n a c l o

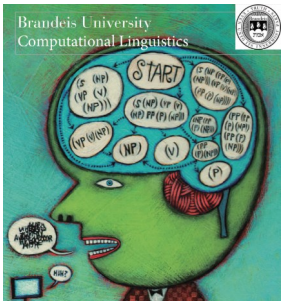
Linguistic Society of America

NAACL

Carnegie Mellon University
Language Technologies Institute

Massachusetts Institute of Technology

MIDDLE TENNESSEE STATE UNIVERSITY



NATIONAL SCIENCE FOUNDATION

Yale University
LUX ET VERITAS

The Fourteenth Annual

North American Computational Linguistics Open Competition 2020

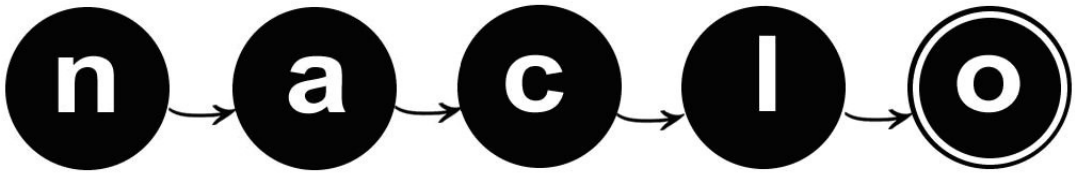
www.nacloweb.org

Invitational Round

March 5, 2020

"Serious language puzzles that are surprisingly fun!"

-Will Shortz, Crossword editor of the *New York Times* and Puzzlemaster for NPR



Welcome to the fourteenth annual North American Computational Linguistics Open Competition! You are among the few, the brave, and the brilliant 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 four hours long and includes ten problems, labeled I to R.
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 the Invitational Round, some questions require explanations.** Please read the wording of the questions carefully.
6. All your answers should be 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 students from each country (USA and Canada) will be invited to the next round, which involves team practices before the international competition in Latvia.
9. Each problem has been thoroughly checked by linguists and computer scientists as well as students like you for clarify, accuracy, and solvability. 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. If we have done our job well, very few people will solve all these problems completely in the time allotted. So, don't be discouraged if you don't finish everything.
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 2020 Organizers

Program Committee:

Adam Hesterberg — Massachusetts Institute of Technology
Aleka Blackwell — Middle Tennessee State University
Ali Sharman — University of Michigan
Andrés Salanova — University of Ottawa
Annie Zhu — Harvard University
Babette Verhoeven — Aquinas College
Ben LaFond — Harvard University
Corinne Soucy — Université de Montréal
Daniel Lovsted — McGill University
David Mortensen — Carnegie Mellon University
Dick Hudson — University College London
Dragomir Radev — Yale University
Elisabeth Mayer — Australian National University
Ethan Chi — Stanford University
Harold Somers — All Ireland Linguistics Olympiad
Heather Newell — UQAM
James Pustejovsky — Brandeis University
James Hyett — University of Toronto
Ji Hun Wang — Stanford University
Kevin Liang — University of Pennsylvania
Lori Levin — Carnegie Mellon University
Margarita Misirpashayeva — Massachusetts Institute of Technology
Mary Laughren — University of Queensland
Matthew Gardner — Carnegie Mellon University
Mihir Singhal — Massachusetts Institute of Technology
Oliver Sayeed — University of Pennsylvania
Patrick Littell — National Research Council of Canada
Pranav Krishna — Massachusetts Institute of Technology
Shuli Jones — Massachusetts Institute of Technology
Sonia Reilly — Massachusetts Institute of Technology
Tom McCoy — Johns Hopkins University
Tom Roberts — University of California, Santa Cruz

NACLO 2020 Organizers

Problem Credits:

Problem I: Ethan Chi
Problem J: Sam Ahmed
Problem K: Ethan Chi
Problem L: Ben LaFond
Problem M: Ethan Chi
Problem N: Daniel Lovsted
Problem O: Harold Somers
Problem P: Ethan Chi
Problem Q: Lori Levin and Tom McCoy
Problem R: Daniel Lovsted

Organizing Committee:

Adam Hesterberg — Massachusetts Institute of Technology
Aleka Blackwell — Middle Tennessee State University
Ali Sharman — University of Michigan
Andrés Salanova — University of Ottawa
Annie Zhu — Harvard University
Ben LaFond — Harvard University
Brian Xiao — Massachusetts Institute of Technology
Corinne Soucy — Université de Montréal
Daniel Lovsted — McGill University
David Mortensen — Carnegie Mellon University
Dragomir Radev — Yale University
Ethan Chi — Stanford University
James Pustejovsky — Brandeis University
Ji Hun Wang — Stanford University
Kevin Liang — University of Pennsylvania
Lori Levin — Carnegie Mellon University
Margarita Misirpashayeva — Massachusetts Institute of Technology
Matthew Gardner — Carnegie Mellon University
Mihir Singhal — Massachusetts Institute of Technology
Patrick Littell — National Research Council of Canada
Pranav Krishna — Massachusetts Institute of Technology
Ryan Guan — Stanford University
Shuli Jones — Massachusetts Institute of Technology
Sonia Reilly — Massachusetts Institute of Technology
Tom McCoy — Johns Hopkins University
Tom Roberts — University of California, Santa Cruz

NACLO 2020 Organizers (cont'd)

US Team Coaches:

Aleka Blackwell, Middle Tennessee State University
Dragomir Radev, Yale University
Lori Levin, Carnegie Mellon University
Patrick Littell, National Research Council of Canada

Canadian Coordinators and Team Coaches:

Daniel Lovsted, McGill University

USA Contest Site Coordinators:

Brandeis University: James Pustejovsky, Sarah Irwin, Nikhil Krishnaswamy
Brigham Young University: Deryle Lonsdale
California State University: Michael Ahland
Carnegie Mellon University: Mary Jo Bensasi, Lori Levin
College of William and Mary: Dan Parker
Columbia University: Kathy McKeown, Brianne Cortese, Smaranda Muresan
Cornell University: Abby Cohn, Sam Tilsen
Emory University: Jinho Choi, Phillip Wolff
Georgetown University: Emma Manning
Indiana University: Sandra Kuebler
Johns Hopkins University: Rebecca Knowles, Pamela Shapiro, Tom McCoy
Kansas State University: Natasha Rojkovskaia
Keiser University, Jacksonville: Deborah Williams
Long Island University Post: Rachel Szekely
Massachusetts Institute of Technology: Margarita Misirpashayeva, Sonia Reilly
Middle Tennessee State University: Aleka Blackwell
Minnesota State University Mankato: Rebecca Bates, Dean Kelley, Louise Chan
Montclair State University: Anna Feldman
Northeastern Illinois University: Ariana Bancu, Lewis Gebhardt
Northwestern University: Alaina Arthurs
Ohio State University: Michael White, Micha Elsner, Marie de Marneffe
Oregon State University: Liang Huang
Princeton University: Christiane Fellbaum, Eve Fleisig, Rishi Mago, Abinitha Gourabathina

NACLO 2020 Organizers (cont'd)

USA Contest Site Coordinators (cont'd):

San Diego State University: Rob Malouf

Southern Illinois University: Vicki Carstens, Jeffrey Punske

Stanford University: Daisy Leigh, Eli Strauss-Reis, Matthew Early

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

UCLA: Sue Han, Lily Kawaoto

Union College: Kristina Striegnitz, Nick Webb

University of California, Irvine: Sameer Singh, Zhengli Zhao, Kristen Salsbury

University of Colorado at Boulder: Silva Chang

University of Houston: Thamar Solorio, Deepti Bhat, Giovanni Molina

University of Illinois at Urbana-Champaign: Benjamin Leff, Greg Sigalov, Julia Hockenmaier

University of Kentucky: Kevin McGowan, Andrew Byrd

University of Maryland: Kasia Hitczenko, Laurel Perkins, Aaron Doliana, Sigwan Thivierge, Nadiya Klymenko

University of Massachusetts, Lowell: Anna Rumshisky, David Donahue, Willie Boag

University of Massachusetts: Andrew Lamont, Anastasia Chobany

University of Memphis: Vasile Rus

University of Michigan: Steven Abney, Sally Thomason, Marcus Berger

University of Nebraska, Omaha: Parvathi Chundi

University of North Carolina Charlotte: Wlodek Zadrozny, Hossein Hematialam, Kodzo Wegba

University of North Texas: Rodney Nielsen

University of Notre Dame: David Chiang

University of Pennsylvania: Chris Callison-Burch, Mitch Marcus, Cheryl Hickey, Oliver Sayeed, Marianna Apidianaki, Derry Wijaya, Anne Cocos

University of Southern California, ISI campus: Jon May, Nima Pourdamghani

University of Texas at Dallas: Vincent Ng, Jing Lu, Gerardo Ocampo Diaz

University of Utah: Brendan Terry, Jessica Larsen, Justin Nistler, Joselyn Rodriguez

University of Washington: Jim Hoard, Joyce Parvi

University of Wisconsin, Milwaukee: Joyce Boyland, Hanyon Park, Gabriella Pinter, Anne Pycha

Western Washington University: Kristin Denham

Yale University: Raffaella Zanuttini

NACLO 2020 Organizers (cont'd)

Canada Contest Site Coordinators:

Dalhousie University: Vlado Keselj, Dijana Kosmajac

Simon Fraser University: Ashley Farris-Trimble

University of British Columbia: Yadong Liu, Jozina Vander Klok

University of Ottawa: Andrés Pablo Salanova

University of Toronto: Minh-Tam Nguyen, Lola Bradford

University of Western Ontario: Janis Chang

University of Calgary: Dennis Storoshenko

University of Alberta: Herbert Colston, Sally Rice

McGill University: Lisa Travis

NACLO 2020 Organizers (cont'd)

Booklet Editors:

Ji Hun Wang, Stanford University
Ethan Chi, Stanford University
Tom McCoy, Johns Hopkins University
Dragomir Radev, Yale University
Shuli Jones, Massachusetts Institute of Technology

Sponsorship Chair:

James Pustejovsky, Brandeis University

Sustaining Donors:

Linguistic Society of America
North American Chapter of the Association for Computational Linguistics
National Science Foundation

University Donors:

Brandeis University
Carnegie Mellon University
Massachusetts Institute of Technology
Middle Tennessee State University
Yale University

Many generous individual donors

Special thanks to:

The hosts of the 200+ High School Sites

(I) Kãa, Kai, and Khai (1/1) [10 Points]

Thai is a Kra-Dai language spoken by over 36 million people in Thailand and Cambodia. Below are some phrases in Thai, with their English translations in a different order. Usually, Thai is written in the Thai script: for example, the word *ngãam*, which means 'beautiful,' is written งาม. However, here, all of the words have been written in Roman script for you. Note that diacritic marks (˜, ˋ, ´, ˆ) over vowels represent tones.

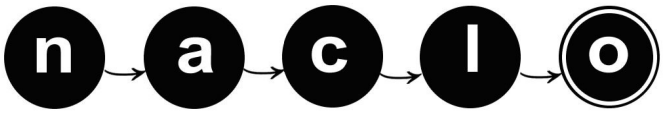
1.	*at*	A.	carbonated drink
2.	*chiãng mai*	B.	Chiang Mai[1]
3.	*dòm*	C.	chicken egg
4.	*kãa dòm*	D.	coffee pot
5.	*kãa kãafãae*	E.	cold air
6.	*khàao mãn kai*	F.	egg
7.	*khai*	G.	Hainan chicken rice[2]
8.	*khai dòm kháeng*	H.	hard-boiled egg
9.	*khai kai*	I.	ice
10.	*lõm nâao*	J.	kettle (for boiling water)
11.	*mâai at*	K.	line
12.	*mai*	L.	line of trees
13.	*mãn*	M.	new
14.	*nãeo*	N.	new price
15.	*nãeo mâai*	O.	oil, grease
16.	*nâm*	P.	plywood[3]
17.	*nâm at lõm*	Q.	the price of gasoline
18.	*nâm kháeng*	R.	to boil
19.	*rãakhãa mai*	S.	to compress
20.	*rãakhãa nâm mãn*	T.	water

I1. Determine the correct correspondences. Write your answers in the Answer Sheets.

I2. Translate the following into English: *lõm*; *kãa*; *chiãng*. Write your answers in the Answer Sheets.

I3. Translate the following into Thai: beautiful line. Write your answers in the Answer Sheets.

[1] Chiang Mai (literally "New City") is a city in Northern Thailand.
[2] Hainan chicken rice is a popular Southeast Asian dish, made of rice with chicken fat.
[3] Plywood is a composite material made by stacking together thin slices of wood.



(J) You Made Me See (1/1) [15 Points]

Iñapari is a critically endangered language of the Arawakan family, spoken by only 4 people in the town of the same name on the border of Peru, Bolivia and Brazil. Below are some verbs in Iñapari with their translations into English. Note that (sg., pl., m., f.) after pronouns represent the singular, plural, masculine, and feminine versions of those pronouns, respectively.

	Iñapari	**English**
1.	*awunahamanahari*	We saw him again.
2.	*ipinimarona isa*	You (pl.) healed them (f.).
3.	*nechaparapirâî isa*	I want to visit you (pl.).
4.	*aarunahapiramanahanona*	They (f.) do not want to see me again.
5.	*ijamapirarona*	He wants to hear them (f).
6.	*aanynapomaro*	I did not find her.
7.	*ajamachajimanahai*	We made you (sg.) hear again.
8.	*rynapomâî isa*	He found you (pl.).
9.	*rechaparamanaharina*	They (m.) visited him again.
10.	*rupinichajimanahâî*	She made us heal again.
11.	*unahachajimano isa*	You (pl.) made me see.
12.	*punahamanaharo*	You (sg.) see her again.
13.	*aanopinipirai*	I do not want to heal you (sg.)
14.	*rupinimanahanana*	They (f.) healed them (m.) again.

J1. Translate the following Iñapari verbs into English. Write your answers in the Answer Sheets.

(i) *pijamamâî*

(ii) *aaipinipiramanahanana*

(iii) *aaawechaparachajimâî isa*

J2. Translate the following English phrases into Iñapari. Write your answers in the Answer Sheets.

(i) She wants to visit you (pl.) again.

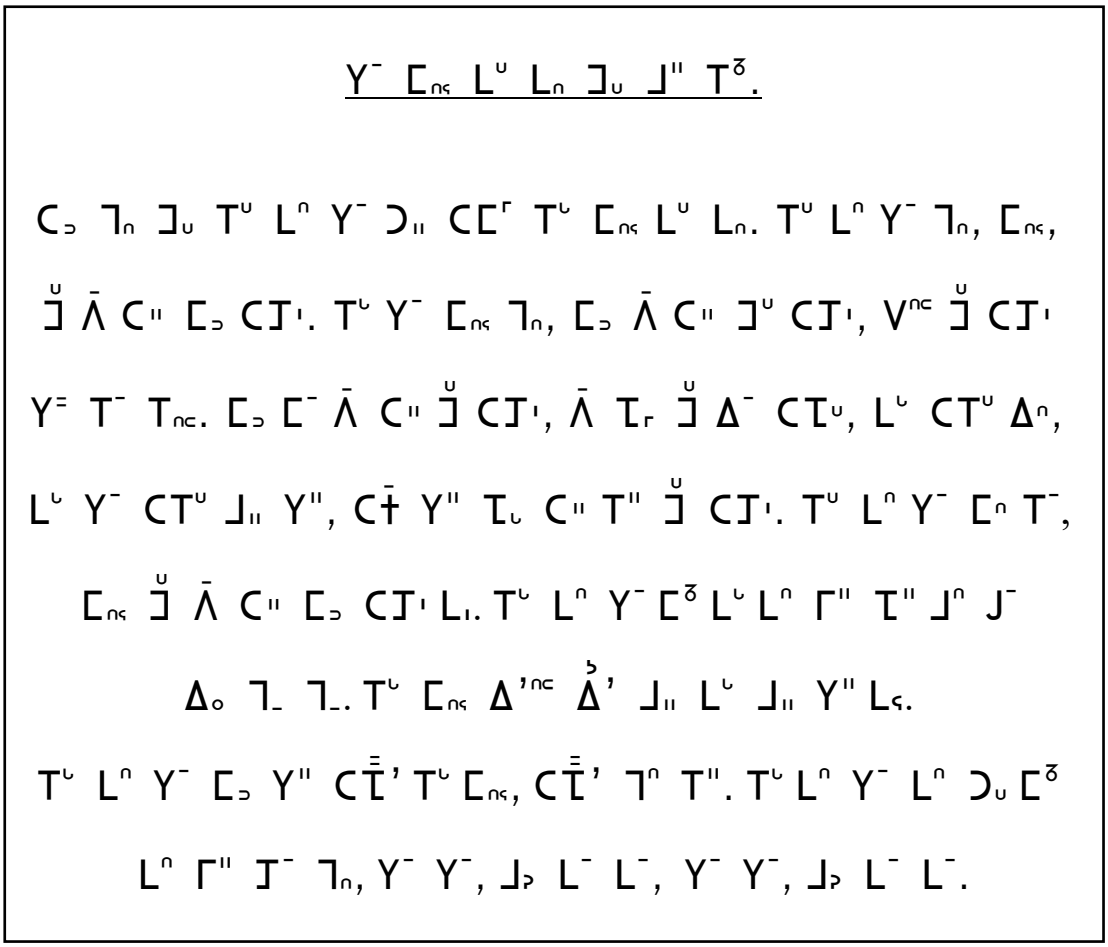
(ii) I made him hear.

(iii) You (sg.) did not heal them (f.).

J3. Explain your solution. Write your explanation in the Answer Sheets.

(K) The Pimpled Toad (1/2) [10 Points]

Below is a famous Miao folk tale, “The Pimpled Toad who was Wise,” written in the Large Flowery Miao language, which is spoken by about 300,000 people in Weining Autonomous County, Guizhou, China. Large Flowery Miao is generally written in two different scripts: the *Pollard* script, created in 1904 by the English missionary Samuel Pollard, as well as Miao Pinyin, which uses Roman script. Below is the Pollard script version.



On the next page is an English translation of the “The Pimpled Toad.” Of course, not every word in the English translation has an equivalent word in Miao; on the other hand, some English words may be translated by more than one Miao word. Also, note that constructions of the form “I want to eat you / you want to eat me” are written in Miao as “I want to eat your flesh / you want to eat my flesh.”

(K) The Pimpled Toad (2/2)

The Pimpled Toad that was Wise.

They say that a crow went to meet a pimpled toad. The crow said, "Toad, I want to eat you." The toad said, "You want to eat me, but my flesh is very bitter. If you want to eat me, you want to carry me over there, to the stream, to the bank of the pool of water. Wash me in water, then eat to get my flesh." The crow replied, "Toad, I want to eat you, so there!" The crow stretched his head upwards, laughing *ha-ha*. The toad jumped, hopped down to the pool water, and was gone. The crow waded in the water to seek the toad, sought him, but did not get him. The crow simply stretched his head, cawed, and said, "Ah-ah, all for nothing! Ah-ah, all for nothing!"

Finally, here is the Miao Pinyin transcription. However, there are some changes:

- The sentences, including the title, are in random order.
- Five of the sentences have been divided into two parts.
- Punctuation has been removed.
- The numbered blanks (*e.g.* **18**) mark places where one or more words are missing.

A. *tái* **1** *jỳ áo ⁿtrīeh tái jìeu ⁿtrīeh hì táo*
B. *jỳ* **2** *yā não gū ⁿghãi yā drìh gū dlá ⁿdrũ*
C. *tú* **3** *jĩ-tá jìeu*
D. *gū yā não jỳ ⁿghãi* **4**
E. *tú lí-á* **5** *jìeu*
F. *gū yā não jỳ ⁿghãi*
G. *nỳ hì gù tú lí-á* **6** *ⁿjíh tái jìeu lú-lì*
H. *ⁿzā* **7** *drài* **8** *táo gū ⁿghãi*
I. *á-jìeu lú-lì* **9** *báo-táng*
J. **10** *bí-xá dlò hà-hà*
K. *tái jìeu tlíe tlw̄ bào lái* **11** *áo lèu*
L. *tái lí-á lí-mù* **12** *lí-fáo ghá hì*
M. *á á* **13** *bè-lá-lá*
N. *tái lí-á* **14** *lái lí-fáo*
O. *tái á-jìeu hì jỳ* **15** *não gú* **16**
P. *víe gū ⁿghãi íeh tá-tìe*
Q. *lái* **17** *lái á-ⁿtú bào áo*

Note that the diacritics (accent marks) ´, `, ¯, ˜ represent tones; *ieh* and *w* are vowels/vowel sequences; *gh*, *dl*, *tl*, *dr*, and *tr* are consonants; ⁿ marks nasalization of the following consonant, that is, air flow escapes through the nose and the mouth simultaneously during the production of the consonantal sound.

K1. Restore the missing blanks, and translate each of the missing sections into English. Write your answers in the Answer Sheets.

K2. Write the following line from another Miao story in Pollard script. If there are multiple ways to write a word, any choice is acceptable. Write your answers in the Answer Sheets.

nỳ hì, á bõ lào jỳ yā xão jỳ tá ghái ⁿdrìh.

('They said, "Old woman, you must look after your chicken carefully."')

K3. Describe your observations on the structure of Pollard script and Large Flowery Miao grammar. Write your answers in the Answer Sheets.

(L) Shiva Sutras (1/3) [5 Points]

The following 14-line poem is one of the four sections of a grammar of Sanskrit, an ancient Indian language, written by the 4th-century BCE Indian grammarian Pāṇini. It is called the *Akṣarasamāmnāya* or *Śivasūtras*, and it functions as an ordering of the sounds of the Sanskrit language[1] – like the English "A, B, C…" with some special properties.

1.	*a*	*i*	*u*			*Ṇ*
2.				*ṛ*	*ḷ*	*K*
3.		*e*	*o*			*Ñ*
4.		*ai*	*au*			*C*
5.	*h*	*y*	*v*	*r*		*Ṭ*
6.					*l*	*Ṇ*
7.	*ñ*	*m*	*n̄*	*ṇ*	*n*	*M*
8.	*jh*	*bh*				*Ñ*
9.			*gh*	*ḍh*	*dh*	*Ṣ*
10.	*j*	*b*	*g*	*ḍ*	*d*	*Ś*
11.	*kh*	*ph*	*ch*	*ṭh*	*th*	
			c	*ṭ*	*t*	*V*
12.	*k*	*p*				*Y*
13.		*ś*	*ṣ*	*s*		*R*
14.	*h*					*L*

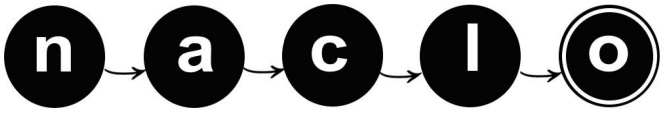
NOTE: ṛ and ḷ are vowels; *ñ, n̄, ṇ, ḍ, ṭ, ś*, and *ṣ* are consonants. A consonant with a letter *h* after it (e.g. *jh*) is considered a separate 'sound' from the consonant without the *h* (e.g. *j*). The vowels *a i u* each have a long counterpart, *ā ī ū*, which for purposes of the *Śivasūtras* is considered equivalent with the short form.

The organization of the *Śivasūtras* allows us to give names to certain groups of sounds. For example, the single syllable *aC* refers to the vowels (*a i u ṛ ḷ e o ai au*). Similarly, *haL* refers to the consonants (all the sounds that are not vowels), and *yaṆ* refers to a specific class of consonants (*y v r l*). Each of these single-syllable words (and the group of sounds that it describes) is known as a *pratyāhāra*.

L1. To what do the following *pratyāhāras* refer? List the sounds in the Answer Sheets:

(i) ...*iK*?
(ii) ...*haṆ*?
(iii) ...*khaY*?

[1] An ancient Indo-European language of India from which many northern Indian languages are derived



(L) Shiva Sutras (2/3)

L2. Give the *pratyāhāras* for the following classes. Write your answers in the Answer Sheets:

(i) ...*ñ m ṅ ṇ n* (nasal consonants)
(ii) ...*ai au* (diphthongs)
(iii) ...all sounds

L3. Explain how to form a *pratyāhāra*. Write your explanation in the Answer Sheets.

You may have noticed that, in English, the same thing can be pronounced differently in different contexts. For example, the words *a* and *an* mean the same thing, but we use *a* before consonants and *an* before vowels. Such rules can often be described as a substitution operation performed under a specified set of conditions, such as "substitute *an* for *a* before a vowel."

An advantage of the *pratyāhāras* is that they can be used to efficiently describe such sound change processes, which often operate on the types of sound groupings that can be expressed as *pratyāhāras*.[2] Approximately 4,000 rules describing the sound change processes of Sanskrit are laid out in another section of Pāṇini's grammar, known as the *Aṣṭādhyāyī*.

Here is an example of a rule from the *Aṣṭādhyāyī*:

6.1.77 *iKaḥ yaṆ aCi*

This rule contains three *pratyāhāras*: *iK*, *yaṆ*, and *aC*, which you have already seen in the previous part of this problem. Each of these *pratyāhāras* is followed by an ending (the ending may be empty, in which case it is marked by the symbol ø):

6.1.77 *iK-aḥ yaṆ-øaC-I*

Here are a few of the changes triggered by this rule 6.1.77; the underlying form is the form before the rule has been applied, while the written form is the result of applying the rule:

Underlying form	Written form	Translation
muniāśrama	*munyāśrama*	'the sages' hermitage'
devīeva	*devyeva*	'the goddess herself'
madhuiva	*madhviva*	'like honey'
pitṛaśva	*pitraśva*	'the father's horse'

[2] In technical linguistic terminology, groups of sounds that have meaningful linguistic roles, such as the set of consonants or the set of vowels, are known as natural classes; *pratyāhāras* are generally natural classes.

(L) Shiva Sutras (3/3)

However, the following forms are not affected by rule 6.1.77 (although they may be affected by other rules):

Underlying form	Translation
munitapas	'the sages' asceticism'
kanyāeva	'the girl herself'
dhenusiva	'like a cow'
kimcit	'something'

In everyday contexts in Sanskrit, the endings seen above (-aḥ, -ø, -i) are used to mark the role of a noun in a sentence. For example, from the noun manas 'mind' the following forms are derived:

Form	Translation	Role in the sentence
manas-ø	'the mind (does, is, etc.)'	subject
manas-aḥ	'of the mind'	possessor
manas-i	'on the mind'	location

However, within the *Aṣṭādhyāyī* these endings have a slightly different meaning.

L4. Express in your own words the meaning of rule **6.1.77**. Write your explanation in the Answer Sheets.

L5. How would you translate the meaning of the following endings as they are used in the *Aṣṭādhyāyī*?

(i) …-aḥ?

(ii) …-Ø?

(iii) …-i?

L6. The following is a simplified version of rule 8.4.53 of the *Aṣṭādhyāyī*: *jhaLaḥ jaś jhaśi*. For each of the following underlying forms, write the corresponding written form; if the form is unaffected, write "no change." Write your answers in the Answer Sheets.

(i) *jagatdhana* 'the wealth of the universe'

(ii) *tatduḥkha* 'that sorrow'

(iii) *bhrātṛnāman* 'brother's name'

Note that the designations of the rules (6.1.77, 8.4.53) refer to book, chapter, and line numbers of the *Aṣṭādhyāyī*.

(M) Adjupectiheaval! (1/3) [10 Points]

You are the administrator of the newest and greatest restaurant review site, whelp.com, which compiles reviews from the most noted gastronomical connoisseurs from around the world.

Recently, you've discovered that dishonest restaurants have been sneakily trying to increase their rating on Whelp! To do this, they're posting thousands of reviews written by spambots, small computer programs that pretend to be human reviewers. To ensure quality, you need to constantly delete these fake reviews. However, being just one administrator, you obviously can't read all of them manually.

Thankfully, spambots make some common mistakes in their fake reviews. Even if a review is grammatically correct, the review still might not make sense; some errors of this category can easily be spotted by anti-spam programs. For example, consider the following two reviews:

(A) At this restaurant, the cake is delicious yet satisfying.
(B) At this restaurant, the cake is delicious and satisfying.

One of these was probably written by a spambot, while the other could plausibly be a real review.

M1. Identify which sentence is spam. Write your answer in the Answer Sheets.

Sometimes, the mistakes made by a spambot may be more subtle. For example, the following sentence is quite reasonable:

The cracker is crunchy and delicious.

But the following sentence is probably not written by a human (or, if so, one with bad taste):

The pudding is crunchy and delicious.

Of course, being able to make these judgements requires some knowledge of the foods involved.[1]

[1] More generally, this form of reasoning aided by human real-world knowledge is termed *knowledge-aware NLP*.

(M) Adjupectiheaval! (2/3)

Having managed to filter out English-language spambots, you've decided to start investigating reviews in Bahasa Indonesia, the national language of Indonesia. However, your task is complicated by the fact that you don't speak Indonesian! In order to write filtering software, you first examine some reviews written by real humans, about popular Indonesian foods such as *kemplang* and *poffertjes*.

1. *Kemplang manis namun berminyak.*
2. *Rengginang manis dan lezat.*
3. *Poffertjes manis serta lezat.*
4. *Rempeyek lezat dan menggugah selera.*
5. *Lemang menggugah selera dan manis.*
6. *Onde-onde lezat namun mahal.*
7. *Poffertjes baik namun mahal.*
8. *Kemplang baik dan sehat.*
9. *Lemang sehat serta manis.*
10. *Rempeyek berminyak dan hambar.*
11. *Rengginang tidak sehat serta mahal.*
12. *Onde-onde berminyak dan tidak sehat.*

Despite not knowing anything about the food items mentioned in the reviews, or anything about the Indonesian language itself, you realize that this is enough to filter out some spam reviews!

M2. Below are six reviews. Three of them are almost certainly spam, while the other three could have been written by a human. Indicate whether the review is real or spam, where "real" means it could be a real review and "spam" means it's probably spam. Record your answers in the Answer Sheets.

13. *Kemplang menggugah selera serta baik*.
14. *Rengginang hambar namun sehat*.
15. *Poffertjes baik namun tidak berminyak*.
16. *Rempeyek tidak manis serta berminyak*.
17. *Lemang manis dan sehat*.
18. *Onde-onde sehat namun tidak menggugah selera*.

(M) Adjupectiheaval! (3/3)

The algorithm you've designed using this knowledge works well, but you find that there are still some words and reviews that stump it. Here are some examples of real (non-spam) sentences in Indonesian that confuse your algorithm:

19. *Onde-onde halus dan manis.*
20. *Rengginang halus serta hambar.*
21. *Rempeyek garing serta baik.*
22. *Lemang tidak mahal namun garing.*
23. *Lemang halus dan tidak mahal.*
24. *Onde-onde berminyak dan garing.*
25. *Renggingang lembut namun lezat.*
26. *Lemang lembut namun mahal.*
27. *Rempeyek garing dan sehat.*

You quickly realize that to fully understand these sentences, you're going to have to read up more about these food items. Unfortunately, you only have access to a monolingual Indonesian dictionary (entries below): [2]

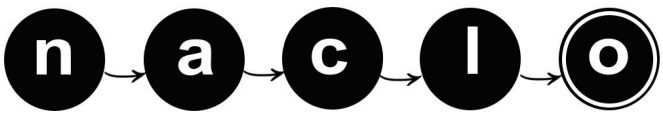
- ***Kemplang*** *adalah sebuah kerupuk ikan yang umum ditemukan di belahan selatan Sumatra, Indonesia. Kerupuk kemplang dikeringkan dan kemudian dipanggang atau digoreng.*
- ***Lemang*** *adalah kue dari beras ketan yang dimasak dalam seruas bambu, setelah sebelumnya digulung dengan selembar daun pisang.*
- ***Rempeyek*** *adalah sejenis makanan pelengkap dari kelompok gorengan. Fungsi rempeyek sama dengan kerupuk yaitu sebagai pelengkap hidangan.*
- ***Rengginang*** *adalah sejenis kerupuk tebal yang terbuat dari beras ketan dibentuk bulat yang digoreng panas dalam minyak goreng.*
- ***Onde-onde*** *adalah sejenis kue yang populer di Indonesia. Ini sangat terkenal di daerah Mojokerto yang disebut sebagai kota onde-onde sejak zaman Majapahit.*
- ***Poffertjes*** *adalah kue tradisional yang empuk dari Belanda. Penampilannya mirip panekuk, tetapi lebih kecil dan manis.*

With this new information, you find that you can deduce which reviews are real or spam!

M3. For each of reviews 28-31, indicate whether the review is real or spam, where "real" means it could be a real review and "spam" means it's probably spam. Record your answers in the Answer Sheets.

28. *Onde-onde halus serta mahal.*
29. *Rempeyek lembut namun tidak sehat.*
30. *Kemplang garing dan tidak berminyak.*
31. *Poffertjes garing serta hambar.*

[2] Adapted from Bahasa Indonesia Wikipedia.



(N) You Can't Handle the Truth (1/3) [5 Points]

Alfred, a student at North Semantick High School, is upset about how many falsehoods pass for truth these days. So, he sets about building TruthBot, a talking robot which will only ever say true things.

Alfred starts by making a file in TruthBot's memory called the *True List*, and loading three true statements into the list. Here is what the *True List* now looks like:

> *True List*
> The United Kingdom contains 4 countries.
> The Senators hockey team plays home games in Ottawa.
> Theodore Roosevelt fought in the Spanish-American War.

Then he programs TruthBot to say statements from the *True List*, and nothing else. This works perfectly; when Alfred turns TruthBot on, the machine says things like:

> (1) TruthBot: "The Senators hockey team plays home games in Ottawa."

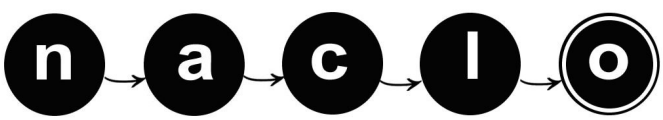
But this is not very exciting — no matter how long TruthBot runs, it only says three distinct statements! What's worse, each time Alfred adds one new statement to the *True List*, TruthBot only says one more new statement.

So, Alfred modifies TruthBot. First, he updates the *True List*, adding three new true statements. Here is what it looks like now:

> *True List*
> The United Kingdom contains 4 countries.
> The Senators hockey team plays home games in Ottawa.
> Theodore Roosevelt fought in the Spanish-American War.
> 2 + 2 = 4.
> Ottawa is the capital city of Canada.
> Theodore Roosevelt was the 26th President of the U.S.

Then, Alfred changes TruthBot's programmed instructions. He keeps the original instruction, but adds another, slightly more complex instruction. When he turns TruthBot on this time, it says things like:

> (2) TruthBot: "The Senators hockey team plays home games in the capital city of Canada."
> (3) TruthBot: "2 + 2 = 2 + 2."
> (4) TruthBot: "Theodore Roosevelt fought in the Spanish-American War."
> (5) TruthBot: "The capital city of Canada is Ottawa."



(N) You Can't Handle the Truth (2/3)

Alfred counts 18 total distinct statements that TruthBot now says. Better yet, they're all true (even if some are a little less informative than others)!

N1. State the new instruction that Alfred added to TruthBot's programming. You may describe the instruction however you like (using words, symbols, or anything else), as long as your answer is clear and accurate. Write your answer in the Answer Sheets.

N2. Before he updated TruthBot, Alfred observed that one addition to the *True List* always produced one new, distinct statement uttered by TruthBot. After the update, how many new statements will TruthBot utter after a single addition is made to the *True List*? The answer might vary depending on the statement; if so, explain as fully as you can what the number will be for different types of statements. Write your answer in the Answer Sheets.

Alfred lends TruthBot to his friend Ruth for testing. The next day, Ruth reports back to Alfred. "I'm sorry, Alfred," she says, "but your TruthBot is badly broken – it doesn't always tell the truth!"

"No way!" exclaims Alfred. "You're telling me that TruthBot says falsehoods?"

"Sometimes, yes," says Ruth, "but other times it says things that are just weird. I don't know whether some of its utterances are true or false, because I'm not sure how to interpret them."

"I don't believe it!" replies Alfred. "Did you modify its programmed instructions?"

"No," says Ruth. "All I did was add a few statements to its True list. But I can guarantee that I only added true statements – I know for sure, since the statements I added were about me."

N3. For each of the statements below, give a *new* statement TruthBot will utter when that statement (and just that statement) is added to the six-item *True List* Alfred passed over to Ruth. For example, for (d), give a statement, different from (d), that TruthBot says when (d) becomes the seventh item on the *True List*, but not before. Write your answers in the Answer Sheets.

(a) Ruth has 4 siblings.
(b) Ruth knows the capital city of Canada.
(c) Ruth prefers the Washington Capitals to the Ottawa Senators.
(d) Ruth named her stuffed, toy bear after Theodore Roosevelt.
(e) Ruth named her stuffed, toy bear Theodore Roosevelt.

(N) You Can't Handle the Truth (3/3)

N4. For each statement you gave in N3, assess whether it would have seemed true, false, or "just weird" to Ruth and Alfred (since they agreed on all their assessments, when they listened together), and explain why. If you think the statement would have seemed either true or false, but don't know enough about Ruth to say which, select false, and say so in your explanation. (Note: your explanation is more important than your assessment.) Write your answers in the Answer Sheets.

N5. How could Alfred and Ruth modify TruthBot's instructions, so that it still utters true statements like (2)-(5), but makes fewer mistakes (i.e., says fewer false or weird statements)? (This is a difficult task! You don't need to cover every possible case where TruthBot might make a mistake, and you may describe what TruthBot would have to know or be able to do without saying exactly *how* that knowledge or ability could be programmed into it.) Write your answer in the Answer Sheets.

(O) We're Counting on Yoruba (1/1) [10 Points]

Here are some numbers in Yoruba, a language spoken in West Africa by 30 million people:

èji	2
ẹ̀rin	4
àrun	5
ẹ̀rinlá	14
eéjìdilogun	18
ẹ́ẹ̀rìndilogóji	36
ẹ̀rìndogóji	44
àádorin	70
ẹ́ẹ̀tàdilogórin	77
ẹ̀tàdogórin	83

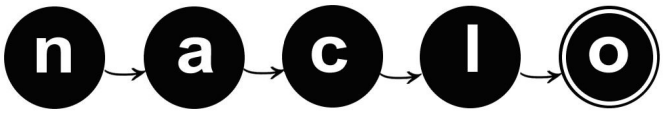
The accents on the vowels represent tones; *e* and *ẹ* represent different vowel sounds.

O1. In the Answer Sheets, write the following Yoruba numbers in Arabic numerals, as written in the second column.

a. **àádota**
b. **àrúndogórin**
c. **aárùndilogórin**
d. **ẹ̀tàdogórun**
e. **òkándilogóji**

O2. In the Answer Sheets, write the following Arabic numerals in Yoruba.

a. 12
b. 90
c. 57
d. 45
e. 99



(P) Doubling Up on Nakanai (1/2) [15 Points]

Nakanai (*Lakalai*) is a language spoken by around 13,000 people of the Nakanai tribe in West New Britain, Papua New Guinea. Below are some words in Nakanai given along with their so-called "reduplicated form," a grammatical form which has a large number of different uses (e.g. plural verb marking, habituative mood[1], and collective plurals).
However, these uses are irrelevant to the derivation of the reduplicated form.

Underlying Form	Reduplicated Form	Meaning
basi	*baibasi*	'bandicoot (a small animal)'
beta	*babeta*	'wet'
galolo	*galololo*	'constantly'
gove	*goegove*	'mountain'
Kise	*Kekise*	(a name)
baimopu	*baimoumopu*	(a kind of fish)
seku	*seuseku*	'to scoop up'
hugu	*hugugu*	'to carry'
sasa	*sasasa*	'one'
mari	*marimari*	'to know'
toa	*tatoa*	'to tread, to kick'
kirosi	*kirosirosi*	'angry'
ligi	*ligiligi*	'to hurt'
kebo	*kokebo*	'weak, exhausted'
kevemuki	*kevemuimuki*	(a name)
baharu	*bahararu*	'widow'
golu	*golugolu*	'thing'
mota	*mamota*	'vine'
kedi	*keikedi*	'to be careful'
vigilemuli	*vigilemulimuli*	'to tell a story'

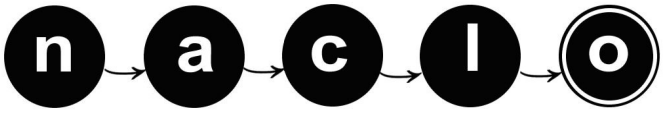
[1] A verbal feature denoting that the action takes place habitually; e.g. English "*I **usually** go on walks*"

(P) Doubling Up on Nakanai (2/2)

P1. Provide the reduplicated forms of the following words. Write your answers in the Answer Sheets.

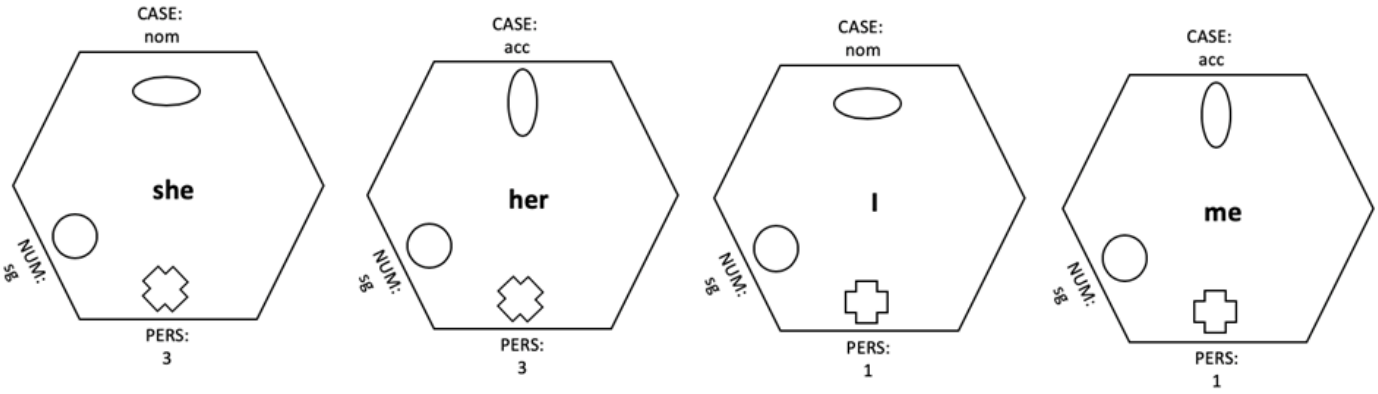
Underlying Form	Reduplicated Form	Meaning
tahalo	*?*	'man'
sekela	*?*	'one at a time'
pita	*?*	'muddy'
bake	*?*	(a kind of fish)
loke	*?*	'to break (a rope)'
voro	*?*	'to pound'
valolohoka	*?*	'to warn someone of trouble'
pasi	*?*	'extremely'
kusa	*?*	'to shout'
bebe	*?*	'butterfly'
hilo	*?*	'to see'
sivo	*?*	'to descend'
rabu	*?*	'charred wood'
tarile	*?*	'tree'
sobe	*?*	'young woman'
vitaumetari	*?*	'younger sibling'
vituga	*?*	'to walk'

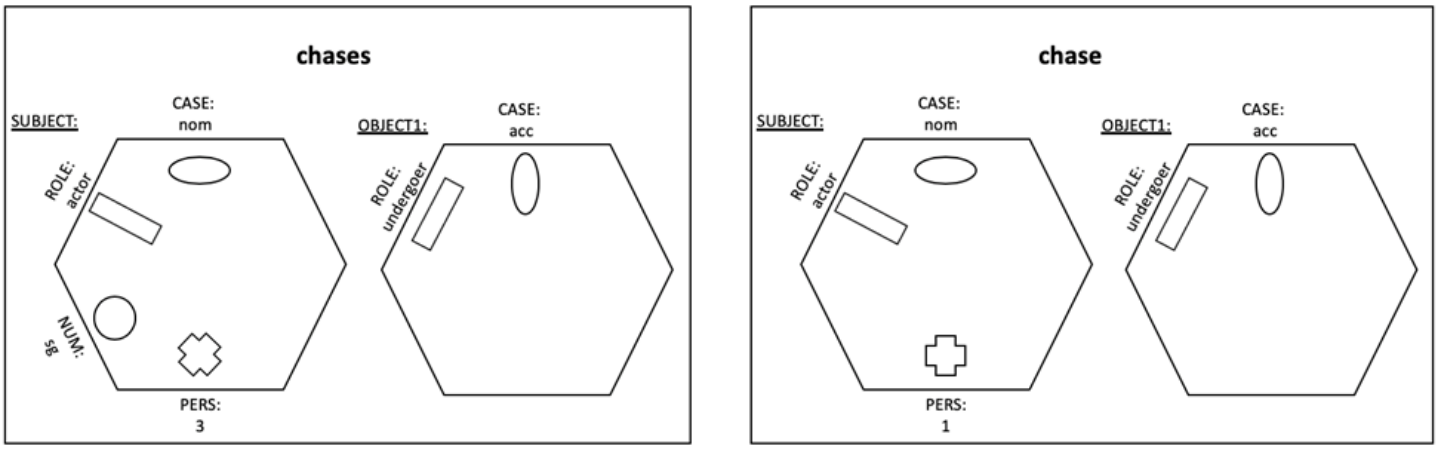
P2. Explain your solution. Write your explanation in the Answer Sheets.



(Q) Cut to the Chase (1/5) [10 Points]

One way for computers to understand language is to form structures that show how words in a sentence relate to each other. Unification Grammar is one way to build such structures. The structures of words are combined to make the structures of sentences. Here are the structures for the words *she*, *her*, *I*, *me*, *chases*, and *chase* in Unification Grammar:

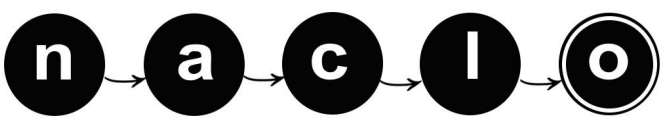




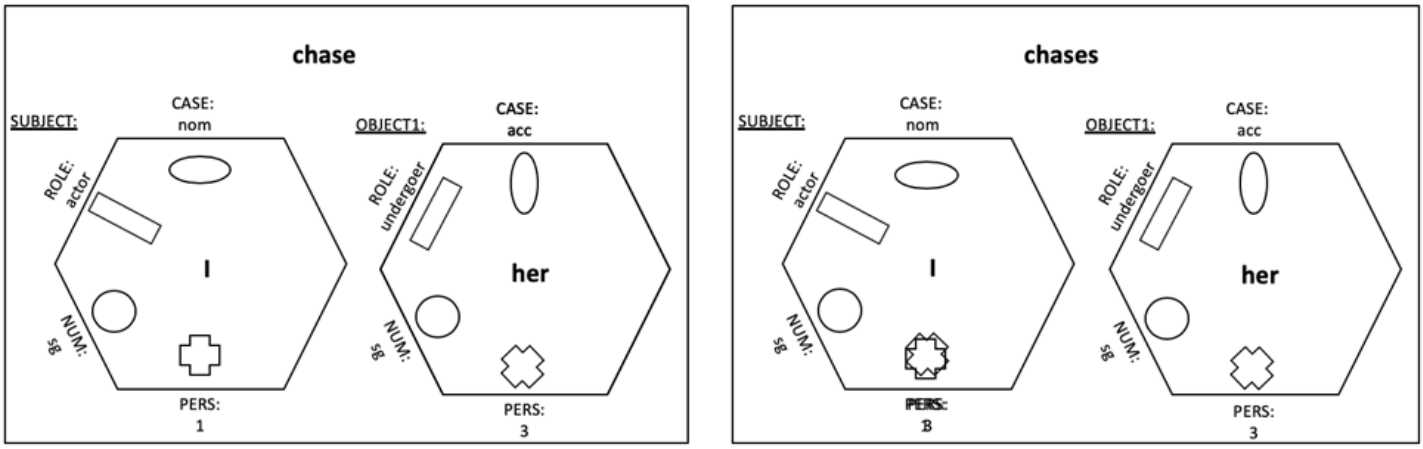
You can do this problem without knowing words like CASE, NUM(ber), PERS(on), SUBJECT, and OBJECT. In fact, if you know what these words mean, be careful because linguists define them in a special way. However, the words *actor* and *undergoer* are important to this problem; the actor does something, and something happens to the undergoer.

How to make a sentence with Unification Grammar

Unify (combine) the structures for nouns with the SUBJECT and OBJECT structures of verbs. Visually, this works by placing the structures for the words being unified on top of each other. Unification only works when all of the information is compatible. For example, on the left at the top of the next page, you can see the unified structure for *I chase her*, which is a successful unification. However, on the right, you can see that you cannot unify *I chases her* because the PERS feature for *I* clashes with the PERS feature for the subject of *chases*.



(Q) Cut to the Chase (2/5)



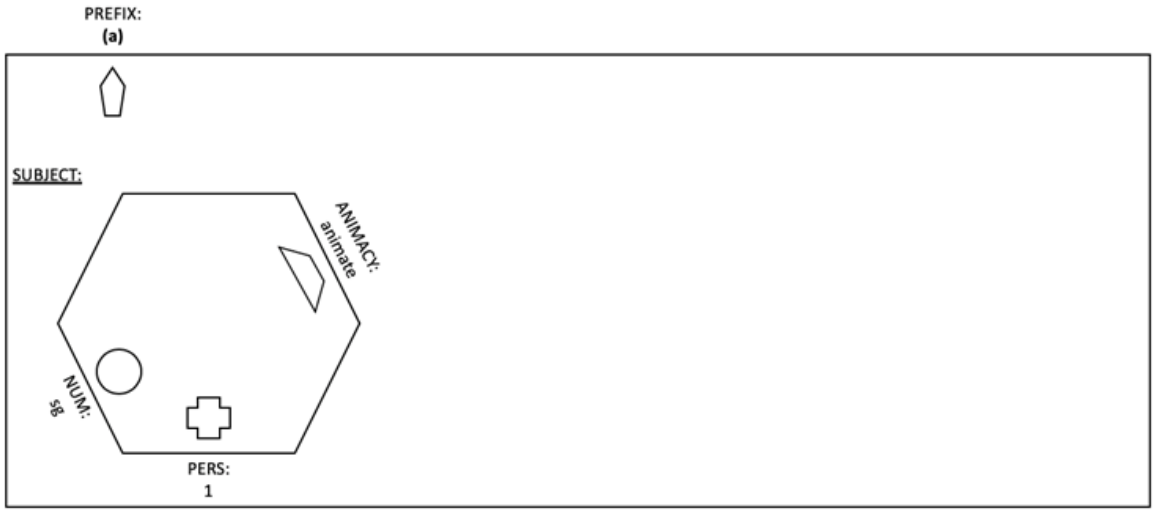
The unification grammar does two things: (1) when we build a structure, we can see who (actor) chases who (undergoer) (for example, in the structure on the left above, we can see that the actor is *I* and the undergoer is *her*); and (2) if we cannot build a structure, we know that there is no English sentence that combines those words in that particular way (as is the case with *I chases her* on the right above).

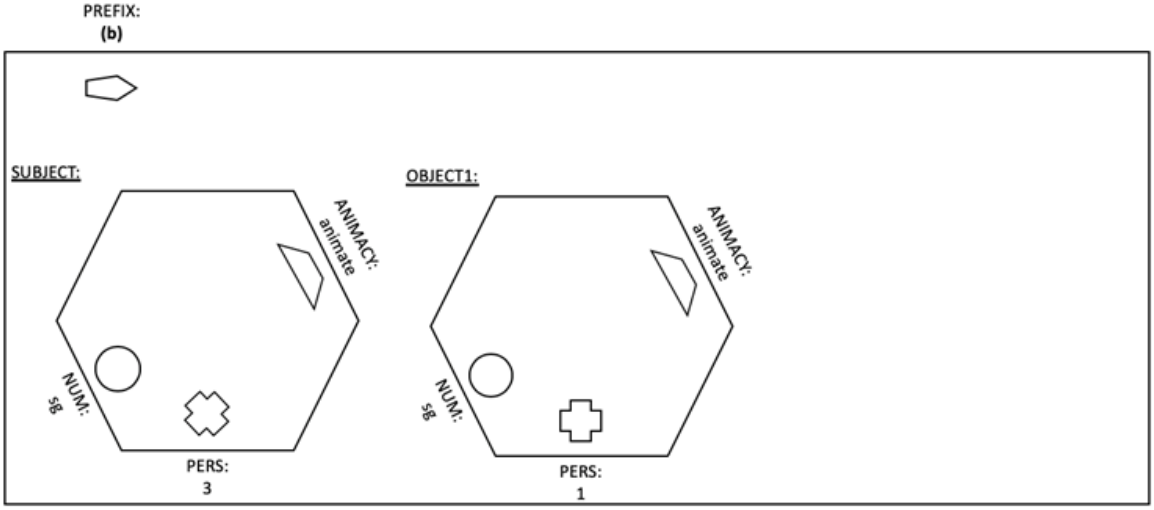
Unification Grammar for Maasai[1] words

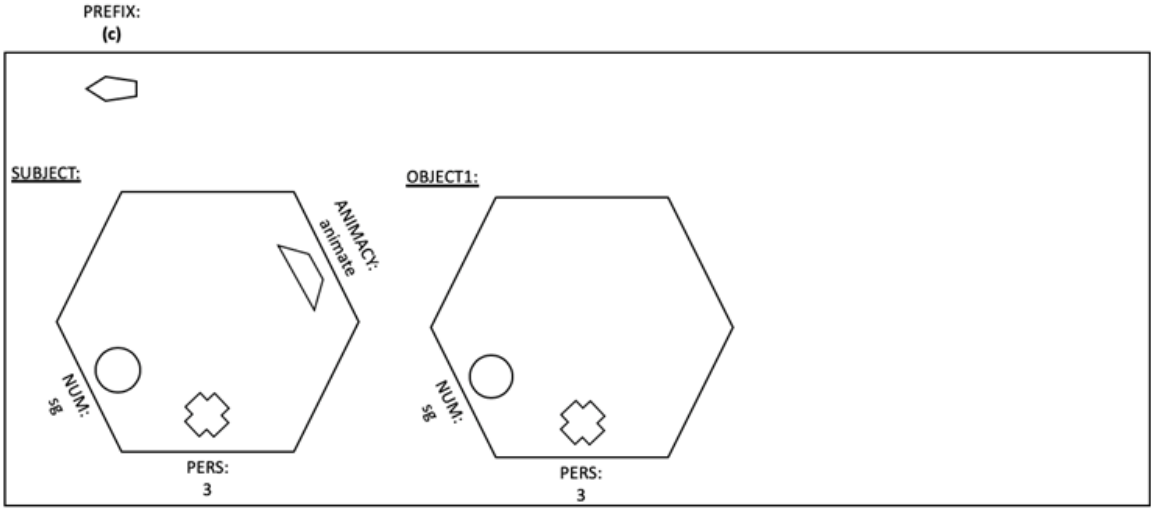
Now you will work with a unification grammar for the Maasai language. There are three new things in the Maasai grammar, (1) There is a new ROLE, *beneficiary*, indicating *for someone*, (2) Some structures have OBJECT1 and OBJECT2, and (3) There are structures for verb prefixes and suffixes, including a special structure that you use when there is no suffix on the verb.

[1] Maasai, Masai, or Maa language is spoken in southern Kenya and northern Tanzania by about 900,000 people.

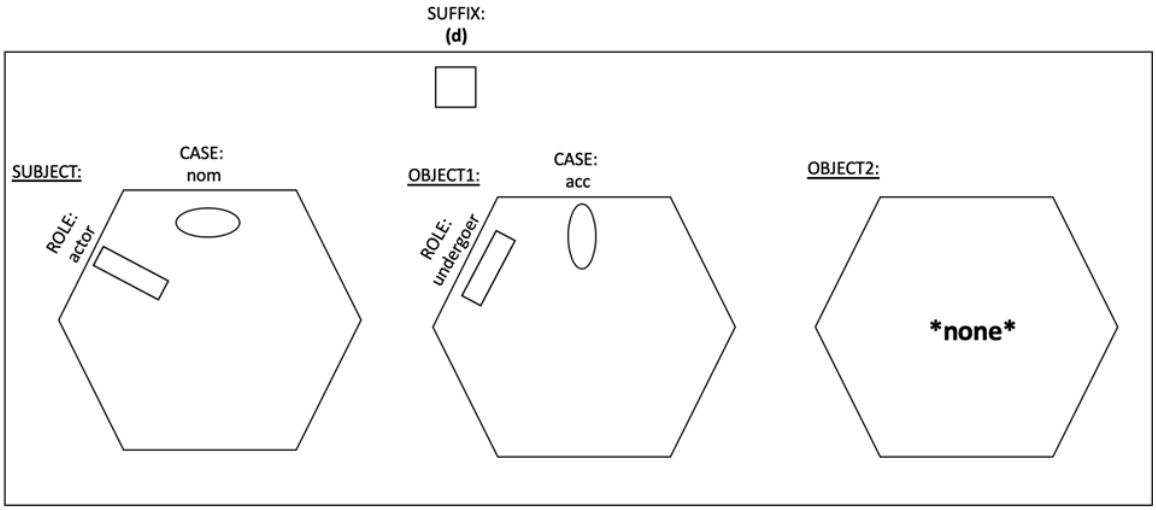
(Q) Cut to the Chase (3/5)

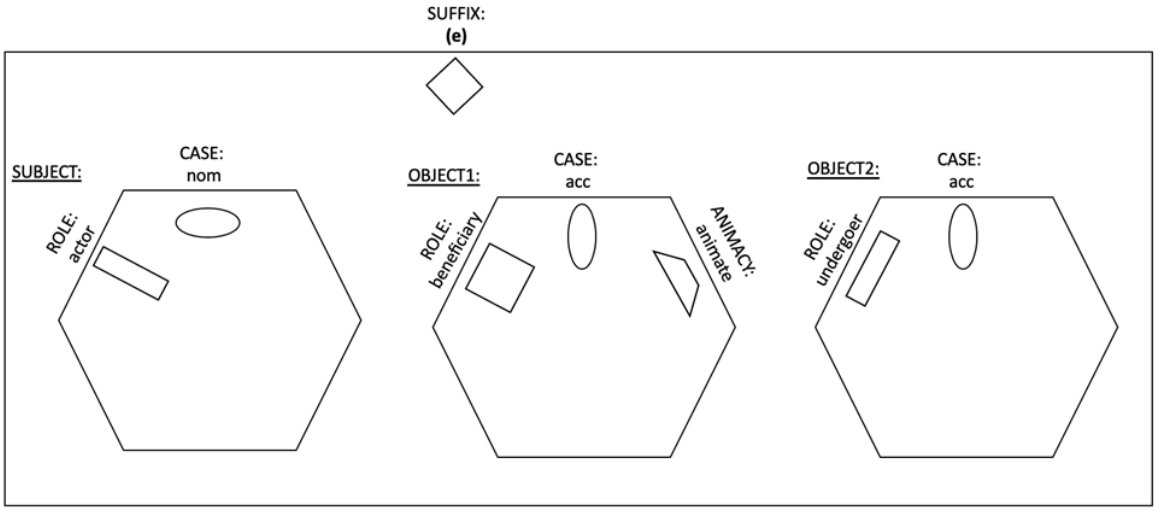


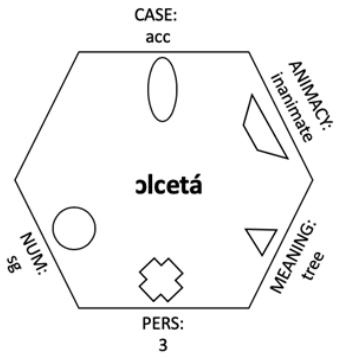


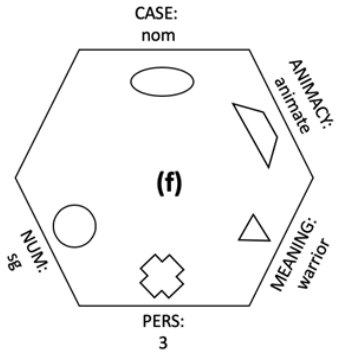


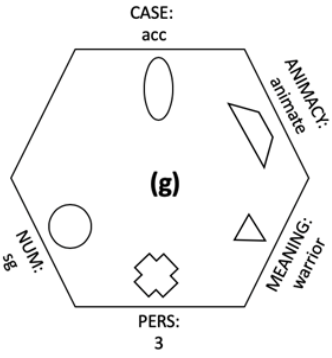
(Q) Cut to the Chase (4/5)

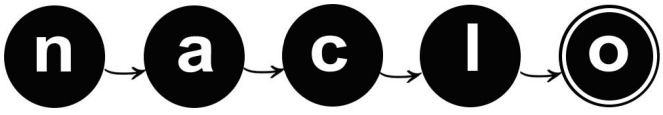












(Q) Cut to the Chase (5/5)

Here are 11 sentences. 9 of them are valid Maasai sentences, but 2 of them are invalid because there is no possible way to unify the words:

A. **ádúŋ ɔlmʊraní ɔlcɛtá**
B. **ádúŋokí ɔlmʊráni ɔlcɛtá**
C. **ádúŋ ɔlmʊráni**
D. **áaduŋokí ɔlmʊráni**
E. **áaduŋokí ɔlcɛtá**
F. **áadúŋ ɔlmʊráni**
G. **áadúŋ ɔlmʊraní**
H. **édúŋ ɔlmʊraní ɔlcɛtá**
I. **édúŋokí ɔlmʊráni ɔlcɛtá**
J. **édúŋokí ɔlmʊraní**
K. **édúŋ ɔlmʊráni**

Q1. Match the missing components of the structures above (indicated by letters (a) through (g)) with the Maasai words and word parts below. Write your answers in the Answer Sheets. [HINT: 1, 2, and 3 go with (a), (b), and (c) (not necessarily in that order); 4 and 5 go with (d) and (e) (not necessarily in that order).]

1. **áa-**
2. **á**-
3. **é-**
4. ***none***
5. **-okí**
6. **ɔlmʊraní**
7. **ɔlmʊráni**

Q2. Which two sentences (from A through K above) are not valid Maasai sentences? Indicate the letters of those sentences in the Answer Sheets.

Q3. Translate the 9 valid sentences in the Answer Sheets. You should leave blank the 2 sentences that were your answer to Question 2. Notes:

- Some sentences may have more than one valid translation; if that is the case, you only need to provide one valid translation.
- In some structures, there will be no meaning specified for the subject, object1, or object2. Here are the translations you should use in those cases:

Features	Translation
PERS: 1, NUM: sg, ANIMACY: animate, CASE: nom	I
PERS: 1, NUM: sg, ANIMACY: animate, CASE: acc	me
PERS: 3, NUM: sg, ANIMACY: animate, CASE: nom	he/she
PERS: 3, NUM: sg, ANIMACY: animate, CASE: acc	him/her
PERS: 3, NUM: sg, CASE: nom	he/she/it
PERS: 3, NUM: sg, CASE: acc	him/her/it

(R) The Obviative Solution (1/1) [10 Points]

Arapaho is an Algonquian language spoken by about 1,000 people in Wyoming and Oklahoma. Here are some Arapaho nouns in several forms and their English translations. Note that the shaded cells indicate that the form does not exist.

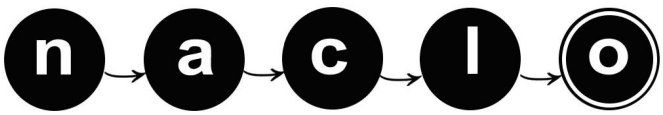
Singular	Plural	Obviative Singular	Locative Singular	Meaning
hisei	hiseinoʔ	hisein	hiseinewe ʔ	*'woman'*
hotii	hotiiwoʔ	hotiiw	hotiiwowe ʔ	*'car'*
nebi	nebihoʔ	hibio	nebihewe ʔ	*'one's older sister'*
neicet	neicetino		neicetine ʔ	*'one's hand'*
nooku	nookuhoʔ	nookuo	nookuhowe ʔ	*'beaver'*
hiseeθ	hiseetoʔ	hiseet	**a.**	*'pine tree'*
b.	ooto		oote ʔ	*'leg'*
beiciθ	beicito		beicite ʔ	*'tooth'*
coox	**c.**	**d.**	**e.**	*'enemy'*
ceʔeinox	ceʔeinoθo		ceʔeinoθe ʔ	*'bag'*
hinen	hinenino ʔ	**f.**	**g.**	*'man'*
wotoo	**h.**	**i.**	wotoohe ʔ	*'pair of pants'*
j.	woθonohoeno	**k.**	woθonohoene ʔ	*'book'*
l.	**m.**	niiʔeihiio	**n.**	*'eagle'*
ceʔibes	ceʔibexo	**o.**	**p.**	*'block (of wood)'*
benes	**q.**	**r.**	**s.**	*'arm'*
t.	nesihoʔ	**u.**	**v.**	*'one's uncle'*

Notes: ʔ and θ are both consonants. ʔ is a glottal stop, the sound in the middle of "uh-oh," and θ is pronounced like the "th" in the English word "think." Arapaho pronunciation also involves tones, which have not been included in this problem.

The *obviative*, sometimes called the "fourth person," is a noun form used in some languages to express how relevant an entity is. If some third-person entities (i.e., ones that are neither the speaker nor the listener) are less important than others to the conversation, they will be given the obviative marking, while the more important ones will take the standard third-person marking. The *locative* is a noun form used to indicate a location – e.g., the locative form of "field" would mean "in the field" or "on the field."

R1. Fill in the missing cells (a., b., etc.). If you think the form does not exist, write N/A. Write your answers in the Answer Sheets.

R2. Explain your solution in the Answer Sheets.



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

Contest Booklet

REGISTRATION NUMBER				

Name: _____

Contest Site: _____

Site ID: _____

City, State: _____

Grade: _____

Please also make sure to **write your registration number and your name on each page** that you turn in.

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 (1/10)

(I) Kãa, Kai, and Khai

1.

1. ☐	2. ☐	3. ☐	4. ☐	5. ☐
6. ☐	7. ☐	8. ☐	9. ☐	10. ☐
11. ☐	12. ☐	13. ☐	14. ☐	15. ☐
16. ☐	17. ☐	18. ☐	19. ☐	20. ☐

2. a. *lõm* ☐

b. *kãa* ☐

c. *chiãng* ☐

3. “beautiful line” ☐

(J) You Made Me See

1. (i) ☐

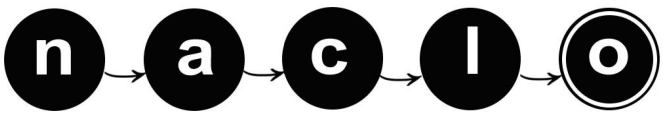
(ii) ☐

(iii) ☐

2. (i) ☐

(ii) ☐

(iii) ☐



YOUR NAME: REGISTRATION #

Answer Sheets (2/10)

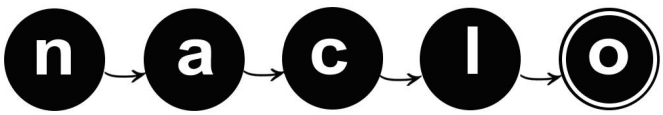
(J) You Made Me See (continued)

3.

(K) The Pimpled Toad

1.

	Missing Miao Pinyin	English		Missing Miao Pinyin	English
1			**10**		
2			**11**		
3			**12**		
4			**13**		
5			**14**		
6			**15**		
7			**16**		
8			**17**		
9					



YOUR NAME: REGISTRATION #

Answer Sheets (3/10)

(K) The Pimpled Toad (continued)

2.

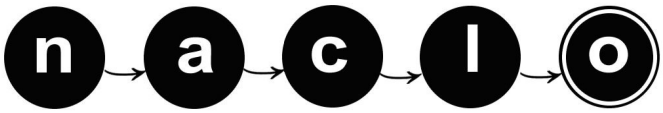
3.

YOUR NAME: REGISTRATION #

Answer Sheets (4/10)

(L) Shiva Sutras

1. (i)
   (ii)
   (iii)
2. (i)
   (ii)
   (iii)
3.
4.
5. (i)
   (ii)
   (iii)
6. (i)
   (ii)
   (iii)



YOUR NAME: REGISTRATION #

Answer Sheets (5/10)

(M) Adjupectiheaval!

1. Sentence ( ) is spam.

2. Circle the correct answer:

(13)	Real	Spam
(14)	Real	Spam
(15)	Real	Spam
(16)	Real	Spam
(17)	Real	Spam
(18)	Real	Spam

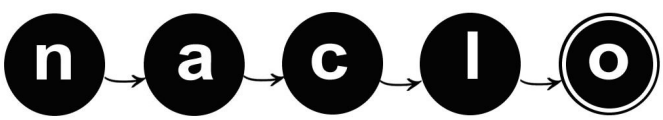
3.

(28)	Real	Spam
(29)	Real	Spam
(30)	Real	Spam
(31)	Real	Spam

(N) You Can't Handle the Truth

1.

2.



YOUR NAME: REGISTRATION #

Answer Sheets (6/10)

(N) You Can't Handle the Truth (continued)

3. (a)

(b)

(c)

(d)

(e)

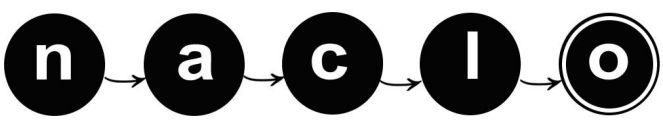
4. (a) Select one: True False Weird

(b) Select one: True False Weird

(c) Select one: True False Weird

(d) Select one: True False Weird

(e) Select one: True False Weird



YOUR NAME: REGISTRATION #

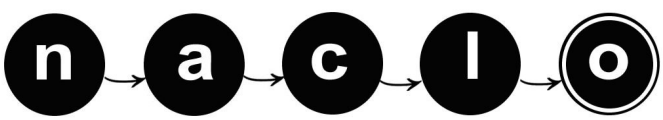
Answer Sheets (7/10)

(N) You Can't Handle the Truth (continued)

5.

(O) We're Counting on Yoruba

1. (a)
   (b)
   (c)
   (d)
   (e)
2. (a)
   (b)
   (c)
   (d)
   (e)



YOUR NAME: REGISTRATION #

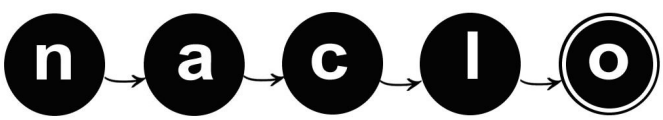
Answer Sheets (8/10)

(P) Doubling Up on Nakanai

1.

Underlying Form	Reduplicated Form	Meaning
tahalo		'man'
sekela		'one at a time'
pita		'muddy'
bake		(a kind of fish)
loke		'to break (a rope)'
voro		'to pound'
valolohoka		'to warn someone of trouble'
pasi		'extremely'
kusa		'to shout'
bebe		'butterfly'
hilo		'to see'
sivo		'to descend'
rabu		'charred wood'
tarile		'tree'
sobe		'young woman'
vitaumetari		'younger sibling'
vituga		'to walk'

2.



YOUR NAME: REGISTRATION #

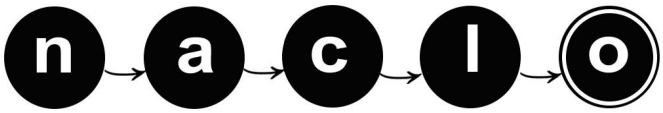
Answer Sheets (9/10)

(Q) Cut to the Chase

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

2. ☐ and ☐

3. Sentence A: ☐
   Sentence B: ☐
   Sentence C: ☐
   Sentence D: ☐
   Sentence E: ☐
   Sentence F: ☐
   Sentence G: ☐
   Sentence H: ☐
   Sentence I: ☐
   Sentence J: ☐
   Sentence K: ☐



YOUR NAME: REGISTRATION #

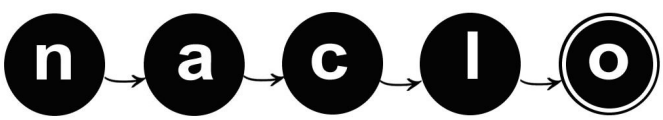
Answer Sheets (10/10)

(R) The Obviative Solution

1.

a.		**l.**	
b.		**m.**	
c.		**n.**	
d.		**o.**	
e.		**p.**	
f.		**q.**	
g.		**r.**	
h.		**s.**	
i.		**t.**	
j.		**u.**	
k.		**v.**	

2.



YOUR NAME: REGISTRATION #

Additional Space for Answers