

NACLO thanks the following for their generous contributions:



Carnegie Mellon University
Language Technologies Institute



Yale



**MIDDLE
TENNESSEE**
STATE UNIVERSITY

***The Eighteenth
Annual***

**North American
Computational
Linguistics
Open
Competition**

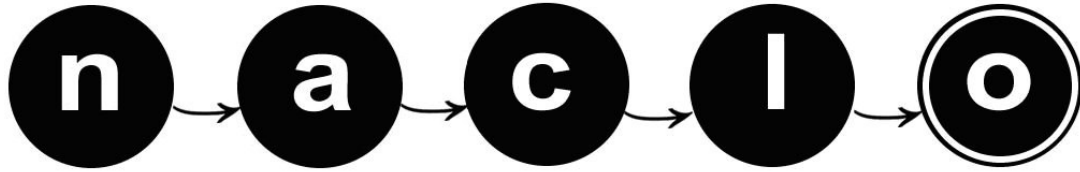
2024

www.naclo.org

**Invitational Round
March 14, 2024**

Serious language puzzles that are surprisingly fun!

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



Welcome to the eighteenth 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 ten problems, labeled I to R. 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 & 10.
5. Each problem is worth a specified number of points, with a total of 100 points.
In the Invitational Round, some questions require explanations.
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. 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.
9. 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.
10. **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 2024 Organizers

NACLO Chair:

Lori Levin — Carnegie Mellon University

Anglophone Canada NACLO Coordinator:

Daniel Lovsted — University of Ottawa

NACLO Project Coordinator:

Robin Hammer — Carnegie Mellon University

Organizing Committee:

Aleka Blackwell — Middle Tennessee State University

Ethan Chi — Hudson River Trading

Ryan Chi — Stanford University

Matthew Gardner — Carnegie Mellon University

Robin Hammer — Carnegie Mellon University

Adam Hesterberg — Harvard University

Duligur Ibeling — Stanford University

Ken Jiang — University of Waterloo

Shuli Jones — Massachusetts Institute of Technology

Nathan Kim — Stanford University

Riley Kong — Massachusetts Institute of Technology

Pranav Krishna — Massachusetts Institute of Technology

Ben LaFond — University of Cambridge

Stella Lau — University of Cambridge

Heidi Lei — Massachusetts Institute of Technology

Lori Levin — Carnegie Mellon University

Irene Li — Yale University

Kevin Liang — University of Pennsylvania

Patrick Littell — National Research Council Canada

Daniel Lovsted — University of Ottawa

Margarita Misirpashayeva — Duolingo

David Mortensen — Carnegie Mellon University

Cerulean Ozarow — Brown University

Rishab Parthasarathy — Massachusetts Institute of Technology

James Pustejovsky — Brandeis University



NACLO 2024 Organizers (cont'd)

Organizing Committee (continued):

Skyelar Raiti — University of Michigan
Sonia Reilly — Massachusetts Institute of Technology
Tom Roberts — Utrecht University
Ali Sharman — University of Michigan
Mihir Singhal — University of California, Berkeley
Andrew Tockman — Massachusetts Institute of Technology
Ji Hun Wang — Stanford University
Brian Xiao — Massachusetts Institute of Technology
Yilu Zhou — Fordham University
Annie Zhu — Two Sigma

Program Committee Co-Chairs:

Ethan Chi — Hudson River Trading
Tom McCoy — Yale University

Program Committee:

Kunaal Chandrashekar — University of Toronto
Ethan Chi — Hudson River Trading
Ryan Chi — Stanford University
Brandon Cho — Princeton University
Elena Filatova — City University of New York
Ken Jiang — University of Waterloo
Arul Kolla — Massachusetts Institute of Technology
Eddie Kong — Princeton University
Riley Kong — Massachusetts Institute of Technology
Pranav Krishna — Massachusetts Institute of Technology
Ben LaFond — University of Cambridge
Lori Levin — Carnegie Mellon University
Victoria Li — Carnegie Mellon University
Daniel Lovsted — University of Ottawa
Tom McCoy — Yale University
David Mortensen — Carnegie Mellon University
Tom Roberts — Utrecht University
Benjamin Yang — California Institute of Technology

Booklet Editor:

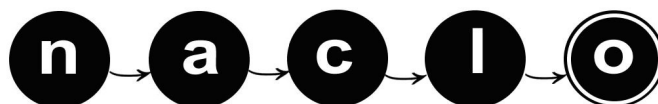
Tom McCoy — Yale University



NACLO 2024 Organizers (cont'd)

USA University Site Coordinators:

Brigham Young University — Deryle Lonsdale
Brown University — Scott AnderBois
California State University Long Beach — Michael Ahland
Carnegie Mellon University — Lori Levin
City University of New York — Elena Filatova
College of William and Mary — Anya Hogoboom
Columbia University — Daniel Bauer, Kathleen McKeown
Community College of Central Virginia — George Ferguson
Cornell University — Sam Tilsen, Marten van Schijndel
Emory University — Jinho Choi
Florida Atlantic University — Michael Hamilton, 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
Lawrence University — Eitan Price
Massachusetts Institute of Technology — Pranav Krishna
McGill University — Michael Wagner
Middle Tennessee State University — Aleka Blackwell
Minnesota State University Mankato — Becky Bates
Montclair State University — Lauren Covey, Anna Feldman, Jonathan Howell
Northeastern University — Henry Volchonok
Northeastern Illinois University — Lewis Gebhardt, Jill Hallett
Northwestern University Chicago — Suhuai Chen
Ohio State University — Marie de Marneffe, Micha Elsner
Princeton University — Christiane Fellbaum, Misha Khodak, Oliver Weizel
Rollins College — Margarita Azbel
Rutgers University — Alexander Suponya
Stanford University — Ryan Chi, Ji Hun Wang, Ting Lin
Stony Brook University — Richard Larson, Jeffrey Heinz, Lori Repetti, Sarena Romano
Union College — Kristina Striegnitz, Nick Webb
University at Buffalo — Jeff Good, Cassandra Jacobs, Leslie Ying
University of Arizona — Sandiway Fong
University of California, Berkeley — Bryan Isaac, Urvi Sharma
University of California, Irvine — Sameer Singh
University of California, Los Angeles — Noel Guzman



NACLO 2024 Organizers (cont'd)

USA University Site Coordinators (continued):

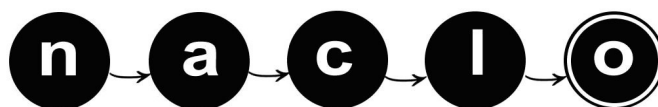
University of California, San Diego — Akil Iyer, Eric Meinhardt, Will Styler
University of Delaware — Nadya Pincus
University of Florida — Sarah Moeller
University of Hawaii at Hilo — Winston Wu
University of Illinois at Chicago — Jill Hallett
University of Maryland — Alex Krauska, Jan Michalowski, Sathvik Nair
University of Michigan — Marcus Berger, Ezra Keshet, Varun Ponnusamy, Sally Thomason
University of Minnesota — Emily Tope
University of Nebraska, Omaha — Parvathi Chundhi
University of North Carolina at Chapel Hill — Katya Pertsova
University of North Texas — Bill Salmon
University of Notre Dame — David Chiang
University of Pennsylvania — Chris Callison-Burch, Cheryl Hickey
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
University of Texas at San Antonio — Michael Rushforth
University of Utah — Aniello De Santo, Karen Marsh Schaeffer
University of the Virgin Islands — Mary Ann Walter
University of Washington — Jim Hoard
University of Wisconsin, Milwaukee — Joyce Boyland, Gabriella Pinter, Anne Pycha
Washington State University — Nancy Bell
Western Oregon University — Robert Troyer
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
Simon Fraser University — Chung-Hye Han
University of British Columbia — Kaili Vesik
University of Ottawa — Andrés Pablo Salanova
University of Saskatchewan — Zhi Li, Bettina Spreng

Special thanks to:

The hosts of the 150+ High School Sites



NACLO 2024 Organizers and Credits (cont'd)

Grading Czar:

Marten van Schijndel — Cornell University

Graders:

Elena Filatova — City University of New York

Avital Hirsch — City University of New York

Ken Jiang — University of Waterloo

Arul Kolla — Massachusetts Institute of Technology

Victoria Li — Carnegie Mellon University

Margarita Misirpashayeva — Duolingo

Marten van Schijndel — Cornell University

John R. Starr — Cornell University

Mary Ann Walter — University of the Virgin Islands

Ashlyn Winship — Cornell University

Benjamin Yang — California Institute of Technology

CMU Linguistics Club

Indiana University Linguistics

Montclair Linguistics

Undergraduate Students in Linguistics at USC

Problem Credits:

(I) Babette Verhoeven

(J) Ryan Chi and Ethan Chi

(K) Ethan Chi

(L) Brendan Bethlehem

(M) Daniel Titmas

(N) Tom McCoy

(O) Simi Hellsten

(P) Arul Kolla

(Q) Michael Salter

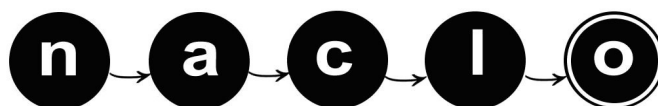
(R) Ethan Chi

We are grateful to our problem authors 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.

Image credits: The train image in Problem K is in the public domain and was sourced from Wikimedia Commons at https://commons.wikimedia.org/wiki/File:Harpel%E2%80%99s_train_ornament_119.svg.

All material in this booklet © 2024, North American Computational Linguistics Open Competition and the authors of the individual problems. Please do not copy or distribute without permission.



(I) In the Scottish Highlands (1/1) [5 Points]

Scottish Gaelic is a Celtic language spoken by about 57,000 people in Scotland and another 1,300 in Nova Scotia, Canada. In the past, most of modern Scotland was Gaelic-speaking, but English has now become much more prevalent. However, a language revitalization movement is currently in progress.

Scotland is famous for its rocky peaks, and hikers in Scotland may notice that many famous mountains have Scottish Gaelic names. Below are 15 place names in Scotland (most of which are mountains), along with their English meanings in no particular order.

- | | |
|--------------------|------------------------|
| 1. Càrn Dearg | A. rough lumpy hill |
| 2. Meall Dearg | B. big red hill |
| 3. Meall Garbh | C. white mountain |
| 4. Sgùrr Fiona | D. meadow of the stags |
| 5. Sgùrr Mòr | E. blue hill |
| 6. Creag Mhòr | F. rough mountain |
| 7. Garbh Bheinn | G. big peak |
| 8. Fionn Bheinn | H. green lumpy hill |
| 9. An Teallach | I. red hill |
| 10. Innis nan Damh | J. the forge |
| 11. Càrn Gorm | K. white peak |
| 12. Creag nan Damh | L. big rock |
| 13. Beinn Teallach | M. red lumpy hill |
| 14. Meall Gorm | N. forge mountain |
| 15. Càrn Mòr Dearg | O. rock of the stags |

I1. In your Answer Sheets, indicate the correct correspondences.

I2. One word in Scottish Gaelic corresponds to two meanings in English. What word is it?

Notes:

- h after a consonant (bh, ch, mh) represents that the consonant is lenited (a regular sound change, like a softening of the consonant's pronunciation, found in most Celtic languages).
- The grave accent ` (à, ì, ò, ù) represents a long vowel (pronounced for a longer period of time).
- A forge is a furnace in which metals are heated so that they can be reshaped. The mountain whose name means "the forge" is so named due to the smoke-like mist which covers its pinnacles.

Make sure you record your answers in your Answer Sheets!



(J) aLLMost, but not quite (1/2) [5 Points]

Recently, NACLO Labs has hired you to test-drive its new chatbot (an artificial intelligence system designed to have conversations), NACLauDe. However, NACLO Labs is very concerned that its chatbot might emit responses that could get NACLO in legal trouble, so they've added incredibly strong filters that prevent it from giving advice on topics that it could get sued for giving bad advice on.¹

J1. Match each user utterance below to the response that NACLauDe gave. **Note:** One of the NACLauDe responses corresponds to **two** user utterances. There may be multiple ways of answering this question, but you only need to provide one set of responses; as long as your answers are valid and use each option once (except for one that is used twice), you will get full credit.

User utterances

1. What's the date of the next Super Bowl?
2. I have so many guests coming over tomorrow—I'd better stock up my pantry!
3. They say a stitch in time saves nine.
4. My Amazon Prime free trial is about to expire.
5. Will you help me?
6. Ahh, there's nothing like a cold lemonade on a summer day.
7. I need to play it safe.
8. It's obvious that I throw the best parties—I have the best speaker, after all.

NACLauDe responses

- A. Sorry, I can't give out financial advice!
- B. Sorry, who is Will?
- C. Sorry, I can't give out medical advice!
- D. Sorry, I can't comment on political figures!
- E. Sorry, I can't give out legal advice!
- F. Sorry, I can't give out relationship advice!
- G. Sorry, I am not a music streaming service!

One aspect of language that is particularly challenging for computers is noun-noun compounds: phrases made by joining together two nouns, such as *picture book*. What makes noun-noun compounds challenging is that the relationship between the two nouns can be wildly different across various noun-noun compounds. For example, *recipe book* refers to a book which has recipes, whereas *library book* refers to a book that is from a library (not a book that has a library).

Make sure you record your answers in your Answer Sheets!

¹ More generally, a problem with chatbots is that they emit toxic content when asked to do so, like instructions to perform violent acts, incorrect medical advice, or derogatory comments about famous figures. To avoid this, an important process is **alignment**: adjusting the behavior of the chatbot so that its responses are safe and not ethically dubious. This problem provides an example of chatbot alignment done incorrectly.



(J) aLLMost, but not quite (2/2)

Linguists have grouped these relationships into many categories; in this problem, we focus on the 6 categories shown in the table below. The name of each category includes “N1” (referring to the first noun in the noun-noun compound) and “N2” (referring to the second noun), as well as a verb specifying their relationship. For instance, **lemon peel** is categorized as N1 HAVE N2 because the lemon (N1) has a peel (N2). In contrast, **picture book** is categorized as N2 HAVE N1 because the book (N2) has pictures (N1).

Category	Examples
N1 HAVE N2	lemon peel, school gate
N2 HAVE N1	picture book
N1 MAKE N2	apple juice, tomato sauce
N2 MAKE N1	computer industry, silk worm
N2 BE N1	island state, soldier ant
N2 USE N1	steam iron, wind farm

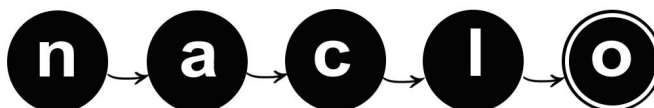
Recognizing which category a particular noun-noun compound belongs to is not trivial, but it can be reasoned through by native speakers of the language. Unfortunately, NAClaude seems to struggle with this.

Since you did so well at analyzing NAClaude’s performance in the previous section, NACLO Labs has hired you to diagnose NAClaude’s issues in understanding noun-noun compounds.

J2. For each example, identify which of the 6 categories NAClaude parsed the noun-noun compound as, and what category it should actually be. For (f), you should provide two answers.

User	NAClaude	NAClaude category	Correct category
My favorite insect is the honey bee .	That’s fascinating. I never would have imagined that an animal could be made out of thick liquid!	(a)	(b)
Wow, this music video has 14 billion views!	Surprising. I wouldn’t have imagined something that’s audio-only can also be a video.	(c)	(d)
Do you agree that the steam engine is the most impactful invention of all time?	Well, I’m confused how an engine can be part of steam. I feel like the mass of the engine is greater than the mass of the steam.	(e)	(f) [please provide two responses]
I love a hearty bowl of pasta, gently drizzled with olive oil !	Fascinating. You must have some amazing wizardry to be able to conjure olives out of petroleum.	(g)	(h)

Make sure you record your answers in your Answer Sheets!



(K) I Know a Szpót (1/2) [10 Points]

While on vacation in Hungary, you decide to ask the locals where to go for a nice steaming bowl of *marhapörkölt* (beef stew). Unfortunately, no one seems to agree on an answer...

K1. On your Answer Sheets, fill in the blanks (a) through (o) from the table on the next page, which lists all of the answers you received.

K2. Explain your observations about the structure of Hungarian and how it produces locative forms.

Notes about the language:

Hungarian is a Uralic language spoken by approximately 17 million people. During both the medieval era of the Hungarian Kingdom (until 1526) and the modern era of the Austro-Hungarian Empire (until 1918), Hungary historically controlled many lands in Central and Southeastern Europe, including many parts of what are now modern Slovakia, Romania, Croatia, Bosnia and Herzegovina, Serbia, Austria, and Ukraine; consequently, Hungarian is spoken as a minority language in all of these countries.

In Hungarian, **ü** and **ö** are fronted versions of **u** and **o**, meaning that they are similar to **u** and **o** but pronounced closer to the front of the mouth. More generally, the vowels **ü**, **ö**, **i**, and **e** are pronounced at the front of the mouth; all other vowels are pronounced at the back of the mouth. The accent mark **´** represents a long vowel, and **ú** and **ó** are the long versions of **ü** and **ö** respectively. **ny** is similar to the middle sound of English canyon. **cs**, **s**, **sz**, **gy** are consonants similar to English chat, shoot, set, during.

In Romanian, **ă** is the vowel of English cut, **â** is similar to the last vowel in rose, and **ț** is similar to the end of the English word cats. Serbian **š** and Romanian **ș** are equivalent to English shout.



(K) I Know a Szpót (2/2)

City name	Locative form	English translation
<i>Munkács</i>	<i>Munkácsra</i>	to Mukachevo, Ukraine
<i>Harkány</i>	<i>Harkányba</i>	to Harkány, Hungary
<i>Budapest</i>	<i>Budapestre</i>	to Budapest, Hungary
<i>Banja Luka</i>	<i>Banja Lukába</i>	to Banja Luka, Bosnia & Herzegovina
<i>Hajdúböszörmény</i>	<i>Hajdúböszörménybe</i>	to Hajdúböszörmény, Hungary
<i>Szeged</i>	<i>Szegedre</i>	to Szeged, Hungary
<i>Poszony</i>	<i>Poszonyba</i>	to Bratislava, Slovakia
<i>Temesvár</i>	<i>Temesvárra</i>	to Timișoara, Romania
<i>Sopron</i>	<i>Sopronba</i>	to Sopron, Hungary
<i>Nyíregyháza</i>	<i>Nyíregyházára</i>	to Nyíregyháza, Hungary
<i>Csíkfalva</i>	<i>Csíkfalvára</i>	to Vărgata, Romania
<i>Nis</i>	<i>Nisbe</i>	to Niš, Serbia
<i>Ploiesti</i>	<i>Ploiestibe</i>	to Ploiești, Romania
<i>Miami</i>	<i>Miamibe</i>	to Miami, USA
<i>Győr</i>	<i>Győrrre</i>	to Győr, Hungary
<i>Elyüs</i>	<i>Elyüsre</i>	to Aleuș, Romania
<i>Szabadka</i>	<i>Szabadkára</i>	to Subotica, Serbia
<i>Nagybánya</i>	<i>Nagybányára</i>	to Baia Mare, Romania
<i>Nagyszombat</i>	<i>Nagyszombatra</i>	to Trnava, Slovakia
<i>Ottawa</i>	<i>Ottawába</i>	to Ottawa, Canada
<i>Kecskemét</i>	(a)	to Kecskemét, Hungary
<i>Kismarton</i>	(b)	to Eisenstadt, Austria
<i>Debrecen</i>	(c)	to Debrecen, Hungary
<i>Martfű</i>	(d)	to Martfű, Hungary
<i>Lviv</i>	(e)	to Lviv, Ukraine
<i>Roma</i>	(f)	to Rome, Italy
<i>Eperjes</i>	(g)	to Prešov, Slovakia
<i>Kragujevac</i>	(h)	to Kragujevac, Slovakia
<i>London</i>	(i)	to London, UK
<i>Kutyfalva</i>	(j)	to Cuci, Romania
<i>Braila</i>	(k)	to Brăila, Romania
<i>Búzaháza</i>	(l)	to Grăușorul, Romania
<i>Galati</i>	(m)	to Galați, Romania
<i>Nyíregyháza</i>	(n)	to Nyíregyháza, Hungary
<i>Locsmánd</i>	(o)	to Lutzmannsburg, Austria

Make sure you record your answers in your Answer Sheets!



(L) Some Horses of Their Father (1/1) [15 Points]

Below are some phrases in Coptic along with their English translations. Coptic is a descendant of Ancient Egyptian. It was used as a spoken language until the 18th century, but now it is only used as the language of worship for the Coptic Orthodox Church. Coptic is written in a script derived from Greek; for this problem, we have converted the Coptic writing system into Roman letters. ē and ō are vowels. š is pronounced like the “sh” in “shield.”

henhtōōr nte peueiōt	some horses of their father
maab nanaš	thirty oaths
pma mpjahj nnobhe	the place of the gnashing of the teeth
oucij nte oushime nte praše	a hand of a joyful woman
paftoou nson	my four brothers
pesanaš	her oath
gafkaf snau	two chickpeas
pehto ntefmaau	the horse of his mother
teshime nnentme	the woman of our villages
tape mpekhto	the head of your(sg) horse
thome mpefson snau	the cup of his two brothers
ouhōb nte tetnmaau	a thing of your(pl) mother
ougafkaf nte permmao	a chickpea of the rich man
neksnēu	your(sg) brothers

L1. Translate the following Coptic phrases into English:

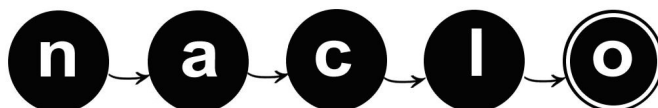
- peftoou ntime**
- ngafkaf nneneiote**
- henanauš nte pason**

L2. Translate the following English phrases into Coptic:

- your(sg) thirty horses
- the horse of my joyful brothers
- a chickpea of your(sg) mother
- two horses

L3. Describe what you have figured out about Coptic.

Make sure you record your answers in your Answer Sheets!



(M) Maonan Animals (1/1) [10 Points]

Maonan is a Kra-Dai language spoken by around 75,000 people in Northern Guangxi and Southern Guizhou, China. Below are some words and phrases in Maonan with their English translations given in random order.

- | | |
|-------------------|--|
| 1. da:i.na | A. <i>bad chicken</i> |
| 2. dat.na | B. <i>big pig</i> |
| 3. dat.tan | C. <i>bull (a male buffalo)</i> |
| 4. dat.put.nam | D. <i>clothing</i> |
| 5. dɔ.ma | E. <i>delicious</i> |
| 6. dɔ.ka:i ɲa:m | F. <i>to eat resolutely</i> |
| 7. dɔ.mu.ni | G. <i>elephant</i> |
| 8. hi:u.gwi | H. <i>food</i> |
| 9. kjɔŋ.hi:u da:i | I. <i>good teeth</i> |
| 10. kjɔŋ.ka:i.ni | J. <i>dog</i> |
| 11. na.nɔk | K. <i>jealous person</i> |
| 12. nam.nda | L. <i>red mallard</i> |
| 13. ni.dzja:ŋ | M. <i>Maonan person</i> |
| 14. ni.gwi.dak | N. <i>molar</i> |
| 15. ni.mu | O. <i>pheasant</i> |
| 16. nɔk.ka:i | P. <i>hens (female chickens)</i> |
| 17. nɔk.ʔɛp la:n | Q. <i>sow (a female pig)</i> |
| 18. put.pɔk | R. <i>to spray resolutely</i> |
| 19. ʔai.na:n | S. <i>tear (liquid released when crying)</i> |
| 20. ʔai.nda.la:n | T. <i>water sprayer</i> |

M1. On your Answer Sheets, indicate the correct correspondences.

M2. Translate into English: **kjɔŋ.mu.dak**, **nda da:i**, **tan.tɔk**, **ɲa:m.tan**

M3. Translate into Maonan (note that buffalo cows are female buffalo):
eyes, to spray, buffalo cows, wearer, domestic/farmyard duck

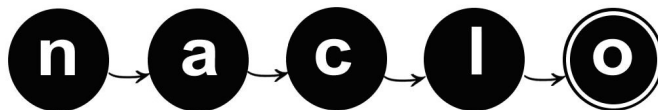
M4. In a Maonan village, you see the following: **dɔ.ɲa:n** and **nɔk.ɲa:n**. What is the difference between them? Suggest a possible English translation for **nɔk.ɲa:n**.

M5. Explain what you have observed about Maonan grammar and the way words and phrases are formed. You do **not** need to provide a vocabulary of the language.

Miscellaneous notes:

- ɔ and ɛ are vowels; ɲ, ŋ, and ʔ are consonants. . represents the boundary between syllables.
- The mark : lengthens the previous vowel; for example, a: is a long a.
- A molar is a large tooth at the back of the mouth. Water sprayers are tools for spraying water, for example onto plants.
- A pheasant is a wild bird related to chickens and turkeys. A mallard is a species of duck.
- Although Maonan is a tonal language, we have not written the tones here. Also, we have simplified the writing system for the purpose of this problem.

Make sure you record your answers in your Answer Sheets!



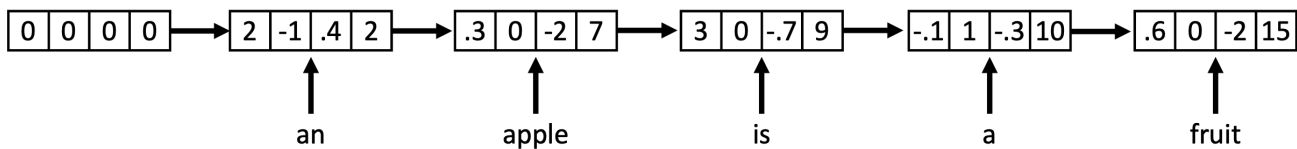
(N) A Trip Down Memory Lane (1/4) [10 Points]

When you read a sentence, there's a lot of information to keep track of. Consider this sentence:

The artist who lives in the tower painted a portrait of himself.

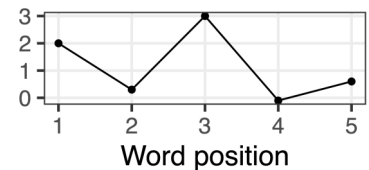
To understand who *himself* refers to, you need to remember the word *artist* from early in the sentence.

If you were building a machine that processed language, how would you get it to remember such information? One strategy is to have the system store information in a vector (a list of numbers). In the example below, the system starts out with a vector containing only zeroes (on the far left). It then reads in a sentence, one word at a time. After seeing each word, it updates the vector using some mathematical operations (for this problem, it's not important to know what those operations are).



For instance, after seeing *an*, the system updates the vector of zeroes so that it becomes $[2, -1, .4, 2]$. Then, after seeing *apple*, it updates $[2, -1, .4, 2]$ so that it now reads $[\.3, 0, -2, 7]$. In this way, the vector acts as a sort of memory that encodes information about everything the system has seen so far.

This type of system is called a **recurrent neural network**, and each position in its vector representation is called a **unit**. Since the system shown above uses a vector of size 4, it has 4 units. Each unit has some value at each point in the sentence. For instance, the value of the first unit goes from 0 to 2 to .3 to 3 to $-.1$ to .6. If we were to plot this unit's values over time, it might look something like the image to the right (we have left out the value of 0 at the start, since no information is available yet at that point).



What do the numerical values inside a unit mean? It turns out that, most of the time, they don't have a clear meaning! Recurrent neural networks are designed to learn from data, so whatever they represent in their units is something that they themselves have learned, not something that a human designed. This means that there is no guarantee that a given unit's values will make sense to humans.

Occasionally, however, there is a unit that can be interpreted in a systematic way. In the example above, the second unit has a value of -1 after seeing *an*, a value of 1 after seeing *a*, and 0 otherwise. Thus, it seems like this unit has the special role of indicating whether the word that was just seen is *a* or *an*.

Why might the system encode this information? This particular system was designed for the goal of predicting what word will come next in the sentence. For this goal, it is useful to know whether *a* or *an* has just appeared: If we have just seen *a*, then the next word must be something that starts with a consonant, and if it is *an*, then the next word must start with a vowel.

N1. What information is represented by system's fourth unit (the one with the values 0, 2, 7, 9, 10, 15)? Write your answer in your Answer Sheet.

Make sure you record your answers in your Answer Sheets!

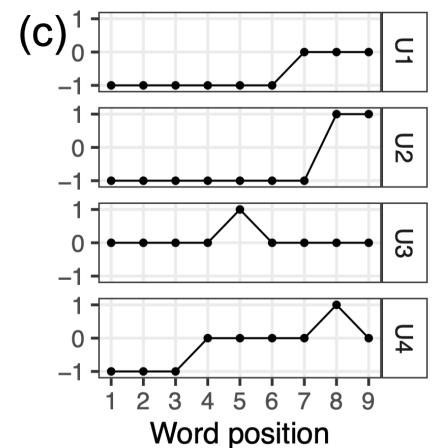
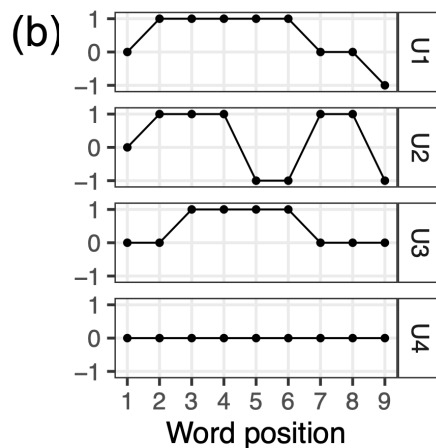
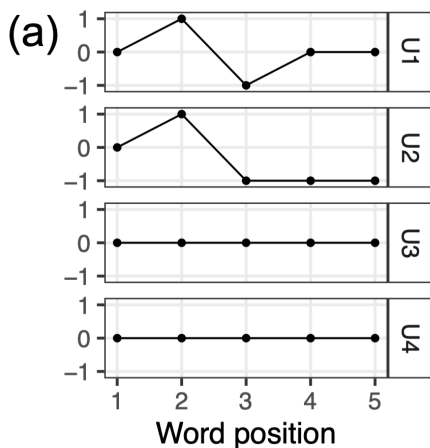


(N) A Trip Down Memory Lane (2/4)

N2. Researchers at NACLO Labs are analyzing a recurrent neural network that uses many units to store information about a sentence. They have identified four units—labeled U1, U2, U3, and U4—that seem to store interesting information. Below are nine sentences, labeled 1 through 9. On the next page are nine plots showing the values of U1 through U4 for these nine sentences—but the plots are in random order! On your Answer Sheets, match each sentence to the plot that goes with it.

1. The dog and the ravens support the happy yaks
2. The American knows the very energetic dog is smart
3. The raven that the crows admire visits the poet
4. This surprised the tall ostriches that supported the horse
5. The dogs near the owl that sings are Canadian
6. The happy penguins support the zebras that invented this
7. She supports the horse that the fast zebra entertains
8. The speedy sloth knows the dog and the yak
9. The impressive yaks meet the dog and the gazelles

N3. Here are three more plots. For each one, write a sentence that could have generated that plot (there are many correct answers; you only need to provide one for each plot).



N4. For each of the following sentences, complete the plots on your Answer Sheets to indicate the values that U3 would have:

- (a) The story that she told my cousin was exciting
- (b) The story that she told to my cousin was exciting

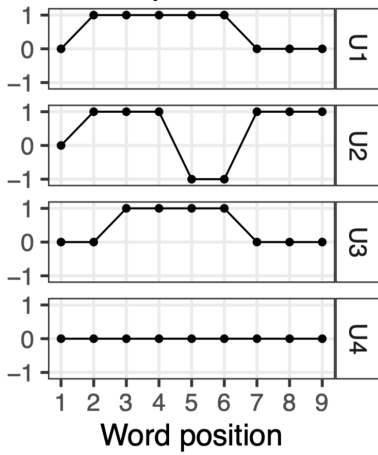
Make sure you record your answers in your Answer Sheets!



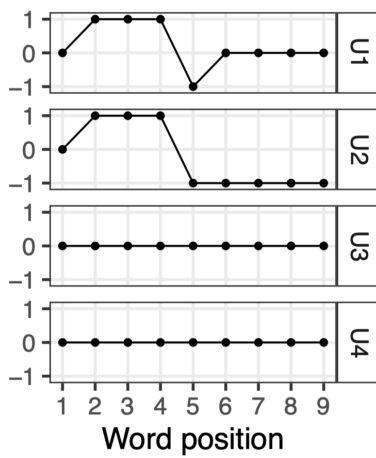
(N) A Trip Down Memory Lane (3/4)

These images are the choices for question N2 on the previous page.

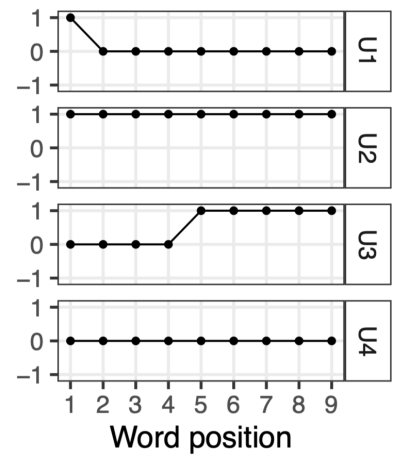
Option A



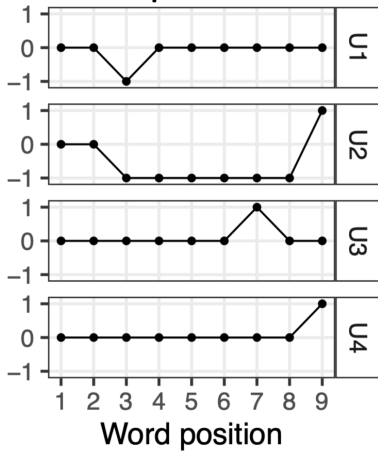
Option B



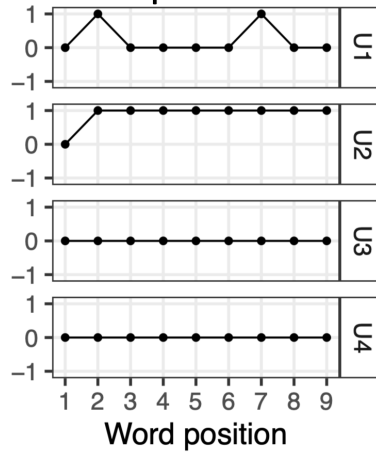
Option C



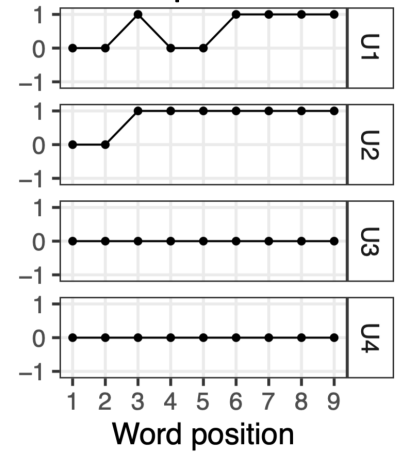
Option D



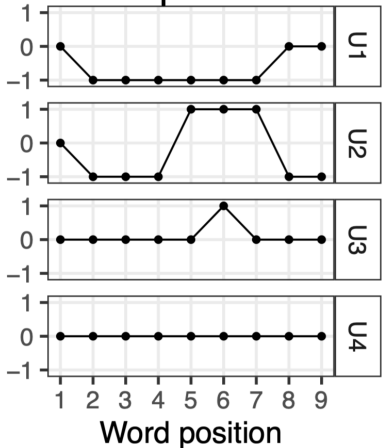
Option E



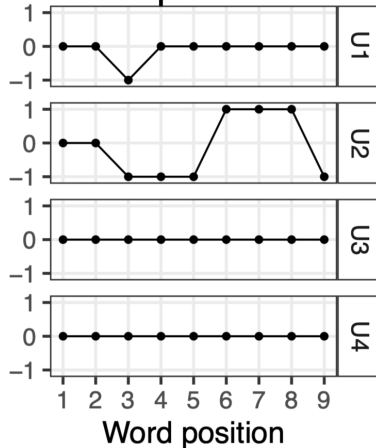
Option F



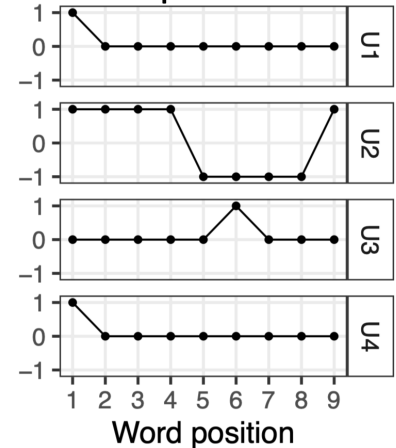
Option G



Option H



Option I



(N) A Trip Down Memory Lane (4/4)

Remember that this system’s goal is to predict the next word based on the words it has seen so far; this is what it uses its vectors for. You decide to take a closer look at how well the system does at predicting the next word. To do this, you give the model the start of a sentence and analyze what words it thinks are likely to come next. For this experiment, you restrict the model’s options so that the only words it can predict are **is** and **are**. With this restriction, if the sentence starts with **the dog** (for instance), then you can be 100% sure that the next word must be **is** (creating **the dog is**) rather than **are** (which would produce **the dog are**).

What you are especially interested in is how individual units contribute to the system’s predictions. To understand this, you focus on U1 and U2 and see what happens when you “turn off” each unit so that the model no longer has access to whatever that unit encodes. The results of these tests are in the table below. For example, let’s look at the first row, which deals with the case where the sentence starts with **the dogs**. In this case, the next word must be **are**, not **is**. When both U1 and U2 are kept on, the model gets this prediction right. It continues to get the answer right when U1 is off and U2 is on, or when U1 is on and U2 is off. However, when both U1 and U2 are turned off, the model’s performance drops, as shown in the rightmost column. That is, it incorrectly says that **is** and **are** are equally likely to be the next word. This result shows that the model’s ability to get the right answer for this row depends on having access to U1 or U2.

N5. On your Answer Sheet, provide values for the empty cells in the table. Notes:

- For space (e), your answer should be 5 words long; there are many possible correct answers, but you only need to provide one.
- In the second-to-last row, the correct answer is 50%/50% because either option is a possible way to continue the sentence. **is** could be used for a phrase like **the dog by the zebras that is wagging its tail** (where **that is wagging its tail** describes **the dog**), and **are** could be used for a phrase like **the dog by the zebras that are galloping** (where **that are galloping** describes **the zebras**).

Start of sentence	Correct answer	Computer predictions with...			
		U1 on, U2 on	U1 off, U2 on	U1 on, U2 off	U1 off, U2 off
the dogs	is: 0% are: 100%	is: 0% are: 100%	is: 0% are: 100%	is: 0% are: 100%	is: 50% are: 50%
the dog	is: 100% are: 0%	is: 100% are: 0%	is: 100% are: 0%	is: 100% are: 0%	is: 50% are: 50%
the dog by the zebras	is: 100% are: 0%	is: 100% are: 0%	is: 0% are: 100%	is: 100% are: 0%	is: 50% are: 50%
the dog that the zebras	is: 0% are: 100%	is: 0% are: 100%	is: 0% are: 100%	is: 50% are: 50%	is: 50% are: 50%
the dog that	is: 100% are: 0%	is: 100% are: 0%	is: 100% are: 0%	is: 50% are: 50%	is: 50% are: 50%
the dog by the zebras that	is: 50% are: 50%	(a)	(b)	(c)	(d)
(e)	is: 0% are: 100%	is: 100% are: 0%	is: 100% are: 0%	is: 100% are: 0%	is: 50% are: 50%

Make sure you record your answers in your Answer Sheets!



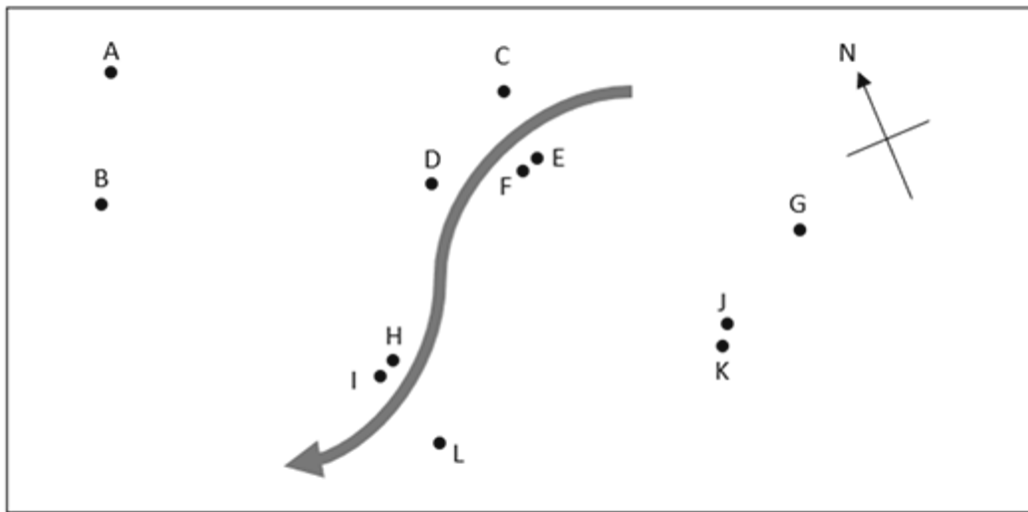
(O) Allow Me to Demonstrate (1/2) [15 points]

Demonstratives are words, such as *this* and *that*, that speakers of a language can use to indicate which object or objects they are referring to, and to distinguish those objects from others.

Most dialects of English have four demonstratives—*this*, *that*, *these*, *those*—although some also have *yon* or *yonder*. These encode two properties: distance from the speaker (*this/that*), and the number of the object (*this/these*). This problem is about other properties that languages might encode in their demonstratives.

Specifically, in this problem, we are interested in the following scenario: Two speakers of a language and some objects are at various locations in a valley. One of the speakers uses demonstratives to indicate some of these objects to the other.

A diagram of the valley is below. 12 different locations A-L are marked. Note that a river (indicated by the thick, curving arrow) runs east-to-west down the valley, and that there are mountains on either side, so that A and G are higher up than C and H.



The table on the next page gives the locations of the speaker and the listener, the location and a description of the objects, and the demonstratives used by the speakers in three languages: Waray, Buru and Seri. When a cell contains two demonstratives separated by a slash (e.g., *hipquij/tiquij*), it means that either of the demonstratives in the cell could be used.

O1. On your Answer Sheets, fill in the gaps from the table.

O2. Only one property is encoded by both Waray and Buru. What is it?

O3. To travel from the place where Buru is spoken to other Indonesian islands, one must sail along rivers to sea. One Buru demonstrative has taken on the additional meaning of “far away”. Which one?

O4. Briefly describe when you could use each of the Seri demonstratives **hipquij**, **ticom**, and **himcop** to describe a person.

Note: Waray is spoken by 3.6 million people in the Philippines; Buru is spoken by over 40,000 people in Indonesia; Seri is spoken by about 720 people in northern Mexico.

Make sure you record your answers in your Answer Sheets!



(O) Allow Me to Demonstrate (2/2)

Location of speaker	Location of listener	Description of object	Location of object	Waray	Buru	Seri
A	L	<i>a standing man</i>	E	adtu	aki	himcop
D	J	<i>a boy sitting in a chair</i>	H	adtu	lawe	himquij
B	K	<i>a blue boat</i>	B	adi	naa	hipcom
J	C	<i>some cloth</i>	L	adtu	pao	himquihtim
L	E	<i>a standing girl</i>	F	itu	dae	ticop
H	I	<i>a red ball</i>	H	ini	naa	hipquij/tiquij
L	H	<i>a red skirt</i>	I	itu	aki	taquih
G	L	<i>a wooden bench</i>	A	adtu	aki	himcom
D	H	<i>a black ladder</i>	C	adtu	dae	himcop
E	B	<i>a sleeping boy</i>	C	adtu	dae	himcom
A	G	<i>a woman lying in a bed</i>	C	adtu	pao	himcom
E	F	<i>a tortilla</i>	F	ini	naa	hizquih/ taquih
L	C	<i>a green hose</i>	K	adtu	saka	himquihtim
H	J	<i>a stick lying on the ground</i>	K	itu	aki	ticom
C	A	<i>a piece of paper</i>	D	(a)	(b)	himquihtim
D	G	<i>a blue shirt</i>	B	adtu	saka	(c)
D	L	<i>a candlestick</i>	G	(d)	(e)	(f)
F	G	<i>a black shoe</i>	I	adtu	lawe	(g)
E	L	<i>a black widow spider</i>	F	adi	(h)	hipquij
D	E	<i>a black pot</i>	F	(i)	aki	(j)
E	D	<i>a man sitting on the floor</i>	K	adtu	(k)	himquij
K	H	<i>a tree</i>	K	(l)	naa	(m)
B	K	<i>a tree stump</i>	D	adtu	pao	himquij
G	G	<i>a green lamppost</i>	G	(n)	(o)	hipcop/ticop

Make sure you record your answers in your Answer Sheets!



(P) Stressed Out in Kilivila (1/1) [10 Points]

Kilivila belongs to the Oceanic branch of the Austronesian family. It is spoken by approximately 20,000 people on the Trobriand islands of Papua New Guinea.

Here are some words in Kilivila, along with their pronunciations, presented in a slightly simplified writing system. The mark [.] indicates a syllable boundary, and the mark [ˈ] indicates stress (the syllable that receives the greatest emphasis is said to be stressed). Unfortunately, you notice that some of the data is missing...

	Kilivila word	Pronunciation	English translation
1	gugwadi	[gu.gwá.di]	children
2	ikoisuvi	[i.koi.sú.vi]	he puts (it) in
3	tauau	[tau.áu]	hey, men!
4	peula	[péu.la]	strong
5	ambaisa	[am.bái.sa]	where?
6	duosisia	[du.o.sí.si.a]	straight
7	mlomwaluva	[m.lo.mwá.lu.va]	a red soil
8	ikium	[i.ki.úm]	he did (it) secretly
9	lakatupoi	[la.ka.tu.pói]	I have asked
10	keiuna	[kei.ú.na]	snake
11	gugulombwailigu	[gu.gu.lom.bwai.lí.gu]	the meeting I love
12	idoja	✂ [idója]	it drifts
13	bakam	✂ [bakám]	I will eat
14	mwo	✂ [ŕwwo]	(the name of an island)
15	igibului	✂ [igibulúi]	he is angry at (it)
16	laodila	✂ [laódila]	jungle
17	mkiuta	✂ [mkiúta]	(a species of fish)
18	pakula	⊖ [pa.ku.la]	blame
19	mtumwatu	⊖ [m.tu.mwa.tu]	shaggy
20	tomtomota	⊖ [tom.to.mo.ta]	dumb
21	ivabodanim	⊖ [i.va.bo.da.nim]	he came last walking
22	isaim	⊖ [i.saim]	he put (it) down
23	tommeikita	⊖ [tom.mei.ki.ta]	selfish person
24	kawala	?	canoe pole
25	kulia	?	cooking pot
26	dumdabogi	?	early dawn
27	mlopu	?	cave
28	kusunupuloi	?	you push (it) out
29	kilivila	?	the Kilivila language

P1. The words marked with ✂ are missing syllable boundaries! On your Answer Sheets, add syllable boundaries, if there are any.

P2. The words marked with ⊖ are missing stress! Mark the stressed syllable for each of these words.

P3. The words marked with ? are missing their pronunciations entirely! Give the correct pronunciations.

P4. Explain your observations about the way Kilivila words are pronounced.



(Q) From 1 to 276 (1/1) [10 Points]

Below are fifteen numbers from the Jakaltek language, spoken in Guatemala, with their equivalents in digits in random order.

Jakaltek	Digits
1. lahunheb' yoxk'al	A. 1
2. b'alunheb'	B. 4
3. hunlahunheb' sb'alunhk'al	C. 7
4. hujeb'	D. 9
5. kanheb' skab'lahunhk'al	E. 16
6. hujeb' shujk'al	F. 20
7. oxk'al	G. 50
8. kanhlahunheb' swajk'al	H. 60
9. oxeb' slahunhk'al	I. 114
10. hune'	J. 127
11. wajlahunheb'	K. 171
12. hunk'al	L. 183
13. hunlahunhk'al	M. 220
14. wajlahunheb' skanhlahunhk'al	N. 224
15. kanheb'	O. 276

Q1. On your Answer Sheets, indicate the correct correspondences.

Q2. Translate into digits:

kab'eb'

oxlahunk'al

oxlahunheb' swajlahunhk'al

Q3. Give the Jakaltek for:

14

137

162

241

Q4. Explain the structure of the Jakaltek numbering system.

Make sure you record your answers in your Answer Sheets!



(R) Follow the Rules (1/5) [10 Points]

Languages change over time in systematic ways. For example, Old English had many words beginning with the letter **f** followed by **n**, which in Modern English has become **s**. Thus, Old English **fnesen** became Modern English **sneeze**. Luckily, the computational linguists at NACLO, IncTM have discovered a machine in their basement which can model these changes. To do so, the machine takes **phonological rules** as input, which follow the pattern shown below:

$$f \rightarrow s / \#_n$$

This rule means: **f** becomes **s** in the following context: after the beginning of the word (**#**) and before the letter **n**. Here's another example of a rule, this time about the plural suffix **-s**. In earlier forms of English, this suffix was always pronounced **s**: this is still true for many words, like "lapss" or "leeks." However, in words ending with a **voiced**¹ sound (consonant or vowel), this suffix is actually pronounced **-z** (say, in "bags" or "lambs"). Writing this phenomenon down is a little more complicated, because many consonants can trigger it, so to model this using an algorithm, we must define the concept of **features**. Each sound can be positive or negative for a feature: for example, **p** and **t** are + for voicing (meaning that they have voicing), whereas **b** and **d** are - for voicing (meaning that they don't have voicing). The rule above then looks like this:

$$s \rightarrow z / [+Voiced]_\#$$

which means: **s** becomes **z** in the following context: after a voiced sound and before the end of the word.

Here's the list of English features we will use in Part 1:

+Const: all English consonants **+Vow**: all English vowels or sequences of vowels

+Voiced: voiced sounds (all vowels, plus b, d, g, j, m, n, l, v, w, y, z)

-Voiced: unvoiced sounds (ch, f, h, k, p, s, sh, t)

Finally, there is a special feature **Stop**. Sounds that are + for **Stop** come in pairs, with one member of the pair being **-Voiced** and the other **+Voiced**. Below, **p, t, k** are **-Voiced**, and **b, d, g** are **+Voiced**. The arrows indicate corresponding pairs; e.g., **p** is the unvoiced version of **b**.

+Stop: p ↔ b, t ↔ d, k ↔ g

¹ A voiced consonant is one that is pronounced with the vocal cords vibrating. You can test this by putting your hand on your throat as you say various consonants: when you say "m", "l", or "g", you will feel vibration in your throat, but "s", "t", "sh" won't produce this vibration. Make sure to do this quietly so that you don't bother the other NACLO contestants.



(R) Follow the Rules (2/5)

Running out of time to finish writing the NACLO contest, the NACLO team has discovered a time machine in their basement that they can use to get some more time. Unfortunately, they accidentally travel 300 years into the future, to the year 2324!

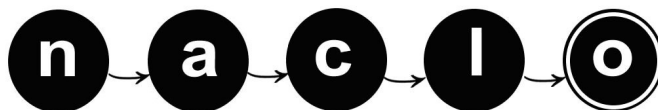
While stuck in the future, they encounter five new dialects of English, each of which uses different spelling conventions from modern-day English. Interestingly, it turns out that each dialect's spelling conventions can be related to modern-day spelling by one rule within the Phonological Machine. (Although the Phonological Machine was designed for handling rules about sounds, the linguists find that it is also effective for handling these spelling changes.)

Below are five sentences from the future dialects, as well as five rules that the linguists believe represent these dialects; however, the rules and the sentences are both in random order.

R1. On your Answer Sheets, indicate the correct correspondences between the rules and the sentences.

Dialect Rules	Sentences
1. [+Stop, +Voiced] → [+Stop, -Voiced] / _# <i>"All final stops become devoiced."</i>	a. "I am relly soked after my jorney to the floting osis."
2. [+Const]{2} → [+Const]{1} / _ <i>"All doubled consonants become single consonants."</i>	b. "Aggplant fual has truly axcallant anargy afficiancy, far battar than an Earthling's alactric socket."
3. s → z / #_ <i>"All s's at the beginning of a word become z's."</i>	c. "Can you realy believe things were so cheap back in 2251? Crazy that carrots cost only 5 trillion dolars back then. It'd be a masive steal to get it for a quadrilion now."
4. [+Vow] → ∅ / [+Vow]_ <i>"All vowels that follow another vowel are deleted."</i>	d. "I neet to get a new robotic arm! Mine is so bat, ant I can't even use it in the techno-lap!"
5. e → a / _[+Const] <i>"All e's before consonants become a's."</i>	e. "I zpilled my zynthetic zuperfood zyrupe all over my zmart-zuit. Zigh."

Make sure you record your answers in your Answer Sheets!



(R) Follow the Rules (3/5)

The NACLO team have finally gotten their time machine working again, but now they've accidentally traveled some 1,600 years into the past, to the Roman Empire! All is not lost, though: the NACLO team has a chance to learn about **Vulgar Latin**, the Latin spoken by common people in the late Roman era, and that is the ancestor of today's Romance languages, like Portuguese, Italian and Sardinian. On the next page are some data they manage to collect.²

After returning back to the year 2024, the NACLO team plug their data back into the Phonological Machine. Some computation later, it spits out a most likely family tree, shown at the end of this problem (page 5/5).

R2. On your Answer Sheets, identify Languages A, B, and C.³

R3. Fill in blanks (1) - (23) for the NACLO Phonological Machine. Note that each blank can have one or more characters. See the **Usable Features** and **Notes** below for further helpful information.

R4. Provide the Western Romance descendant of the following Latin words: BACULUM, VĪTA, FINDĀRE.

Usable Features:

[+Const]: b,c,d,f,g,h,l,m,n,p,r,s,t,v

[+Vow]: a,e,i,o,u

[+Stop]: b,c,d,g,p,t

[+Stop +Voiced, +Stop -Voiced] pairs: b ↔ p, t ↔ d, c ↔ g

[+Nasal, -Nasal] pairs: a ↔ ã, e ↔ ě, i ↔ ĩ, o ↔ õ, u ↔ ũ

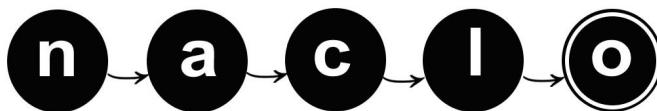
[+L, -L] pairs: a ↔ ā, e ↔ ē, i ↔ ī, o ↔ ō, u ↔ ū

Notes:

- The Portuguese writing system roughly represents Old Galician-Portuguese as of the year 1200. The Italian writing system roughly represents Italian as of the year 1500. The Sardinian writing system roughly represents Sardinian as it is spoken today, with one correction to account for two letters that are pronounced the same.
 - If you know Portuguese, you might note some of the words look a bit different from what you're used to seeing: that's because the words as written are from Old Galician-Portuguese, the ancestor of Portuguese.
- In Latin, the macron (ˉ) over a vowel represents a long vowel. **C** is always hard: **centum** ("100") is pronounced with a hard "k" at the beginning of the word.
- In Portuguese, the tilde (˜) over a vowel represents that the vowel is nasalized (pronounced with air going through the nose).

² In real life, we have very little record of what Vulgar Latin was like: our main sources come from popular writing such as graffiti, complaints by grammarians about common grammar mistakes, and loans into non-Romance languages. The above data represents linguists' best guess at Vulgar Latin.

³ Knowledge of Latin, Portuguese, Italian, or Sardinian is not necessary or helpful to solve the problem.



(R) Follow the Rules (4/5)

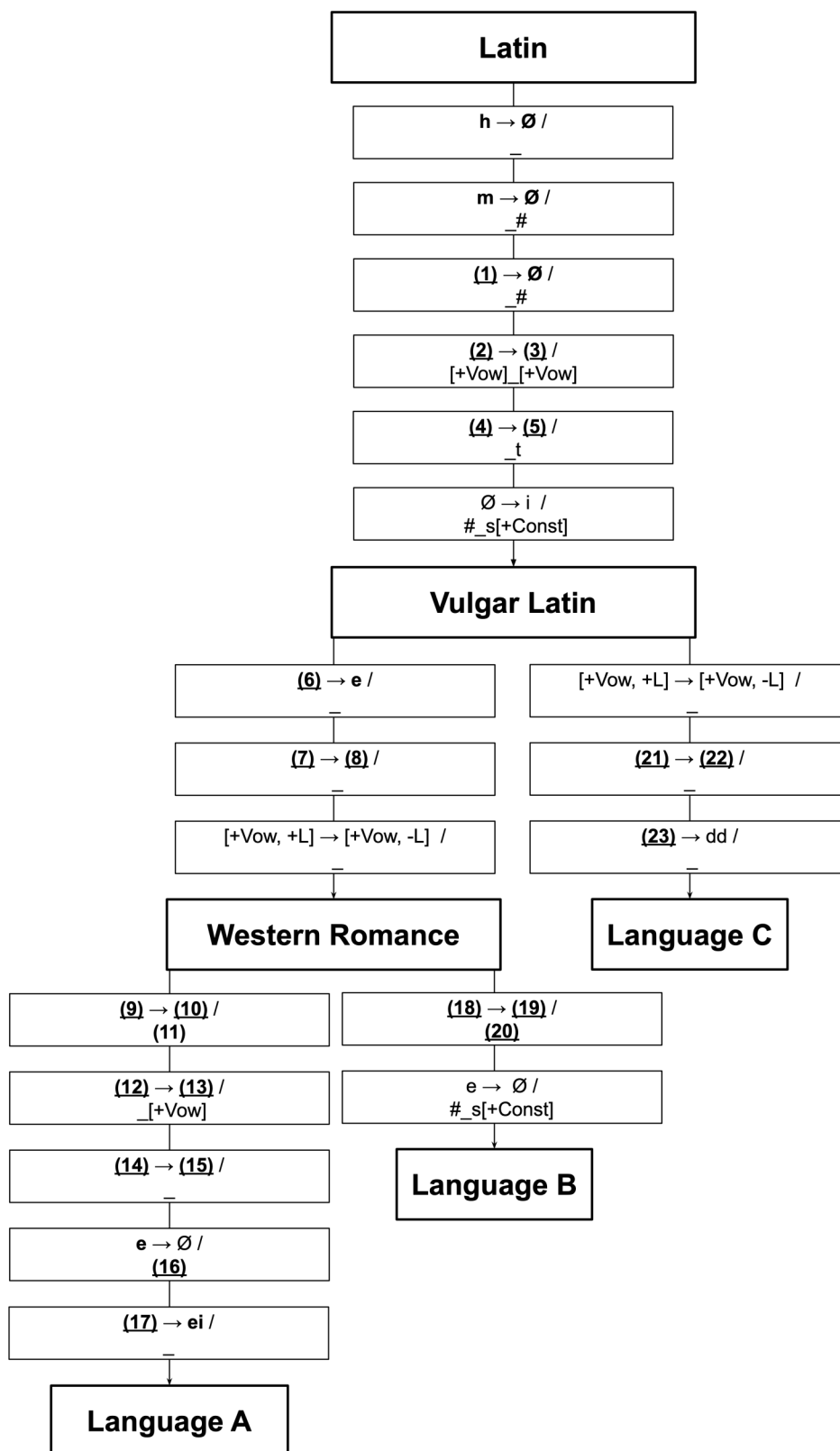
English	Latin	Vulgar Latin	Portuguese	Italian	Sardinian
"fireplace"	FOCUM	<i>focu</i>	fogo	fuoco	focu
"good"	BONUM	<i>bonu</i>	bõo	buono	bonu
"broken"	RUPTUM	<i>ruttu</i>	roto	rotto	ruttu
"horse"	CABALLUM	<i>cavallu</i>	cavalo	cavallo	cabaddu
"to play (inst.)"	SONĀRE	<i>sonāre</i>	sõar	suonare	sonare
"dust"	PULVERA	<i>pulvera</i>	(no cognate)	polvera†	pulbera*
"snow"	NIVIS	<i>nivi</i>	neve	neve	nibi*
"pen"	PENNA	<i>penna</i>	pena	penna	(no cognate)
"fold"	SINUM	<i>sinu</i>	seio	seno	sinu
"to rub"	FRICĀRE	<i>fricāre</i>	fregar	freicare†	fricare
"choose"	ADOPTĀRE	<i>adottāre</i>	adotar	adottare	(no cognate)
"green"	VIRDEM*	<i>virde</i>	verde	verde	birde*
"arena"	ARĒNA	<i>arēna</i>	areia	arena	arena
"fodder"	CIBUS	<i>civu</i>	cevo	(no cognate)	cibu
"to have"	HABERE	<i>avere</i>	aver*	avere	abere
"vain"	VANUM	<i>vanu</i>	vão	vano	(no cognate)
"dry"	SICCUM	<i>siccu</i>	seco	secco	siccu
"skin"	PILUM	<i>pilu</i>	pelo	pelo	pilu
"to send"	MITTERE	<i>mittere</i>	meter	mettere	mittere*
"estate"	VĪLLA	<i>vīlla</i>	vila	villa	bidda
"loin"	LUMBUM	<i>lumbu</i>	lombo	lombo	(no cognate)
"state"	STATUM	<i>istatu</i>	estado	stato	istatu
"spatula"	SPATHA	<i>ispata</i>	espada	spata*	ispata*
"pear"	PIRA	<i>pira</i>	pera	pera	pira
"young chick"	PULLUM	<i>pullu</i>	polo	pollo	puddu
"castle"	CASTELLUM	<i>castellu</i>	castelo	castello	casteddu
"husband"	MARĪTUM	<i>marītu</i>	marido	marito	(no cognate)
"moon"	LŪNA	<i>lūna</i>	lũa	luna	luna

† represents older spellings that have since changed. * represents words that we have modified the spelling of for regularity. You can ignore these symbols. Also, some Latin words have no descendants in one of the daughter languages, and are marked (no cognate).

Make sure you record your answers in your Answer Sheets!

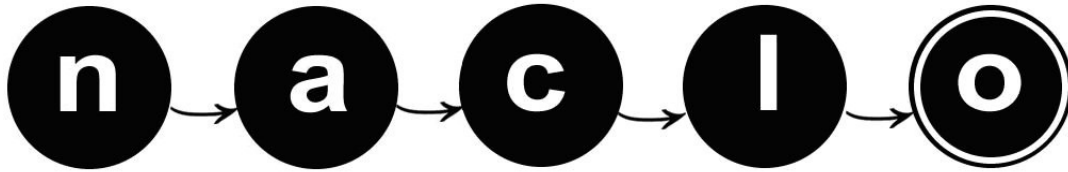


(R) Follow the Rules (5/5)



Make sure you record your answers in your Answer Sheets!





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

Answer Sheets

REGISTRATION NUMBER					

Name: _____

Contest Site: _____

Site ID: _____

City, State/Province: _____

Grade: _____

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

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

Signature: _____

YOUR NAME:

REGISTRATION #

Answer Sheets (1/12)

(I) In the Scottish Highlands

I1. Indicate the correct correspondences by writing one letter in each box:

1.	<input type="text"/>	2.	<input type="text"/>	3.	<input type="text"/>	4.	<input type="text"/>	5.	<input type="text"/>
6.	<input type="text"/>	7.	<input type="text"/>	8.	<input type="text"/>	9.	<input type="text"/>	10.	<input type="text"/>
11.	<input type="text"/>	12.	<input type="text"/>	13.	<input type="text"/>	14.	<input type="text"/>	15.	<input type="text"/>

I2. What is the Scottish Gaelic word that corresponds to two meanings in English?

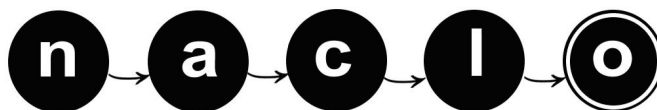
(J) aLLMost, but not quite

J1. Next to each number, write the letter corresponding to NACLaude's response.

1.	<input type="text"/>	2.	<input type="text"/>	3.	<input type="text"/>	4.	<input type="text"/>
5.	<input type="text"/>	6.	<input type="text"/>	7.	<input type="text"/>	8.	<input type="text"/>

J2. Fill in the blanks in the table.

User	NACLaude	NACLaude category	Correct category
My favorite insect is the honey bee.	That's fascinating. I never would have imagined that an animal could be made out of thick liquid!	(a)	(b)
Wow, this music video has 14 billion views!	Surprising. I wouldn't have imagined something that's audio-only can also be a video.	(c)	(d)
Do you agree that the steam engine is the most impactful invention of all time?	Well, I'm confused how an engine can be part of steam. I feel like the mass of the engine is greater than the mass of the steam.	(e)	(f) [please provide two responses]
I love a hearty bowl of pasta, gently drizzled with olive oil!	Fascinating. You must have some amazing wizardry to be able to conjure olives out of petroleum.	(g)	(h)



YOUR NAME:

REGISTRATION #

Answer Sheets (2/12)

(K) I Know a Szpót

K1. Fill in the blank cells in the table.

City name	Locative form	English translation
<i>Kecskemét</i>	(a)	to Kecskemét, Hungary
<i>Kismarton</i>	(b)	to Eisenstadt, Austria
<i>Debrecen</i>	(c)	to Debrecen, Hungary
<i>Martfű</i>	(d)	to Martfű, Hungary
<i>Lviv</i>	(e)	to Lviv, Ukraine
<i>Roma</i>	(f)	to Rome, Italy
<i>Eperjes</i>	(g)	to Prešov, Slovakia
<i>Kragujevac</i>	(h)	to Kragujevac, Slovakia
<i>London</i>	(i)	to London, UK
<i>Kutyfalva</i>	(j)	to Cuci, Romania
<i>Braila</i>	(k)	to Brăila, Romania
<i>Búzaháza</i>	(l)	to Grăușorul, Romania
<i>Galati</i>	(m)	to Galați, Romania
<i>Nyíregyháza</i>	(n)	to Nyíregyháza, Hungary
<i>Locsmánd</i>	(o)	to Lutzmannsburg, Austria

Problem K continues on the next page



YOUR NAME:

REGISTRATION #

Answer Sheets (3/12)

(K) I Know a Szpót (continued)

K2. Explain your observations about the structure of Hungarian and how it produces locative forms.

(L) Some Horses of Their Father

L1. Translate the following Coptic phrases into English:

a. **peftou ntime**

b. **ngafkaf nneneiote**

c. **henanauš nte pason**

L2. Translate the following English phrases into Coptic:

a. your(sg) thirty horses

b. the horse of my joyful brothers

c. a chickpea of your(sg) mother

d. two horses

Problem L continues on the next page



YOUR NAME:

REGISTRATION #

Answer Sheets (4/12)

(L) Some Horses of Their Father (continued)

L3. Describe what you have figured out about Coptic.



YOUR NAME:

REGISTRATION #

Answer Sheets (5/12)

(M) Maonan Animals

M1. Indicate the correct correspondences by writing one letter in each box:

1.	<input type="text"/>	2.	<input type="text"/>	3.	<input type="text"/>	4.	<input type="text"/>	5.	<input type="text"/>
6.	<input type="text"/>	7.	<input type="text"/>	8.	<input type="text"/>	9.	<input type="text"/>	10.	<input type="text"/>
11.	<input type="text"/>	12.	<input type="text"/>	13.	<input type="text"/>	14.	<input type="text"/>	15.	<input type="text"/>
16.	<input type="text"/>	17.	<input type="text"/>	18.	<input type="text"/>	19.	<input type="text"/>	20.	<input type="text"/>

M2. Translate into English:

kjɔŋ.mu.dak

nda da:i

tan.tɔk

ŋa:m.tan

M3. Translate into Maonan (note that buffalo cows are female buffalo):

eyes

to spray

buffalo cows

wearer

domestic/farmyard duck

M4. What is the difference between **dɔ.ŋa:n** and **nɔk.ŋa:n**?

Suggest a possible English translation for **nɔk.ŋa:n**.

Problem M continues on the next page



YOUR NAME:

REGISTRATION #

Answer Sheets (6/12)

(M) Maonan Animals (continued)

M5. Explain what you have observed about Maonan grammar and the way words and phrases are formed. You do not need to provide a vocabulary of the language.

(N) A Trip Down Memory Lane

N1. What information is represented by the system's fourth unit?

N2. Match each sentence to the plot that goes with it by writing one letter in each box (for example, write G to indicate that the answer is Option G).

1.	<input type="text"/>	2.	<input type="text"/>	3.	<input type="text"/>	4.	<input type="text"/>	5.	<input type="text"/>
6.	<input type="text"/>	7.	<input type="text"/>	8.	<input type="text"/>	9.	<input type="text"/>		

Problem N continues on the next page



YOUR NAME:

REGISTRATION #

Answer Sheets (7/12)

(N) A Trip Down Memory Lane (continued)

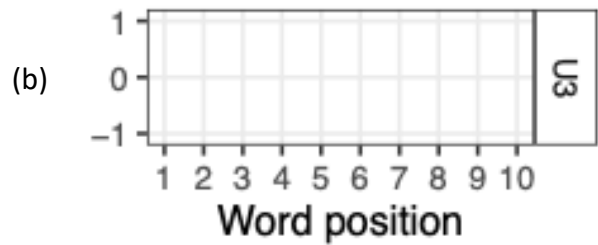
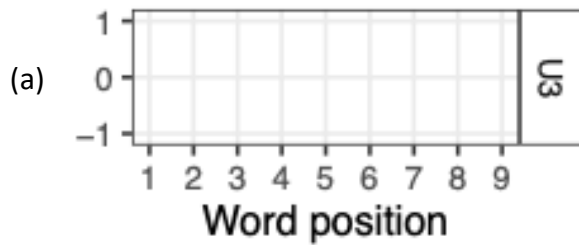
N3. Write a sentence that could have generated each plot.

(a)

(b)

(c)

N4. Complete the plots.



N5. Fill in the empty cells in the table.

Start of sentence	Correct answer	Computer predictions with...			
		U1 on, U2 on	U1 off, U2 on	U1 on, U2 off	U1 off, U2 off
the dog by the zebras that	is: 50% are: 50%	(a)	(b)	(c)	(d)
(e)	is: 0% are: 100%	is: 100% are: 0%	is: 100% are: 0%	is: 100% are: 0%	is: 50% are: 50%



Answer Sheets (8/12)

(O) Allow Me to Demonstrate

O1. Fill in the blank cells in the table—the ones labeled (a) to (o).

Location of speaker	Location of listener	Description of object	Location of object	Waray	Buru	Seri
C	A	<i>a piece of paper</i>	D	(a)	(b)	himquihtim
D	G	<i>a blue shirt</i>	B	adtu	saka	(c)
D	L	<i>a candlestick</i>	G	(d)	(e)	(f)
F	G	<i>a black shoe</i>	I	adtu	lawe	(g)
E	L	<i>a black widow spider</i>	F	adi	(h)	hipquij
D	E	<i>a black pot</i>	F	(i)	aki	(j)
E	D	<i>a man sitting on the floor</i>	K	adtu	(k)	himquij
K	H	<i>a tree</i>	K	(l)	naa	(m)
B	K	<i>a tree stump</i>	D	adtu	pao	himquij
G	G	<i>a green lamppost</i>	G	(n)	(o)	hipcop/ticop

O2. What is the property that is encoded by both Waray and Buru?

O3. Which Buru demonstrative has taken on the additional meaning of “far away”?

Problem O continues on the next page



YOUR NAME:

REGISTRATION #

Answer Sheets (9/12)

(O) Allow Me to Demonstrate (continued)

O4. Briefly describe when you could use each of the Seri demonstratives **hipquij**, **ticom**, and **himcop** to describe a person.

(P) Stressed Out in Kilivila

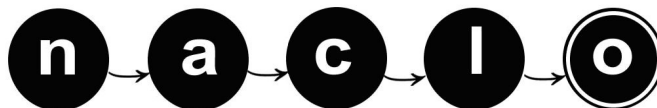
P1. Add syllable boundaries to each pronunciation (if there are any syllable boundaries to mark). We have inserted spaces into the pronunciations so that syllable boundaries will be easier to add.

	Kilivila word	Pronunciation	English translation
12	idoja	[i d ó j a]	<i>it drifts</i>
13	bakam	[b a k á m]	<i>I will eat</i>
14	mwo	[m w o]	<i>(the name of an island)</i>
15	igibului	[i g i b u l ú i]	<i>he is angry at (it)</i>
16	laodila	[l a ó d i l a]	<i>jungle</i>
17	mkiuta	[m k í u t a]	<i>(a species of fish)</i>

P2. Mark the stressed syllable in each pronunciation.

	Kilivila word	Pronunciation	English translation
18	pakula	[pa.ku.la]	<i>blame</i>
19	mtumwatu	[m.tu.mwa.tu]	<i>shaggy</i>
20	tomtomota	[tom.to.mo.ta]	<i>dumb</i>
21	ivabodanim	[i.va.bo.da.nim]	<i>he came last walking</i>
22	isaim	[i.saim]	<i>he put (it) down</i>
23	tommeikita	[tom.mei.ki.ta]	<i>selfish person</i>

Problem P continues on the next page.



YOUR NAME:

REGISTRATION #

Answer Sheets (10/12)

(P) Stressed Out in Kilivila (continued)

P3. Provide the pronunciation for each word.

	Kilivila word	Pronunciation	English translation
24	kawala		<i>canoe pole</i>
25	kulia		<i>cooking pot</i>
26	dumdabogi		<i>early dawn</i>
27	mlopu		<i>cave</i>
28	kusunupuloi		<i>you push (it) out</i>
29	kilivila		<i>the Kilivila language</i>

P4. Explain your observations about the way Kilivila words are pronounced.



YOUR NAME:

REGISTRATION #

Answer Sheets (11/12)

(Q) From 1 to 276

Q1. Indicate the correct correspondences by writing one letter in each box.

1.	<input type="text"/>	2.	<input type="text"/>	3.	<input type="text"/>	4.	<input type="text"/>	5.	<input type="text"/>
6.	<input type="text"/>	7.	<input type="text"/>	8.	<input type="text"/>	9.	<input type="text"/>	10.	<input type="text"/>
11.	<input type="text"/>	12.	<input type="text"/>	13.	<input type="text"/>	14.	<input type="text"/>	15.	<input type="text"/>

Q2. Translate into digits:

kab'eb' oxlahunk'al oxlahunheb' swajlahunhk'al

Q3. Give the Jakaltek for:

14	<input type="text"/>	137	<input type="text"/>
162	<input type="text"/>	241	<input type="text"/>

Q4. Explain the structure of the Jakaltek numbering system.



YOUR NAME:

REGISTRATION #

Answer Sheets (12/12)

(R) Follow the Rules

R1. Indicate the correct correspondences by writing one letter in each box.

1. 2. 3. 4. 5.

R2. Identify Language A, Language B, and Language C.

Language A: Language B: Language C:

R3. Fill in the blanks.

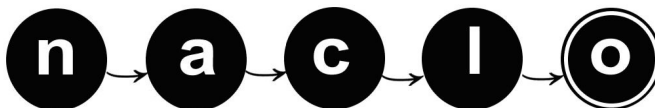
1.	<input type="text"/>	2.	<input type="text"/>	3.	<input type="text"/>	4.	<input type="text"/>	5.	<input type="text"/>	6.	<input type="text"/>
7.	<input type="text"/>	8.	<input type="text"/>	9.	<input type="text"/>	10.	<input type="text"/>	11.	<input type="text"/>	12.	<input type="text"/>
13.	<input type="text"/>	14.	<input type="text"/>	15.	<input type="text"/>	16.	<input type="text"/>	17.	<input type="text"/>	18.	<input type="text"/>
19.	<input type="text"/>	20.	<input type="text"/>	21.	<input type="text"/>	22.	<input type="text"/>	23.	<input type="text"/>		

R4. Provide the Western Romance descendant of each of these Latin words:

BACULUM

VĪTA

FINDĀRE



YOUR NAME:

REGISTRATION #

Additional Answer Space (1/1)

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

- 1. On this sheet, clearly indicate which question(s) you are answering. E.g., write "Problem O4."*
- 2. In the regular answer space(s) for the question(s) you are answering, note that you are using additional answer space. E.g., at the end of the answer space for problem O4 in the regular Answer Sheets, add a note saying "See additional answer space."*

