judge, except that dd is retroflex.

(F) Splash to Save (2/2)

- (7) It blew down and broke the tree.
- (8) The wind tore Ant's house and threw Ant into the river.
- (9) When Small Fish saw, he moved closer to him.
- (10) But some big fish were trying to kill him.
- (11) A catfish was about to swallow him.
- (12) Small Fish quickly splashed with his tail, and with water threw Ant on top of the grass, and he was saved.
- (A) Bunkuttang a mäse ngänygäny e dängkamän.
- (B) Tärko da mängalae källa gokätaemän a ine peyang ankom bom towall toko we daspunän a ttam gogän.
- (C) Ankom bo ma de wel a dapisamän a ankom bom daspunän walle we.
- (D) Be ddob kollba ulleulle da gäz e de ada däganeyo.
- (E) Llo de duduaibnegnän a dattkaemnegnän.
- (F) Tärko da angde ikop dägagän, obo dowae e guinggolän.
- **F2.** Please provide English translations for the following Ende words:
 - a. tärko
 - b. walle
 - c. daspunän
 - d. gongkamän
 - e. yäbäd

F3. Please provide Ende translations for the following words:

- a. Ant
- b. and
- c. tree
- d. catfish

F4. On your Answer Sheet, circle all of the words from the following list that would appear in the Ende translation for "Ant got on top of his house." You should circle at least one word.

tärko, kollba, llo, me, toko, towall, ulleulle, we, wel



(G) Out of Order (1/2) [5 Points]

We are used to hearing sentences one word at a time. Sometimes, this can cause confusion. Think about this sentence:

Katherine believes Lydia is lying.

If you read the sentence from left to right, you would first read *Katherine*, then *believes*, and then *Lydia*. At this point, you would think that Katherine believes Lydia. After reading the rest of the words, you would need to go back and fix your incorrect first impression.

Instead of receiving words one at a time, some computers take in the entire sentence at once. This approach can help avoid the confusion illustrated above. In addition, it is often more efficient because the computer does not have to wait while it processes one word before it can move on to the next one.

For the all-at-once approach to work, the computer needs its input to be in a representation that does not depend on the order of its elements. In this problem, we focus on some representations of this type. To simplify the scenario, we will only talk about sentences as sequences of letters, even though most language technology today is based around larger units such as words. For example, a word-based system would view "good morning" as two words ("good" and "morning"), while a letter-based system would view "good morning" as eleven letters ("g", "o", "o", "d", "m", "o", "r", "n", "i", "n", and "g").

In the Bag

One very simple representation that avoids using a sequence is called a *bag of words* (or, in our case, a bag of letters). For example, the word GREEN would be represented as the bag of letters [E,E,G,N,R]. Note that [E,E,G,N,R] is the same bag of letters as [G,E,N,E,R] or [G,R,E,E,N] - although these look different on the printed page, they are the same to a bag-of-letters computer system.

Using a bag of letters can cause some confusion. Recently, Professor Eliza Shrdlu asked her computer assistant to prepare dinner for an enormous party. The computer asked how many people were attending, and Professor Shrdlu answered OVERFIFTY. Unfortunately, the computer received this answer as the bag of letters [E,F,F,I,O,R,T,V,Y], which the computer misinterpreted: it rearranged these letters to get a smaller number than what Professor Shrdlu had intended. As a result, there was not enough food at the party for all of the guests.

G1. How many people did the computer think would be at the party? You don't need to include any punctuation or spaces in your answer. [HINT: The answer is a number between one and fifty.]

Location, Location, Location

In order to avoid this sort of confusion, we need to insert information about the order of the letters. We can do this by annotating each letter with its position in the sequence.

G2. What message does the following representation encode? You don't need to include any punctuation or spaces in your answer.

[O:16, N:17, E:7, A:21, O:11, E:20, P:10, C:18, M:4, L:19, E:3, D:6, I:15, A:5, I:1, Y:9, S:12, V:2, M:8, T:14, R:22, I:13]



(G) Out of Order (2/2)

Note that, like the bag of letters, the representation at the bottom of the previous page is still not a sequence: the order of the letters must be inferred from the numbers they are associated with, not from the order that the letters appear on the page. The representation would not be changed if we rearranged it as [P:10, E:20, O:11, A:21, Y:9, I:13, T:14, O:16, N:17, I:1, R:22, D:6, M:4, A:5, L:19, V:2, E:3, C:18, E:7, I:15, S:12, M:8]. In other words, this representation provides sequential *information* without sequential *structure*.

G3. There are other possible ways to represent linear position besides the approach shown in G2. What message does the following representation encode? You don't need to include any punctuation or spaces in your answer.

[6:U, 4:P, 10:B, 1:N, 9:A, 5:P, 2:A, 7:K, 3:L, 8:C]

Barking Up the Right Tree

In linguistics, words and phrases are usually analyzed using a tree structure rather than a sequential structure. For example, we might use the following tree for IN THE SEA:



Another way to represent the tree shown above is [E:RRRL, A:RRRR, H:RLLR, I:LL, T:RLLL, S:RRL, N:LR, E:RLR].

G4. What message does the following representation encode? You don't need to include any punctuation or spaces in your answer.

[T:RR, N:LRRL, A:LLRL, U:RLR, R:LLLR, H:LLRRRR, B:LLLL, C:LLRRRL, O:RLL, N:LLRRL, G:LRRR, I:LRL]

Wickelphones

An influential paper by Rumelhart and McClelland (1986) modeled English verbs using representational units called Wickelphones. Wickelphones are named after psychologist Wayne Wickelgren; the *phone* part of the name refers to the fact that they were representing sounds, also known as phones. Since we are instead using letters, we could encode a message using Wickel<u>letters</u>. The word BLUE represented using Wickelletters would be [U:L_E, B:#_L, E:U_#, L:B_U].

G5. What message does the following representation encode? You don't need to include any punctuation or spaces in your answer.

[O:I_N, I:T_O, T:N_O, T:A_T, R:U_S, N:U_D, N:E_T, O:T_Y, N:O_T, D:N_I, A:Y_T, R:R_O, I:D_N, T:T_E, P:#_A, E:T_N, Y:O_O, T:N_I, Y:A_A, G:N_S, S:G_#, U:O_N, O:R_U, N:I_G, O:Y_U, U:S_R, R:U_R, A:P_Y, U:O_R, S:R_U]



(H) From Soup to Nuts (1/2) [15 Points]

Tseltal is a Mayan language spoken by about 590,000 people in the Mexican state of Chiapas. Below are some Tseltal sentences about what people ate, along with their loose English translations. *Sg.* and *pl.* are short for *singular* and *plural*.

Note: A <u>chicken comb</u> is the crest on top of a chicken's head. A <u>gizzard</u> is an organ that birds use for grinding up food. A <u>tamale</u> is a traditional Mesoamerican dish, made of dough that is steamed in a corn husk or banana leaf. An <u>avocado</u> is a fruit that has a large seed and green flesh with green to purple or black skin. As an avocado ripens, it becomes significantly softer; unlike other fruits, however, it does not become sweet. <u>Panela</u> is a block of unprocessed brown sugar. A <u>tortilla</u> is a thin, circular flatbread usually made from corn rather than wheat, as many other breads are. A <u>radish</u> is a crunchy vegetable with a sharp flavor often used in salads. <u>Marrow</u> is a fatty substance contained in bones, sometimes spread over bread or vegetables. <u>Greens</u> refers to green leafy vegetables that cook down into a mushy mass.

	Tseltal	English
1	la jti? te stsalub mute	I ate the chicken comb.
2	la jwe? kaxlan waj	I ate bread.
3	la achik' kaxlan waj ta kajpe	You (sg.) ate bread soaked in coffee.
4	la slo?ik chab	They ate honey.
5	la ati? ?ich	You (sg.) ate chili peppers.
6	la ak'ux kajan	You (sg.) ate my corn.
7	la jwe?tik jwajtik	We ate our tortillas.
8	la sk'ux chenek'	She ate beans.
9	la slo? manko	She ate a ripe mango.
10	la jk'ux te mankoe	I ate the unripe mango.
11	la slo?ik jlo?baltik	They ate our bananas.
12	la ati? te xcha? mute	You (sg.) ate the chicken gizzard.
13	la ak'ux kalwanextik	You (sg.) ate our radishes.
14	la swe?ik te pats'e	They ate the tamales.
15	la sk'ux yaskal	She ate her panela chunks.
16	la jchik'tik ya?lel ti?bal	We ate meat soup.
17	la jk'uxtik awajanik	We ate your (pl.) corn.
18	la awe? te waje	You (sg.) ate the tortillas.
19	la sk'ux bok	She ate raw greens.
20	la jti? awich	I ate your (sg.) chili peppers.
21	la jlo? te ?one	I ate the avocados.
22	la sti?ik te yo?tan mute	They ate the chicken heart.
23	la slo? yonik	She ate their avocados.
24	la alo? achab	You (sg.) ate your (sg.) honey.
25	la sti? sejkub wakax	She ate cow liver.
26	la sk'ux sp'olik	She ate their popcorn.



(H) From Soup to Nuts (2/2)

H1. One of the foods above is an exception: the phrase(s) with this food do not follow the normal rules that determine the rest of Tseltal phrases. Which food is it? Write your answer in English.

Here are some more words in Tseltal:

k'in = kidney, kaxlan chenek' = peanuts, tsukum = stomach, chin bak = marrow, k'oxox = toasted tortillas

H2. Translate the following sentence into English: la jti? sk'in wakax

H3. Translate the following sentences into Tseltal:

- a. We ate the peanuts.
- b. You (pl.) ate cow stomach.
- c. They ate meat.
- d. You (sg.) ate your (pl.) avocados.
- e. I ate my honey.
- f. She ate cooked greens.
- g. She ate bean soup.

H4. One of your friends is learning Tseltal. He translates "I ate marrow" as **Ia jti? chin bak**, and he is told that this translation is incorrect: the correct form has a different word instead of **jti?**. What is this other word?

H5. Another Tseltal learner translates "She ate your (sg.) toasted tortillas" as **la swe? ak'oxox**, but it turns out that this translation is incorrect: the correct form has a different word instead of **swe?**. What is this other word?

H6. Given that ixim means "corn," what plant does kaxlan ixim translate to in English?



(I) A Cornish Conundrum (1/3) [15 Points]

Cornish (known in Cornish as Kernewek) is the Celtic language of the county of Cornwall (Kernow) in South West England. It stopped being actively spoken in the late 1700s, but knowledge of the language survived, and since the early 1900s, attempts have been made to revive it. Currently, there is a growing number of second language speakers, as well as some children who are being brought up speaking Cornish.

11. Complete the table below by filling in the gaps with the correct Cornish words. For the singular English word "frog", the definite singular is "the frog", the plural is "frogs", and the definite plural is "the frogs".

Singular	Definite Singular	Plural	Definite Plural	Meaning
benyn	an venyn	benenes	an benenes	'woman'
gwarier	an gwarier	gwarioryon	an warioryon	'male actor'
kares	an gares	karesow	an karesow	'girlfriend'
pennsevik	an pennsevik	pennsevigyon	an bennsevigyon	'prince'
kwilkyn	an kwilkyn	kwilkynyow	an kwilkynyow	'frog'
davas	an dhavas	deves	an deves	'sheep'
tiek	an tiek	tiogow	an diogow	'male farmer'
gweli	an gweli	gweliow	an gweliow	'bed'
mowes	an vowes	mowysi	an mowysi	ʻgirl'
pluven	an bluven	pluvennow	an pluvennow	'pen'
dehen	an dehen	dehennow	an dehennow	'cream'
dama	an dhama	damyow	a.	'mother'
b.	an gasek	с.	an kasegi	'mare' (female horse)
kulyek	an kulyek	kulyoges	d.	'cockerel' (young male chicken)
myghtern	e.	myghternedh	an vyghternedh	'king'
myghternes	f.	myghternesow	g.	'queen'
tas	h.	tasow	i.	'father'
bogh	j.	boghes	k.	'billy-goat' (male goat)
banow	I.	banowes	m.	'sow' (female pig)
badh	an badh	n.	an badhes	'boar' (male pig)
tevesik	о.	р.	an devesigyon	'adult man'
pons	an pons	ponsyow	q.	'bridge'
maw	an maw	r.	an vebyon	'boy'
S.	an desen	tesennow	an tesennow	'cake'
t.	an tarow	terewi	an terewi	'bull' (male cow)



(I) A Cornish Conundrum (2/3)

Here are some additional Cornish words and phrases:

Cornish	English
war bons glas	'on a blue bridge'
war bons perfydh	'on a perfect bridge'
war desennow perfydh	'on perfect cakes'
pal drosek	'noisy spade' (a spade is a small shovel)
men teg	'beautiful stone'
pel	'ball'
war vyghternedh berfydh	'on perfect kings'
an bal byhan	'the small mine'
bys bras	'thumb'
penn daras	'lintel' (a horizontal beam over the top of a door)
daras melyn	'yellow door'
bys troos	'toe'
war grogen benn	'on a skull'
war dhas deg	'on a beautiful stack'
bys byhan	'pinky finger'
bran bollek	'intelligent crow'
an badhes tronek	'the curious boars'
war dhavas	'on a sheep'
rudhvelyn	'orange' (adjective)
war bluven berfydh	'on a perfect pen'
pel droos	'football'
krogen wynnrudh	'pink shell'
duwes bollek	'intelligent goddess'
gwariores dronek	'curious actress'
bran drosek	'noisy crow'
war weliow glasrudh	'on purple beds'



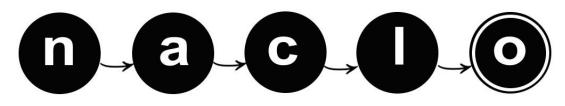
(I) A Cornish Conundrum (3/3)

12. Fill in the blanks in the table below:

Cornish	English	
u.	'the finger'	
v.	'on a beautiful cake'	
w.	'on a door'	
an bal berfydh	х.	
das	у.	
war das	Ζ.	

I3. Some people believe that the English word for a certain animal comes from a phrase meaning "white head" in Welsh, a language that is closely related to Cornish. What is this animal? Write your answer in English.





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

Answer Sheets

REGISTRATION NUMBER					

Name:	
Contest Site:	
Site ID:	
City, State:	
Grade:	

Please also make sure to write your registration number and your name on each page 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 (1/7)

(A) Lines in the Sand

A1. Write one number per box:		
a. filosofi	b. institiut	c. blong
A2. Write one number per box:		
a. village	b. center	c. teach(ing)
d. technology	e. humanity	
A3. The geographical area of the P	acific is:	

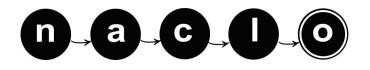
(B) Who Saw the Bear?

B1. Fill in the missing entries:

Fil atladı.	a.
b.	The turtle ran for the fish.

B2. Fill in the missing entries:

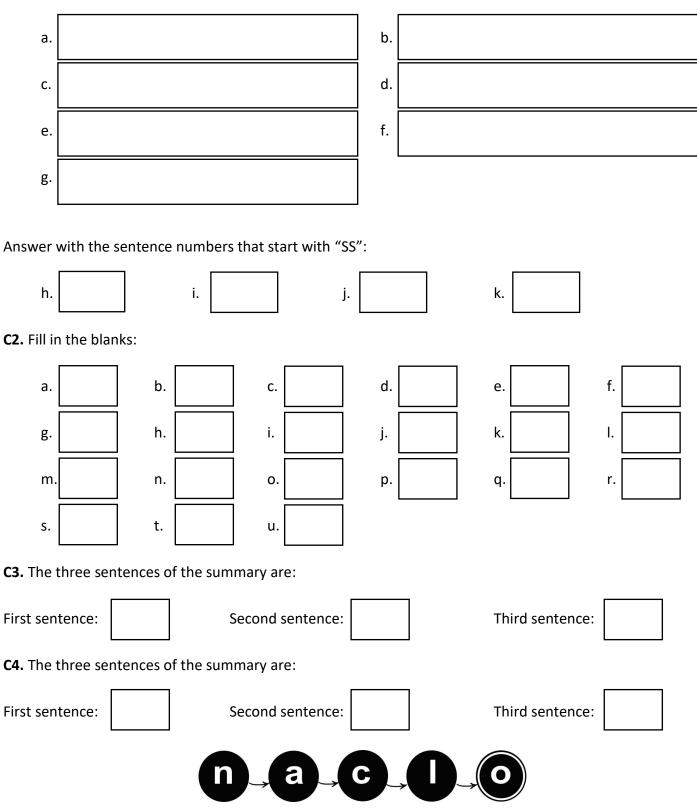
Martı kaplumbağayı gördü.	а.
b.	Who saw the bear?
c.	Whom did the bear see?



Answer Sheets (2/7)

(C) To Make a Long Story Short

C1. Fill in the blanks by choosing from the word bank:



YOUR NAME:

Answer Sheets (3/7)

(C) To Make a Long Story Short (continued)

C5. Answer the following:

- a. The word that should be deleted is:
- b. The two phrases that should be swapped are:
- c. The word that should be replaced is:

It should be replaced by:

C6. Answer the following:

- a. The sentence that the word appears in is:
- b. The word is:

(D) Real Numbers

D1. Fill in the blanks on the blackboard with Kaktovik Iñupiaq numerals:



b.





D2. Give the Iñupiaq word or words for:

a. three	
b. eleven	
c. twenty-two	



YOUR NAME:

REGISTRATION #

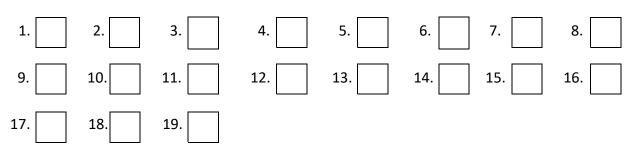
Answer Sheets (4/7)

(D) Real Numbers (continued)

D3. Write in Arabic numerals (i.e., using digits 0-9):								
a. atausiq	. atausiq b. tallimat c. iñuiññaġutaiḷaq							
D4. Write in Arabic numerals (i.e., using digits 0-9):								
a. the Kaktovik Iñupiaq numerals in the top left corner of the blackboard								
b. quliagliaq								

(E) Sleeping in the Shade

E1. Write the letter of the English sentence or phrase that corresponds to each Ik sentence or phrase:



E2. Translate word-for-word into English, in a way that reveals the meaning of each Ik word:

S1. Epa ŋoka kuruo na daą.	Sleeps	in-shade	
S2. Epa ŋoka na ɓets 'ą.			

E3. Translate into Ik. If there are two possible translations into Ik, please provide them both:

- a. these
- b. my huts
- c. I love my wife.
- d. Father is coming from the nice hut.



YOUR NAME:

REGISTRATION #

Answer Sheets (5/7)

(E) Sleeping in the Shade (continued)

E4. Translate into English. If you can think of two structurally different translations into English, provide both:

a. Zeƙwata	oŋorika kuruǫ.				
b. Mina ŋo	ka <i>skaka ntsi</i> ę.				
c. Minima	oŋorika ni epą.				
d. <i>ŋoka na</i>	pcię				
E5. Circle exactly on	e letter to choose th	nat statement a	as the best option:		
(a)	(b)	(c)	(d)		
E6. State a rule that	describes when spe	akers of Ik dev	oice vowels in their	utterances:	
E7. The numbers (1-	19) of two sentence	es or phrases th	at would make peo	pple think this are:	and
(F) Splash to Save					
F1. Write the letter	of the Ende sentenc	e that correspo	onds to each English	sentence:	
7.	8.	9.	10.		12.
F2. Translate into Er	nglish:				
a. tärko			b. walle		
c. daspunän			d. gongkamän		
e. yäbäd					
F3. Translate into Er	Lide:				
a. Ant			b. and		
c. tree			d. catfish		

Answer Sheets (6/7)

(F) Splash to Save (continued)

F4. Circle all of the words from the following list that would appear in the Ende translation for "Ant got on top of his house." You should circle at least one word:

tärk	o	kollba	llo	me	toko	towall	ulleulle	we	wel
(G) Out of	f Order								
G1. The n	umber o	of people t	that the	e comput	er thought	would be at	the party is	:	
G2. The m	iessage i	is:							
G3. The m	iessage i	is:							
G4. The m	iessage i	is:							
G5. The m	iessage i	is:							
(H) From	Soup to	Nuts							
H1. Which	n food is	an excep	tion? W	/rite your	answer in	English:			
H2. Trans	late into	English: I	a jti? sk	'in waka	x				
H3. Trans	late into	Tseltal:							
a.	We ate	e the pea	nuts.						
b.	You (p	l.) ate cov	v stoma	ach.					
c.	They a	ite meat.							
d.	You (s	g.) ate yo	ur (pl.) a	avocados					
e.	l ate m	ny honey.							
f.	She at	e cooked	greens.						
g.	She at	e bean so	up.						

 $1 \rightarrow$

Answer Sheets (7/7)

(H) From Soup to Nuts (continued)

H4. The correct word to use instead of jti? is:

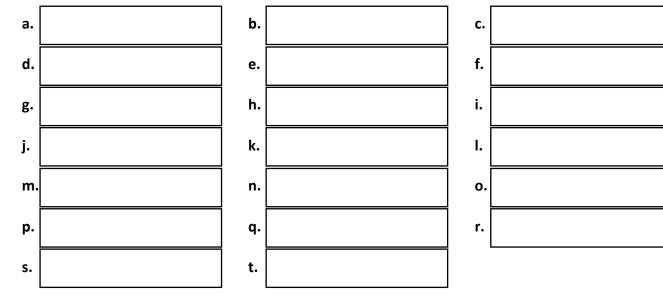
H5. The correct word to use instead of swe? is:

H6. Translate to English: kaxlan ixim

s:

(I) A Cornish Conundrum

11. Fill in the missing places in the table with the correct Cornish words:



I2. Complete the table:

u.	'the finger'
v.	'on a beautiful cake'
w.	'on a door'
an bal berfydh	х.
das	у.
war das	Ζ.

I3. What is the animal? Write your answer in English:

