n a c l o

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

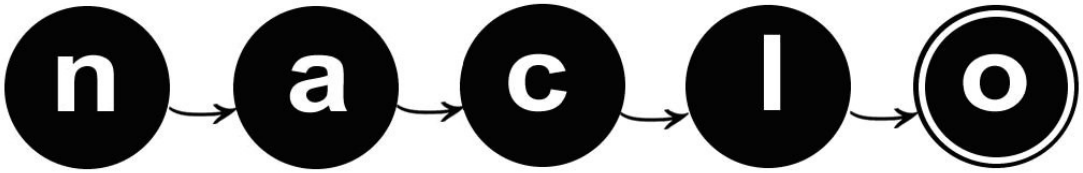
2024

www.naclo.org

Open Round
January 25, 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 eight problems, labeled A to H. Note that this year's contest has one fewer problem than in some previous years.
2. Follow the facilitators' instructions carefully.
3. If you want clarification on any of the problems, talk to a facilitator. The facilitator will consult with the jury before answering.
4. You may not discuss the problems with anyone except as described in items 3 & 11.
5. Each problem is worth a specified number of points, with a total of 100 points. In this year's Open Round, no points will be given for explanations. Instead, make sure to fill out all the answer boxes properly.
6. All your answers should be written clearly in the Answer Sheets at the end of this booklet. ONLY THE ANSWER SHEETS WILL BE GRADED.
7. Write your name and registration number on each page of the Answer Sheets. Here is an example: Jessica Sawyer #850
8. The top 10% of participants (approximately) across the United States and 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 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 — Harvard University

Program Committee Chair:

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

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

NACLO 2024 Organizers (cont'd)

USA University Site Coordinators (continued):

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
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

NACLO 2024 Organizers and Credits (cont'd)

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

Booklet Editor:

Tom McCoy — Yale University

Problem Credits:

(A) Sonia Reilly

(B) Mary Laughren

(C) Michael Salter

(D) Tom McCoy

(E) Arul Kolla

(F) Ethan Chi

(G) Pranav Krishna and Lori Levin

(H) Babette Verhoeven and Simi Hellsten

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

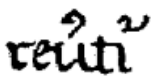
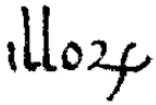
We are grateful for the support of many institutional and individual donors who make this contest possible.

Image credits: The images in Problem E were created by Wikipedia user Widehawk and distributed for free reuse under a CC BY-SA 3.0 license (https://creativecommons.org/licenses/by-sa/3.0/). We modified the images by making them black-and-white; these modified versions may be reused under the same CC BY-SA 3.0 license. The magnifying glass in Problem H was created by Pixabay user OpenClipart-Vectors and is licensed under the Pixabay Content License.

(A) Holy Roman Shorthand (1/3) [10 Points]

Before the printing press was invented, books had to be copied by hand. In Medieval Europe, most of this copying was done by monks. Poor lighting and other factors made this a difficult, often painful, task[1] – so it's no wonder that monks developed an extensive system of abbreviations to speed up their work!

For example, if a monk wanted to write the Latin words *revertitur* and *illorum*, they might have written:

 and 

Below are some examples of abbreviated Latin[2] words you might find in a Medieval manuscript. Note that we've made some formatting changes when typing them up. Next to each one is the full (un-shortened) Latin word it corresponds to. In the right column, one possible English translation of the word is given. The English translations, and knowledge of Latin in general, aren't necessary or helpful to solve this problem.

Abbreviated Latin word	Latin word	English translation
reu^{9}ti^{2}	*revertitur*	she returns
ꝓcul	*procul*	far
su^{9}	*suus*	its own
īp^{9}p7ūū	*imperpetuum*	not lasting
illoꝝ	*illorum*	of theirs
(a)	*conversatio*	conversation
(b)	*fortis*	strong
(c)	*complexio*	combination
(d)	*martis*	of Mars
(e)	*caveret*	he might fear
(f)	*introductum*	introduced

A1. On your Answer Sheets, fill in the blanks (a) to (f) with the number (1-6) of the correct abbreviation from the list below:

1. *f^{o}tis* 2. *īt^{o}ductū* 3. *9u^{9}sa^{o}* 4. *9p^{x}io* 5. *cau^{9}7* 6. *m^{a}tis*

Make sure you record your answers in your Answer Sheets!

1. A quote from the 10th century about manuscript copying reads: "It dims your eyes, makes your back ache, and knits your chest and belly together. It is a terrible ordeal for the whole body."
2. Latin was the language of ancient Rome. In Medieval Europe, Latin continued to be used in many contexts, such as in religious and scholarly settings.

(A) Holy Roman Shorthand (2/3)

A2. There can be many ways to abbreviate a Latin word. Below are 5 abbreviated words. 4 of them stand for the same Latin word. On your Answer Sheet, write the letter of the one that does not match the others.

a. *9probatum* **b.** *cōprobatū* **c.** *comp^{o}batum* **d.** *9probat^{9}* **e.** *9probatū*

In English, we sometimes use the symbol & as an abbreviation for *and.* This symbol originally comes from the Latin word *et*, meaning "and". In some fonts, the way that this symbol is written makes it easier to see how it came from the letter *e* plus the letter *t.* Here is one example:

A3. Some monks used the symbol & in their writing in the same contexts where one of the symbols listed below might be used. Which one? Write the correct letter on your Answer Sheet.

a. *7* **b.** *9* **c.** *ī* **d.** *ō* **e.** *ū*

A4. On your Answer Sheets, give abbreviated forms for the Latin words in the table on the next page, labeled (a) through (f). You should write your abbreviations using the numerical codes shown in the symbol bank at the bottom of the next page, writing one number from 1 to 22 in each box on your Answer Sheet. ***Notes:***

- Some symbols appear multiple times in the symbol bank, always with the same number. You should use a given symbol no more times than it appears in the symbol bank. For example, there are four copies of *n* (number 9), so you should not use the number 9 more than four times in your answers.
- As **A2** showed, there can be more than one way to abbreviate a Latin word. However, for this question (**A4**), there is only one correct abbreviation for each Latin word, because other ways of abbreviating would not satisfy the constraints imposed by the symbol bank.
- Some symbols will be left over in the symbol bank after you complete **A4**—see **A5** for more details.
- There are more boxes on your Answer Sheets than you need; it's okay to leave some blank boxes at the end of an answer.

For example, if you think the answer to (a) is *d* ^{9}, your answer would look like the boxes below, with 3 standing for *d*, 22 standing for ^{9}, and six blank boxes left over:

3	**22**						

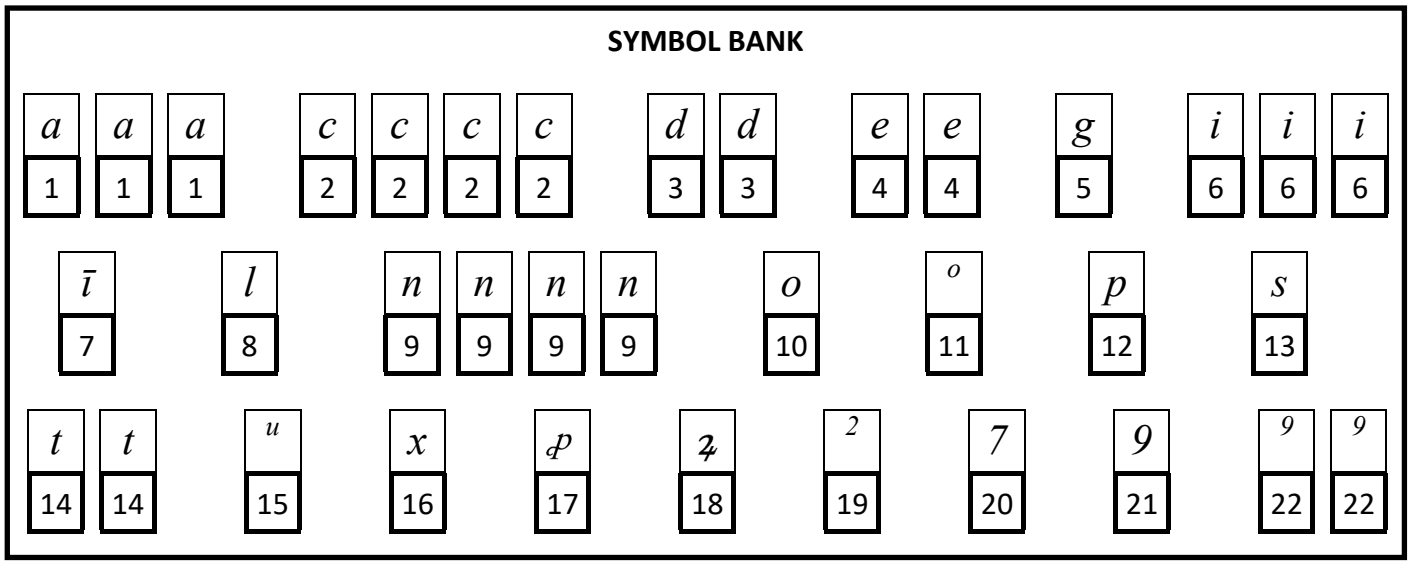
Make sure you record your answers in your Answer Sheets!

(A) Holy Roman Shorthand (3/3)

This table and the symbol bank below it are for problem A4, which is described on the previous page.

	Latin word	English translation
(a)	*procatis*	you (plural) ask
(b)	*notarum*	of the known ones
(c)	*decernimus*	we decide
(d)	*congnitio*	knowledge [alternate form of *cognitio*]
(e)	*crux*	cross
(f)	*petitur*	it is sought for

SYMBOL BANK



A5. If you have completed **A4** correctly, there should be six numbers left over in the symbol bank. The symbols corresponding to these numbers can be arranged into an English word that could be a good birthday gift for a Medieval monk. What is the English word? Write it on your Answer Sheet.

Make sure you record your answers in your Answer Sheets!

(B) Finding Your Place in Warlpiri (1/1) [10 Points]

Warlpiri is a language spoken by about 3,000 people in Central Australia. The Warlpiri people are originally from the Tanami Desert in Australia's Northern Territory; most still live in that area, but many also live in other parts of Australia. In Warlpiri-speaking communities, such as Lajamanu, Yuendumu, Willowra, and Nyirripi, schools offer bilingual education programs in both Warlpiri and English.

B1. Below are some sentences in Warlpiri, along with English translations. On your Answer Sheet, write the sentences that should go in the two blank spaces.

	Warlpiri	**English**
1.	Maliki ka marlungka karrimi.	The dog is standing on the kangaroo.
2.	Maliki ka marlungka nyinami.	The dog is sitting on the kangaroo.
3.	Maliki ka karntangka karrimi.	The dog is standing on the woman.
4.	Marlu ka nantuwurla karrimi.	The kangaroo is standing on the horse.
5.	Karnta ka kuurlurla ngunami.	The woman is lying in the school.
6.	Maliki ka nantuwurla ngunami.	(a)
7.	(b)	The kangaroo is sitting in the school.

Here are some more sentences. Study them, and then answer the questions below.

	Warlpiri	**English**
8.	Nantuwu ka yaparranjirla nyinami.	The horse is sitting on the teenage boy.
9.	Luurnpa ka mardukujarla nyinami.	The kingfisher bird is sitting on the woman.
10.	Yaparranji ka paarlparla nyinami.	The teenage boy is sitting on the calf (of the leg).
11.	Mardukuja ka pirlingka ngunami.	The woman is lying on the hill.
12.	Pirli ka raangka nyinami.	The hill is sitting in the clearing.

B2. Two of the Warlpiri words in the following list have the same meaning as each other. On your Answer Sheet, circle these two words: *karnta, luurnpa, mardukuja, ngunami, pirli*

B3. Answer each of the following questions by writing A or B on your Answer sheet.

i. *Wati* means "man", so "on the man" is:
A. watirla B. watingka

ii. *Warlu* means "fire", so "in the fire" is:
A. warlurla B. warlungka

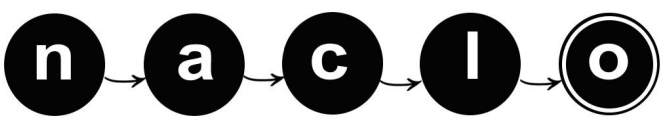
iii. *Jaaji* means "church", so "in the church" is:
A. jaajirla B. jaajingka

iv. *Wulpayi* means "creek", so "in the creek" is:
A. wulpayirla B. wulpayingka

v. *Ngurra* means "home", so "at home" is:
A. ngurrarla B. ngurrangka

vi. *Yama* means "shade", so "in the shade" is:
A. yamarla B. yamangka

Make sure you record your answers in your Answer Sheets!



(C) Happy Birthday! (1/2) [10 Points]

The Ewe language is spoken principally in Ghana, Togo, and Benin. Names in Ewe are often connected to the day on which a person was born. (Not all Ewe names are related to date of birth, but all the names used in this problem are.) Below are some dates along with one possible name for a child born on each date. For your convenience, we have also provided calendars for January 2024 and February 2024 on the next page. All dates in this problem are from 2024. "ɔ" is a vowel sound a little like the "aw" in "paw".

January 7	Kɔsi (male child)	January 27	Kwami (male child)
January 9	Kɔbla (male child)	February 6	Abla (female child)
January 11	Yawa (female child)	February 11	Akosua (female child)
January 14	Kwasi (male child)	February 13	Kwamla (male child)
January 15	Kɔdzo (male child)	February 22	Yao (male child)
January 20	Ami (female child)	February 26	Adzoa (female child)
January 24	Aku (female child)		

C1. On your Answer Sheets, indicate whether each name below would typically be a female name or a male name. You should record your answers by writing one letter in each box: write F for female or M for male.

a. Awusi **b.** Kwadzo **c.** Adzo **d.** Kofi **e.** Afiwa

C2. On your Answer Sheets, match each date below to a name that could be given to a child born on that day. Some of the names listed here also appear in the examples above. Each name will be used exactly once.

(i) January 10	(a) Afua
(ii) January 13	(b) Kɔku
(iii) January 16	(c) Yao
(iv) January 26	(d) Adzoa
(v) January 29	(e) Kɔmi
(vi) February 8	(f) Abla

C3. On which day could a child named Kofi have been born? Write the correct letter on your Answer Sheet.

(a) January 12 (b) January 18 (c) January 23

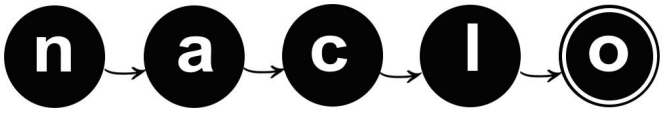
C4. Which of these names might be given to a child born on February 5? Write the correct letter on your Answer Sheet.

(a) Kwadzo (b) Afi (c) Kwaku

C5. In Ewe, the word *gbe* means "day." Translate each of the following Ewe words into English:

i. Dzoɖagbe ii. Fiɖagbe

Make sure you record your answers in your Answer Sheets!



(C) Happy Birthday! (2/2)

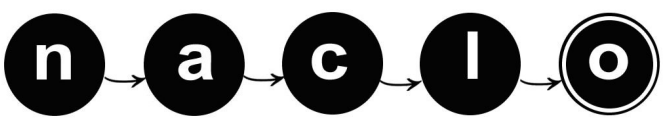
Below are calendars for January and February 2024, which may be helpful for answering the questions on the previous page.

January

Sunday	Monday	Tuesday	Wednesday	Thursday	Friday	Saturday
	1	2	3	4	5	6
7	8	9	10	11	12	13
14	15	16	17	18	19	20
21	22	23	24	25	26	27
28	29	30	31			

February

Sunday	Monday	Tuesday	Wednesday	Thursday	Friday	Saturday
				1	2	3
4	5	6	7	8	9	10
11	12	13	14	15	16	17
18	19	20	21	22	23	24
25	26	27	28	29		



(D) Generally Speaking (1/4) [20 Points]

Think about these two examples:

A. The hummingbird interviewed the penguin.

B. The the interviewed penguin hummingbird.

It's unlikely that you have seen either of these examples before. Nonetheless, you can probably tell that example A is a possible English sentence (it's certainly strange, but it is **grammatical**—that is, it obeys all the rules of English sentence structure), while example B could not be an English sentence.

How are we able to make judgments about sentences we have never seen before? This question relates to the broader topic of **generalization**—applying what you have learned to new situations or examples. Generalization is an important concept in computational linguistics because we want to build computers that can effectively generalize to sentences they have never seen before.

At NACLO Labs, to study generalization, we have created four different computers that are able to learn from data. The computers are named C1, C2, C3, and C4, and they have never encountered any English sentences before our experiment. We first show the computers the 17 sentences below, which we will refer to as the **training set**—the set of examples that the computers learn English from. We tell the computers that all 17 of these sentences are grammatical.

Training set:

1. The linguist met the programmer.
2. The cartoonist saw the spy.
3. The watchmaker is asleep.
4. The happy concierge met the ballerina.
5. The spy saw the woodcarver.
6. The spy is knowledgeable.
7. The main detective visited the ballerina.
8. The main calligrapher saw the astronaut.
9. The programmer visited the blacksmith.
10. The detective met the yodeler.
11. The spy is happy.
12. The ballerina is famous.
13. The linguist is tall.
14. The talented watchmaker visited the astronaut.
15. The famous programmer visited the blacksmith.
16. The cheerful concierge met the ballerina.
17. The haberdasher saw the linguist.

Note that all of the nouns in these sentences are types of occupations (for example, a haberdasher is someone who sells men's clothing), but for this problem you do not need to know the meanings of these words.

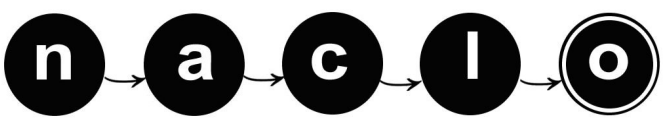
We now show the computers 21 more sentences, which are listed in the table on the next page. For each sentence, we ask each computer whether it thinks the sentence is **grammatical (G)** or **ungrammatical (U)**. Their answers are recorded in the table. We have also included the answers that a human provided when shown the same sentences.

(D) Generally Speaking (2/4)

D1. Some of the cells in the table are empty. On your Answer Sheets, provide the missing entries.

	Sentence	C1	C2	C3	C4	Human
18.	The main detective visited the ballerina.	G	G	G	G	G
19.	The linguist is tall.	G	G	G	G	G
20.	The cartoonist saw the spy.	G	G	G	G	G
21.	The linguist visited the spy.	U	G	G	G	G
22.	The main concierge saw the blacksmith.	U	G	G	G	G
23.	The ballerina is asleep.	U	G	G	G	G
24.	The cartoonist met the programmer.	U	G	G	G	G
25.	The woodcarver visited the programmer.	U	U	G	G	G
26.	The linguist saw the haberdasher.	U	U	G	G	G
27.	The famous concierge visited the watchmaker.	U	U	G	G	G
28.	The linguist is cheerful.	U	U	G	G	G
29.	The calligrapher watchmaker visited met.	U	U	U	G	U
30.	The knowledgeable cheerful talented.	U	U	U	G	U
31.	The tall detective saw the ballerina.	U	(a)	G	G	G
32.	The linguist saw the yodeler.	U	(b)	G	G	G
33.	The visited met linguist programmer.	U	(c)	U	G	U
34.	The yodeler met the woodcarver.	U	(d)	G	G	G
35.	The watchmaker is famous.	U	(e)	G	G	G
36.	The detective met the yodeler.	(f)	(g)	G	(h)	G
37.	The spy is talented.	(i)	(j)	G	(k)	G
38.	The programmer met the cartoonist.	(l)	(m)	G	(n)	G

Make sure you record your answers in your Answer Sheets!



(D) Generally Speaking (3/4)

D2. The table below contains some more sentences, along with C2 and C3's judgments for those sentences. In each sentence, one word has been hidden from you by replacing it with the string HIDDEN_WORD_1, HIDDEN_WORD_2, or HIDDEN_WORD_3. On your Answer Sheeet, write down what each HIDDEN_WORD stands for, choosing from the following options: ***asleep***, ***happy***, or ***main***. Additional notes:

- You should use each option (***asleep***, ***happy***, or ***main***) exactly once.
- Each HIDDEN_WORD stands for the same word in both of the sentences where it appears.

	Sentence	C2	C3
39.	The watchmaker is **HIDDEN_WORD_1**.	G	G
40.	The **HIDDEN_WORD_1** concierge visited the astronaut.	G	G
41.	The watchmaker is **HIDDEN_WORD_2**.	U	G
42.	The **HIDDEN_WORD_2** concierge visited the astronaut.	G	G
43.	The watchmaker is **HIDDEN_WORD_3**.	G	G
44.	The **HIDDEN_WORD_3** concierge visited the astronaut.	U	G

D3. For each sentence in the table above, think about what your own labels would be. Out of these 6 sentences, how many are there where C2 produced the same label as you? How many are there where C3 produced the same label as you? Write your answers on your Answer Sheet.

⇒ ***NOTE:*** Different people may have different judgments about these sentences. Therefore, as long as you write *something* in your Answer Sheet for **D3**, we will give you credit—it doesn't matter what you write! We just want you to think a bit about which computer matches your own judgments the best.

Ideally, we would want our computers to make exactly the same predictions as a human would. If you look back at the previous page (sentences 18 to 38), you will see that C1 and C2 often **undergeneralize**—there are sentences, such as sentence 25, that these computers label as ungrammatical (U) even though the human called them grammatical (G).

In the other direction, C4 sometimes **overgeneralizes**—it labels some sentences as grammatical (G) even though the human labeled them as ungrammatical (U). Depending on how you answered question **D3**, there may also be some examples of undergeneralization or overgeneralization in sentences 39 to 44 as well.

The examples of overgeneralization that we've seen so far were about the domain of sentence structure. Overgeneralization can also happen on the level of individual words. For example, based on pairs of words like *walk* and *walked*, or *laugh* and *laughed*, a computer might overgeneralize and produce the word *thinked* (based on *think*) or *bringed* (based on *bring*).

How might word-level overgeneralization affect our interactions with a computer? As an illustration, the next page provides some interactions between a human and a hypothetical robot. The robot has been trained on a very large amount of English text, and it often overgeneralizes when forming words. Each word that resulted from overgeneralization has been hidden from you by being encoded in a cipher, where each letter has been replaced with a different letter. The same code was used for all examples—for example, if **X** stands for w in the first encoded word, then it will also stand for w everywhere else.

(D) Generally Speaking (4/4)

D4. On your Answer Sheets, decode each encoded word from the conversations below. We are not providing the robot's training set, so it will not be possible to understand the robot as precisely as C1, C2, C3, and C4 from the previous part of the problem. However, if you carefully analyze what the robot says and also figure out the code, it should be possible to determine what each encoded word stands for.

Human: How severe will the storm be? Do you think it will interrupt my WiFi?

Robot: The storm will have a high level of severity. An interruption to your WiFi is likely, especially if lightning strikes somewhere nearby.

Human: Are you able to make this picture smaller for me?

Robot: I would be happy to **QGTRHUU** it for you! The **RKGKHENDKEW** of the smaller version should make it easier to move around. And I would be happy to enlarge it again later.

Human: That's a strange way to phrase it…but thank you! Also, something looks a little odd about the picture—can you tell what it is?

Robot: The picture is currently **UQYETKFQ-DKAPE**. I would be happy to apply a mirror filter to fix that. The website where you got the picture has many shortcomings—one of them is that it often distorts images.

Human: *[showing a picture]* What is the largest object in this picture?

Robot: *[pointing at object in picture]* That!

Human: Where is the water bottle in the picture?

Robot: *[pointing at corner of picture]* There!

Human: Who is the tallest person in the picture?

Robot: *[pointing at person in picture]* **EPX**!

Human: Can you attempt to schedule an appointment with my hairdresser?

Robot: Yes, I will make this **HEEQRMEKXG** soon. I will request an appointment **TXRQCPQG** next week.

Human: *[returning after a two-week vacation]* It's great to be back in the office! Did anything happen while I was away?

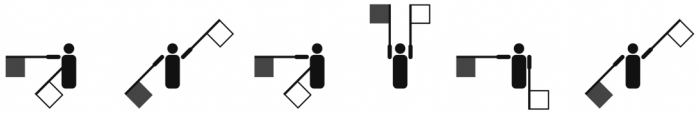
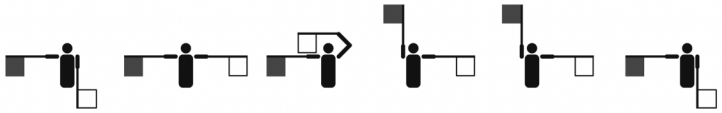
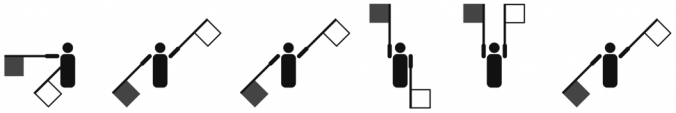
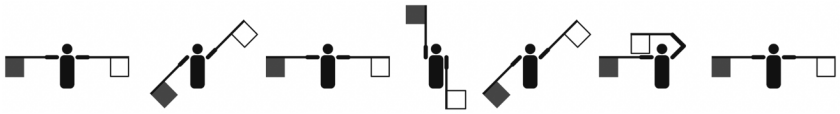
Robot: Not much happened for the first week. But then, **WQTEQDCQQB**, we got two new clients! They thought our company had many **EHUUJXRKGAT**, so they were happy to hire us.

An important challenge in computational linguistics is figuring out how to get computers to generalize in the right way. Generalizing can also be a challenge for humans acquiring languages. For example, when acquiring English, children often overgeneralize by producing words such as *thinked* or *bringed.*

Make sure you record your answers in your Answer Sheets!

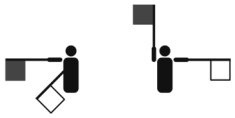
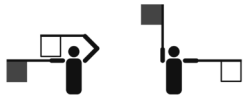
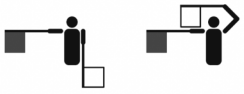
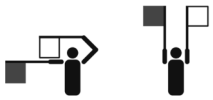
(E) Japanese From Afar (1/1) [10 Points]

Flag semaphore is a technique that uses flags to communicate, best known for enabling sailors to talk across long distances. Below are several Japanese words written in three different ways: in a standard Japanese script called katakana, in the Latin script (the same set of letters that are used to write English), and in the Japanese version of flag semaphore. Definitions are also provided, but they are not needed to solve the problem. Knowledge of Japanese is also not needed.

Katakana	Latin Script	Japanese Flag Semaphore	Definition
ノミ	***nomi***		flea
アフリカ	***afurika***		Africa
コモロ	***komoro***		Comoros
アイツ	***aitsu***		that person
ナオミ	***naomi***		Naomi

E1. Below, numbered (i) through (vii), there are several more katakana characters along with their equivalents in the Latin script. On your Answer Sheets, match each of these to the correct Japanese Flag Semaphore example from A through H. Each of your answers should be written as one letter (A, B, C, D, E, F, G, or H). Be warned: one of these eight letters will remain unused!

i) テ ***te*** ii) ヨ ***yo*** iii) ア ***a*** iv) ミ ***mi*** v) モ ***mo*** vi) ヌ ***nu*** vii) セ ***se***

A. 	B. 	C. 	D. 
E. 	F. 	G. 	H. 

E2. The letter that was not used in **E1** corresponds to one of the characters below. Which one? Circle your answer on your Answer Sheets.

メ ***me*** 井 ***wi*** ク ***ku*** へ ***he*** ホ ***ho*** シ ***si***

<u>Make sure you record your answers in your Answer Sheets!</u>

(F) Paiwan in Progress (1/1) [10 Points]

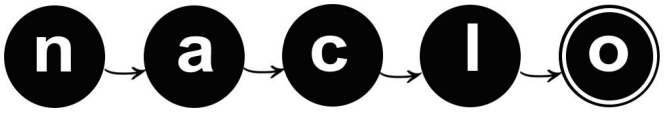
Paiwan is the language of the Paiwan people, one of the indigenous peoples of Taiwan. There are nearly 100,000 Paiwan people, and approximately 15,000 of them are native speakers of the language.

Below are some Paiwan verbs given in two forms: their root form and their progressive form. (The root form is a verb's basic form, and the progressive is the form that, in English, has the ending *–ing*.) The meanings of the verbs are provided for interest, but they are not necessary for solving the problem.

F1. Some entries in the table have been left blank. Fill them in on your Answer Sheet.

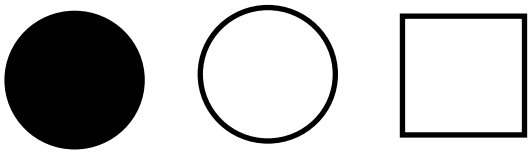
Root Forms		**Progressive Forms**	
keman	*to eat*	**kemakan**	*eating*
ivu	*to talk*	**ivuivu**	*talking*
kivata	*to ask*	**kivatavata**	*asking*
lemumay	*to hit*	**lemumalumay**	*hitting*
temekel	*to drink*	**temeketekel**	*drinking*
kalidjikidj	*to flash*	**kalidjikidjikidj**	*flashing*
paderua	*to hiccup*	**paderuarua**	*hiccupping*
semutjiray	*to spit*	**semutjiratjiray**	*spitting*
kemesa	*to cook*	**kemesakesa**	*cooking*
temulu	*to teach*	(a)	*teaching*
djemavac	*to walk*	(b)	*walking*
lemenguaq	*to fix*	(c)	*fixing*
mikerekel	*to shake*	(d)	*shaking*
takalava	*to wait*	(e)	*waiting*
masasevalit	*to take a turn*	(f)	*taking turns*

Make sure you record your answers in your Answer Sheets!



(G) At a Loss for Words (1/5) [15 Points]

Researchers at NACLO Labs are working on machine-to-machine communication, and they need your help with something! They are trying to program two robots to talk to each other about cards with shapes on them. Here are some shapes they are using:

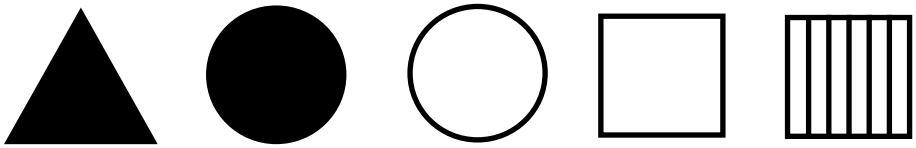


The two robots are called Describo and Selectrix. Describo's job will be to choose a card at random and describe it. Selectrix's job will be to select the correct card that Describo described. It would be easy for Describo to describe each card uniquely. For example, it could say "filled circle", "unfilled circle", and "unfilled square." But NACLO Labs wants the descriptions to be more efficient. They want the robots to follow something called Rational Speech Acts theory. This theory says that **a rational speaker, when deciding what to say, will consider both the cost of their utterance as well as the likelihood that they will be understood**. To make matters simple, we will say that the cost of an utterance will just be the number of words it contains: Saying nothing has a cost of zero; saying "filled", "unfilled", "square", or "circle" has a cost of one; and saying "filled circle", "unfilled circle", or "unfilled square" has a cost of two.

Now the question is how Describo can make the shortest utterance that Selectrix can understand. If Describo wants to refer to the filled circle, it could say "filled circle," but it could instead just say "filled" because there is only one filled shape. Saying "filled" costs less than saying "filled circle," and Selectro can still understand it correctly. Therefore, it would make sense for Describo to say "filled" instead of "filled circle."

G1. Here is a question to ponder: Suppose that, instead of machines, you and one other (rational) person had to communicate in this way, which is logical but not normal for humans. The other person is Describo and you are Selectrix, and the two of you are looking at the set of three cards shown above. The other person told you to choose the "circle." Which object does the other person want you to pick? Circle this object on your Answer Sheet. (Remember, the other person is using the Rational Speech Acts theory).

Now you want to apply Rational Speech Acts theory to a larger set of cards:



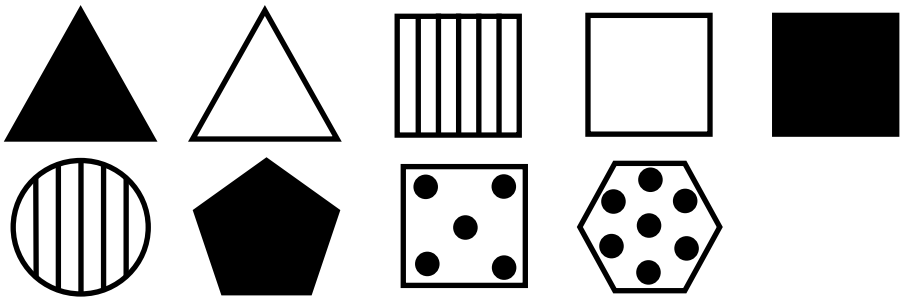
G2. Describe each shape in a way that has the smallest cost while remaining understandable. Your descriptions should come from the list below. You should provide each answer by writing a number from 1 to 11 next to each shape on your Answer Sheet. If multiple descriptions are tied for having the smallest cost while being understandable, you should only write one. If your descriptions were given to a (rational) listener, they should be able to understand exactly which card each description refers to—so make sure to strike the proper balance: you need to be understood, but you should also minimize the cost while doing so!

1. "filled" 2. "unfilled" 3. "striped" 4. "circle" 5. "triangle" 6. "square"
7. "filled triangle" 8. "filled circle" 9. "unfilled circle" 10. "unfilled square" 11. "striped square"

<u>Make sure you record your answers in your Answer Sheets!</u>

(G) At a Loss for Words (2/5)

Now, to push the Rational Speech Acts theory even further, you look at an even bigger set of cards:



G3. How would you describe each of these shapes with the shortest possible utterance that will still be understood? Record each description on your answer sheet by selecting a description from the list below and writing its number next to the shape it would describe. Note: A pentagon is a five-sided shape, and a hexagon is a six-sided shape. If multiple descriptions are tied for being the shortest one that will be understood, you should only write one.

1. "filled" 2. "unfilled" 3. "striped" 4. "spotted"
5. "triangle" 6. "square" 7. "circle" 8. "pentagon" 9. "hexagon"
10. "filled triangle" 11. "unfilled triangle" 12. "striped square"
13. "unfilled square" 14. "filled square" 15. "striped circle"
16. "filled pentagon" 17. "spotted square" 18. "spotted hexagon"

Now we fast-forward several years. Researchers at NACLO are looking through an abandoned lab when they find two robots, along with three envelopes containing cards. The robots had nametags saying "Descriptron" and "Identylizer", and the envelopes were labeled SMALL, MEDIUM, and LARGE. When they looked in the envelopes, they found that they contained the same cards that NACLO Labs had been working with! The workers at the abandoned lab must have been spying on NACLO to steal some of NACLO's research. This made the NACLO researchers curious: What if these researchers made more progress on machine-to-machine communication than NACLO had? NACLO Labs assigns you to investigate.

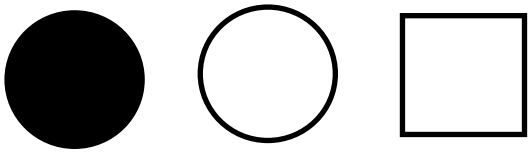
You experiment with Descriptron and Identylizer to find out what they are capable of. You observe that Descriptron does the following:

- It starts by choosing a card at random.
- Sometimes, it then says nothing.
- Other times, it names one property of the shape on the card it chose. For example, it might say "circle", "square", "filled", or "unfilled".
- Still other times, it names two properties of the shape it chose. For example, it might say "filled circle", "unfilled circle", or "unfilled square".
- Whether it names zero, one, or two properties is random.
- It never says anything wrong. For example, it never says "square" when describing a circle or "filled" when describing something unfilled.

Make sure you record your answers in your Answer Sheets!

(G) At a Loss for Words (3/5)

You then notice that Identylizer could respond to what Descriptron says by picking up cards. But it doesn't always pick the card that Descriptron was describing. In order to better understand what is going on, you decide to focus exclusively on the smallest set of cards—all of the rest of the questions will be about this set of cards, which we have repeated here:



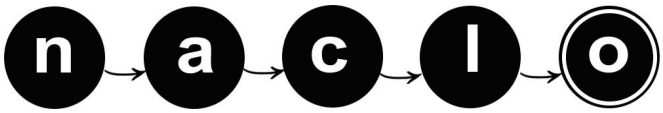
Using these three cards, you make many observations and record them in a table. Each row of the table represents what happens over multiple times when Descriptron chooses the card in the first column and says the thing in the second column. Remember that Descriptron does not always say the same thing in response to the same card. The third column shows Identylizer's error rate (the percentage of times it chooses the wrong card when Descriptron chose the card in column 1 and said the thing in column 2). To be certain that you are getting good observations, you make sure you have seen each pair of a card and a description many times.

For example, the table shows that every time Descriptron chooses the filled circle card and says "filled", Identylizer always picks up the card with a filled circle. Its error rate is zero. Every time Descriptron chooses unfilled circle and says nothing, Identylizer only picks the "correct" card (the unfilled circle) one third of the time. Its error rate is 0.67.

Intended card	**Descriptron's description**	**Identylizer's error rate**
Filled Circle	"Filled"	0.00
Filled Circle	"Circle"	0.50
Unfilled Square	"Unfilled Square"	0.00
Unfilled Circle	""	0.67
Unfilled Circle	"Unfilled Circle"	(a)
Unfilled Square	""	(b)
Unfilled Square	"Unfilled"	(c)

G4. Some cells in the table have been left blank. Use the examples that have been provided to figure out what strategy Identylizer uses to identify a card, and then fill in the missing cells in the table based on that strategy. Write your answers on the Answer Sheets. Each of your answers should be one of the following values: 0.00, 0.33, 0.50, 0.67, 1.00.

Make sure you record your answers in your Answer Sheets!



(G) At a Loss for Words (4/5)

You show your table of data to Descriptron. After it looks at the table, a button starts flashing. The button has the word "Loss" printed on it in bright red letters. You push the button. Now, when you show Descriptron a card, it whirs, then says zero, one, or two words as before, but it also says "weight", "error rate", "cost", and "loss" followed by some numbers. You notice that the values of these numbers can be different each time, so you record these outputs in the table below:

Intended Object	Description	Weight	Error Rate	Cost	Loss
Filled Circle	"Filled"	2	0.00	1	2.00
Filled Circle	"Circle"	3	0.50	1	3.50
Unfilled Square	"Unfilled Square"	0.2	0.00	2	0.40
Unfilled Circle	""	10	0.67	0	0.67

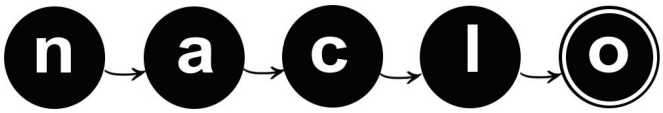
G5. Loss can be computed based on the other outputs. Describe how it is computed by filling in the blanks in the equation below. There are two correct ways to fill in the blanks; you should only pick one (it doesn't matter which one). Each blank should be filled with one of the following terms: **weight**, **error rate**, or **cost**.

Loss = _____(a)_____ + (_____(b)_____ × _____(c)_____)

You notice that Identylizer has a button on it labeled Rational Speech Acts. You push the button. Then you rerun your experiments from before (when you were computing Identylizer's error rate). But this time you get different results, as shown in the table below:

Intended Object	Description	Identylizer's Error Rate (in Rational Speech Acts mode)
Filled Circle	""	0.67
Filled Circle	"Filled"	0.00
Filled Circle	"Circle"	1.00
Filled Circle	"Filled Circle"	0.00
Unfilled Circle	""	0.67
Unfilled Circle	"Unfilled"	0.00
Unfilled Circle	"Circle"	0.00
Unfilled Circle	"Unfilled Circle"	0.00
Unfilled Square	""	0.67
Unfilled Square	"Unfilled"	1.00
Unfilled Square	"Square"	0.00
Unfilled Square	"Unfilled Square"	0.00

Make sure you record your answers in your Answer Sheets!



(G) At a Loss for Words (5/5)

Suddenly you hear another loud whirring, and then Descriptron says "I HAVE DETECTED THAT IDENTYLIZER IS USING RATIONAL SPEECH ACTS MODE. RECALIBRATING... RECALIBRATING…" Then, after a pause, Descriptron says, "PLEASE SELECT A WEIGHT." Seven buttons start flashing, labeled 0.1, 0.5, 0.7, 1, 5, 10, and 100. You push the button labeled 1. Descriptron then prints out an index card with the table below on it:

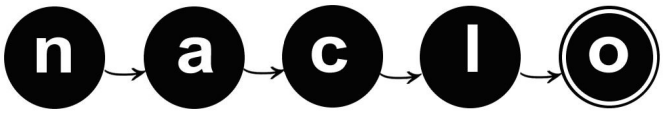
Intended Object	Description	Weight	Error Rate	Cost	Loss
Filled Circle	""	1	0.67	0	0.67
Filled Circle	"Filled"	1	0.00	1	1.00
Filled Circle	"Circle"	1	1.00	1	2.00
Filled Circle	"Filled Circle"	1	0.00	2	2.00
Unfilled Circle	""	1	0.67	0	0.67
Unfilled Circle	"Unfilled"	1	0.00	1	1.00
Unfilled Circle	"Circle"	1	0.00	1	1.00
Unfilled Circle	"Unfilled Circle"	1	0.00	2	2.00
Unfilled Square	""	1	0.67	0	0.67
Unfilled Square	"Unfilled"	1	1.00	1	2.00
Unfilled Square	"Square"	1	0.00	1	1.00
Unfilled Square	"Unfilled Square"	1	0.00	2	2.00

After printing the table, Descriptron says "SWITCHING TO BEST-DESCRIPTION MODE." To figure out what that means, you show Descriptron the card with the unfilled square. Descriptron says, "THE BEST DESCRIPTION FOR THIS ITEM IS: <SAY NOTHING>" After experimenting some more, you figure out that the way Descriptron selects the best description is by finding the description that has the lowest loss. If there is a tie for the lowest loss, Descriptron chooses randomly between the tied options. The reason why it said "<SAY NOTHING>" for the unfilled square is that, in the table above, the option of saying nothing (written as "") has a loss of 0.67, which is lower than the loss for the other options (2.00, 1.00, and 2.00).

G6. Currently, with the weight set to 1, Description says that "<SAY NOTHING>" is the best description for all three objects in the table above. However, by changing the weight, it is possible to get Descriptron to produce a one-word description for all 3 shapes. On your Answer Sheets, circle **all** of the following weight values that would accomplish this. You can circle zero, one, or more than one values.

a. 0.1 b. 0.5 c. 0.7 d. 5 e. 10 f. 100

Make sure you record your answers in your Answer Sheets!



(H) Evidentially, My Dear Tariana (1/3) [15 Points]

Tariana is an endangered language spoken in the Vaupés River area in Brazil, close to the border with Colombia. There are about 100 speakers of Tariana today, while there are a further 1,500 Tariana people who do not speak the language.

One morning, Antônia, Cândido, Margarida, João, Maria, and Emílio were playing a detective game, in which they had to discover who had just "poisoned" their sleeping friend Ismael! They very quickly decided that he had drunk some poisoned *manioc-flour*, a mixture of cassava (or tapioca) flour and water. When asked what happened to Ismael, their responses were as follows:

Speaker	Tariana Statement	English Translation	Evidence for Claim
Antônia	**Diha kawhi-nuku diiraka**	*He drank manioc-flour*	Antônia saw Ismael drink it before he (Ismael) fell asleep.
Cândido	**Mayẽ kawhi-nuku diiranihka**	*Ismael drank manioc-flour*	Cândido saw some manioc-flour on Ismael's upper lip (like a "milk mustache").
Margarida	**Diha kawhi-nuku diiramahka**	*He drank manioc-flour*	Margarida had heard Ismael slurping his manioc-flour in the room next door.
João	**Mayẽ kawhi-nuku diirasika**	*Ismael drank manioc-flour*	João knew that Ismael took some of his sister's manioc-flour every morning.
Maria	**Diha kawhi-nuku diirapidaka**	*He drank manioc-flour*	Ismael told Maria (before falling asleep) that he (Ismael) had drunk some.
Emílio	**Mayẽ kawhi-nuku diiramahka**	*Ismael drank manioc-flour*	Emílio could smell manioc-flour on Ismael's breath.

First, they suspected José, so they went to his house. They investigated the house and spoke to a key witness (José's mom). They all concluded José was innocent—he had an alibi! When asked what José had been doing, their responses were as follows:

Speaker	Tariana Statement	English Translation	Evidence for Claim
Antônia	**Juse irida dimanikaka**	*José was playing soccer*	Antônia saw José playing soccer on the field by the road.
Cândido	**Juse irida dimanikanihka**	*José was playing soccer*	Cândido saw that both José and José's soccer equipment were missing.
Margarida	**Juse irida dimanikasika**	*José was playing soccer*	Margarida knew it was time for José's weekly soccer match.
João	**Naha irida namanikaka**	*They were playing soccer*	João could see José playing soccer with his brother.
Maria	**Naha irida namanikapidaka**	*They were playing soccer*	Maria was told by José's mum that José was playing soccer with his brother.
Emílio	**Naha irida namanikanihka**	*They were playing soccer*	Emílio saw that José's soccer ball and both sets of soccer shoes were missing.

(H) Evidentially, My Dear Tariana (2/3)

Next, they went to Ismael's house. Again they looked around, and they spoke to a second key witness (Ismael's sister, Olívia). They agreed she was acting very suspiciously, as she wouldn't say what she was doing at the time of Ismael's "death". This time, they didn't all agree. When asked what they thought had happened, their responses were as follows:

Speaker	Tariana Statement	English Translation	Evidence for Claim
Antônia	**Óli dukwasika**	*Olívia was sleeping in the hammock*	Antônia believed that Olívia always sleeps all morning.
Cândido	(a)	*She was swimming*	Cândido saw that Olívia's clothes were wet.
Margarida	**Duha duayhamahka**	*She was swimming*	Margarida had heard Olívia splashing around in the river earlier.
João	**Óli kawhi-nuku dunimahka**	*Olívia was making manioc-flour*	João could smell manioc-flour in the kitchen.
Maria	(b)	*She was making manioc-flour*	Maria was told by Ismael's dad that Olívia had been making manioc-flour.
Emílio	**Duha hinipuku dumeyutanihka**	*She was raking the garden*	Emílio saw that the garden had recently been raked.

When João and Maria shared their information, they told Olívia:

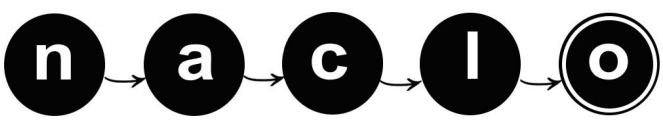
"___(c)___ mayẽ piinunihka!" *You killed Ismael!*

Answer the following questions on your Answer Sheets:

H1. Fill in the gaps (a) and (b) by giving Maria and Emílio's statements in Tariana.

H2. What Tariana word should replace (c)?

<u>Make sure you record your answers in your Answer Sheets!</u>



(H) Evidentially, My Dear Tariana (3/3)

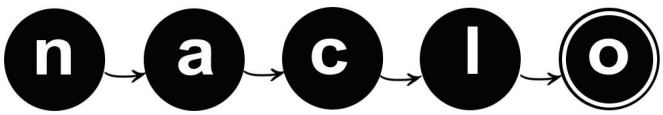
They began explaining all of the things they had come across that morning. They establish that Olívia had also gone swimming, and that Leonardo had raked the garden before going to lie down in the hammock.

H3. For each statement below (numbered 1 through 4), three possible pieces of evidence are given (labelled A, B, and C). On your Answer Sheet, circle which one(s) the speaker may have had in mind. For each question, you may circle zero, one, or more than one options.

	Speaker	Tariana Statement	English Translation	Possible Evidence for Claim
1)	Leonardo	**Óli duayhaka**	*Olívia was swimming*	A. Leonardo knew that Olívia liked swimming. B. Leonardo had seen Olívia swimming in the river. C. Leonardo saw that Olívia's clothes were wet.
2)	Maria	**Leó hinipuku dimeyutasika**	*Leonardo was raking the garden*	A. Maria knew that Leonardo raked the garden every morning. B. Leonardo smelt of the garden. C. Maria saw that someone had raked the garden.
3)	João	**Diha hinipuku dimeyutamahka**	*He was raking the garden*	A. Leonardo smelt of the garden. B. João had heard Leonardo raking the garden. C. João saw that someone had raked the garden.
4)	Emílio	**Leó dikwanihka**	*Leonardo was sleeping in the hammock*	A. Leonardo told Emílio that he (Leonardo) had been asleep. B. Emílio saw that someone had been sleeping in the hammock. C. Leonardo looked well-rested.

H4. On your Answer Sheets, fill in the gaps (d)-(h) with the correct Tariana translation.

Speaker	Tariana Statement	English Translation	Evidence for Claim
Cândido	(d)	*You were swimming*	Cândido had overheard Olívia confirm that she (Olívia) had gone swimming.
Margarida	(e)	*He was raking the garden*	Margarida saw that Leonardo's clothes had bits of grass and leaves on them.
Olívia	(f)	*Leonardo was sleeping in the hammock*	Olívia knew that Leonardo always lies down in the hammock after gardening.
Antônia	(g)	*You drank manioc-flour*	Antônia saw a dirty cup next to the hammock where Leonardo had been sleeping.
Olívia	(h)	*They were playing soccer*	Olívia had heard a soccer game going on while she was swimming.



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 (1/6)

(A) Holy Roman Shorthand

A1. Write a number from 1 to 6 in each box to fill in the blanks in the table:

(a) ☐ (b) ☐ (c) ☐ (d) ☐ (e) ☐ (f) ☐

A2. Write the letter of the abbreviation that does not match the others: ☐

A3. Write the letter for the correct symbol: ☐

A4. In the boxes next to each Latin word, write the numbers that stand for the relevant symbols from the symbol bank. You might not use all of the boxes.

a. *procatis* ☐☐☐☐☐☐☐☐

b. *notarum* ☐☐☐☐☐☐☐☐

c. *decernimus* ☐☐☐☐☐☐☐☐

d. *congnitio* ☐☐☐☐☐☐☐☐

e. *crux* ☐☐☐☐☐☐☐☐

f. *petitur* ☐☐☐☐☐☐☐☐

A5. What is the English word? Write this answer in letters, not numbers. ☐

(B) Finding Your Place in Warlpiri

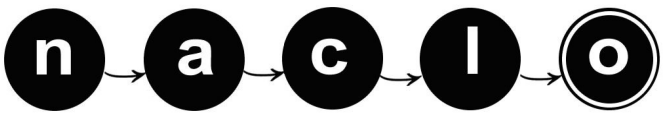
B1. Fill in the blank spaces in the table.

	Warlpiri	**English**
6.	Maliki ka nantuwurla ngunami.	**a.**
7.	**b.**	The kangaroo is sitting in the school.

B2. Circle the two words that have the same meaning:

karnta *luurnpa* *mardukuja* *ngunami* *pirli*

Problem (B) continues on the next page.



YOUR NAME: REGISTRATION #

Answer Sheets (2/6)

(B) Finding Your Place in Warlpiri (continued)

B3. Write A or B to answer each question from i. to vi.:

i. ☐ iv. ☐

ii. ☐ v. ☐

iii. ☐ vi. ☐

(C) Happy Birthday!

C1. Write M or F in each box to indicate the typical gender associated with each name.

a. ☐ b. ☐ c. ☐ d. ☐ e. ☐

C2. Write the letter of the name that could go with each birth date.

i. ☐ ii. ☐ iii. ☐ iv. ☐ v. ☐ vi. ☐

C3. Write the letter for the day on which a child named Kofi could have been born: ☐

C4. Write the letter for the name that might be given to a child born on February 5: ☐

C5. Translate each Ewe word into English.

i. Dzoɖagbe ☐ ii. Fiɖagbe ☐

(D) Generally Speaking

D1. Fill in the empty cells (the ones labeled **a.** through **n.**).

	Sentence	C1	C2	C3	C4	Human
31.	The tall detective saw the ballerina.	U	**a.**	G	G	G
32.	The linguist saw the yodeler.	U	**b.**	G	G	G
33.	The visited met linguist programmer.	U	**c.**	U	G	U
34.	The yodeler met the woodcarver.	U	**d.**	G	G	G
35.	The watchmaker is famous.	U	**e.**	G	G	G
36.	The detective met the yodeler.	**f.**	**g.**	G	**h.**	G
37.	The spy is talented.	**i.**	**j.**	G	**k.**	G
38.	The programmer met the cartoonist.	**l.**	**m.**	G	**n.**	G

YOUR NAME: REGISTRATION #

Answer Sheets (3/6)

(D) Generally Speaking (continued)

D2. Write what each HIDDEN_WORD stands for.

HIDDEN_WORD_1

HIDDEN_WORD_2

HIDDEN_WORD_3

D3. Next to each computer's name below, write how many sentences there were for which that computer produced the same label as you did. Each answer should be a number from 1 to 6. This question will only be graded for completion—as long as you write an answer, you will receive credit no matter what the answer is.

C2 C3

D4. Under each encoded word, write the word that it stands for. You should write one letter in each box. You do not need to write the hyphen in the word that contains a hyphen.

Q G T R H U U

R K G K H E N D K E W

U Q Y E T K F Q - D K A P E

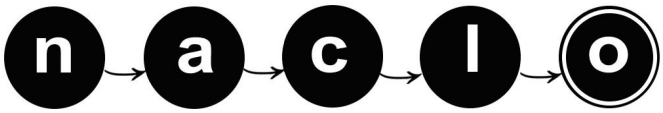
E P X

H E E Q R M E K X G

T X R Q C P Q G

W Q T E Q D C Q Q B

E H U U J X R K G A T



YOUR NAME: REGISTRATION #

Answer Sheets (4/6)

(E) Japanese From Afar

E1. Under each katakana character, write the letter of the Japanese Flag Semaphore example that corresponds to it.

i) テ ***te***	ii) ヨ ***yo***	iii) ア ***a***	iv) ミ ***mi***	v) モ ***mo***	vi) ヌ ***nu***	vii) セ ***se***

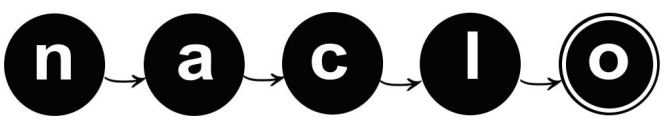
E2. Circle the example that corresponds to the Japanese Flag Semaphore letter that was not used in **E1**.

メ ***me*** 井 ***wi*** ク ***ku*** へ ***he*** ホ ***ho*** シ ***si***

(F) Paiwan in Progress

F1. Fill in the table entries that have been left blank.

Root Forms		**Progressive Forms**	
temulu	*to teach*	**a.**	*teaching*
djemavac	*to walk*	**b.**	*walking*
lemenguaq	*to fix*	**c.**	*fixing*
mikerekel	*to shake*	**d.**	*shaking*
takalava	*to wait*	**e.**	*waiting*
masasevalit	*to take a turn*	**f.**	*taking turns*

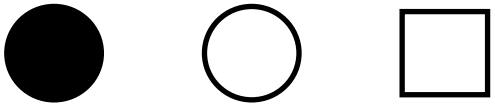


YOUR NAME: REGISTRATION #

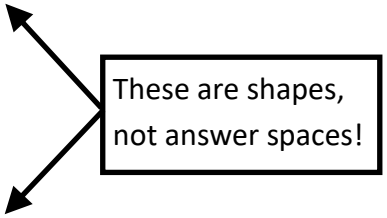
Answer Sheets (5/6)

(G) At a Loss for Words

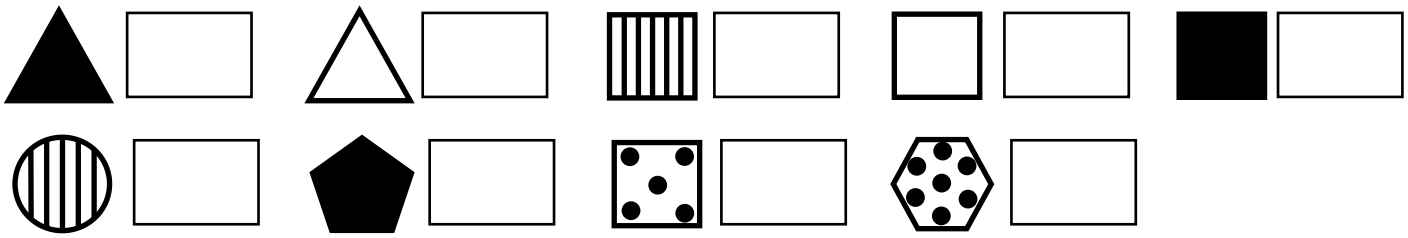
G1. Circle the object that the other person wants you to pick:



G2. Next to each shape, write the number for a description of it that has the smallest cost while still being understandable.



G3. Next to each shape, write the number for a description of it that has the smallest cost while still being understandable.



G4. Fill in spaces **a**, **b**, and **c** in the table. Each value should be 0.00, 0.33, 0.50, 0.67, or 1.00.

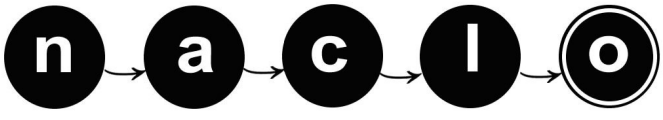
Intended card	Descriptron's description	Identylizer's error rate
Unfilled Circle	"Unfilled Circle"	**a.**
Unfilled Square	""	**b.**
Unfilled Square	"Unfilled"	**c.**

G5. Fill in the blanks from the equation.

a. _____ **b.** _____ **c.** _____

G6. Circle **all** weight values that ensure that Descriptron will produce a one-word description for all three shapes. You can circle zero, one, or more than one answer.

a. 0.1 b. 0.5 c. 0.7 d. 5 e. 10 f. 100



YOUR NAME: REGISTRATION #

Answer Sheets (6/6)

(H) Evidentially, My Dear Tariana

H1. Fill in the empty cells in the table.

Speaker	Tariana Statement	English Translation	Evidence for Claim
Cândido	**a.**	*She was swimming*	Cândido saw that Olívia's clothes were wet.
Maria	**b.**	*She was making manioc-flour*	Maria was told by Ismael's dad that Olívia had been making manioc-flour.

H2. Write the Tariana word that goes in space **c**.

H3. For each number, circle the letter(s) corresponding to the evidence that the speaker may have had in mind. In each case, you may circle zero, one, or more than one option.

1) A. B. C. 2) A. B. C. 3) A. B. C. 4) A. B. C.

H4. Fill in the gaps in the table.

Speaker	Tariana Statement	English Translation	Evidence for Claim
Cândido	**d.**	*You were swimming*	Cândido had overheard Olívia confirm that she (Olívia) had gone swimming.
Margarida	**e.**	*He was raking the garden*	Margarida saw that Leonardo's clothes had bits of grass and leaves on them.
Olívia	**f.**	*Leonardo was sleeping in the hammock*	Olívia knew that Leonardo always lies down in the hammock after gardening.
Antônia	**g.**	*You drank manioc-flour*	Antônia saw a dirty cup next to the hammock where Leonardo had been sleeping.
Olívia	**h.**	*They were playing soccer*	Olívia had heard a soccer game going on while she was swimming.