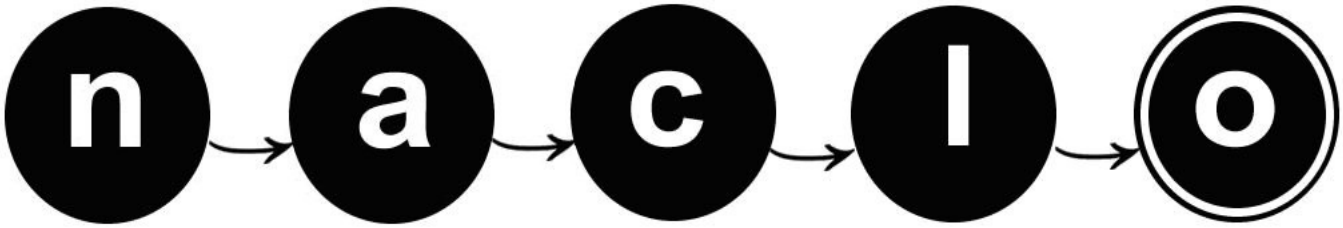


IMPORTANT RULES

To ensure the integrity of the contest:

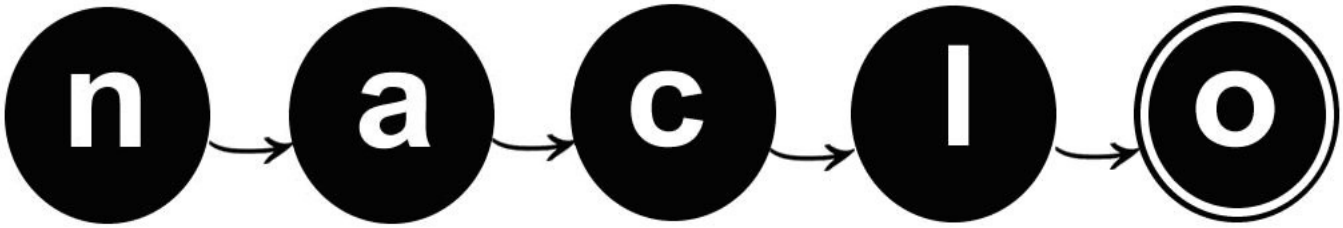
1. Do not discuss the contents of this booklet with anyone during and after the contest (until it has been posted on the NACLO web site in late March). If you have any questions during the contest, talk quietly to the local facilitators, who will relay your questions to the jury and then give you the official jury answer.

2. Students are not allowed to keep any pages of the booklet after the contest is over.

THE ACTUAL CONTEST BOOKLET STARTS ON PAGE 3

Invitational Round
March 10, 2010

THIS PAGE HAS BEEN INTENTIONALLY LEFT BLANK



The Association for Computational Linguistics
North American Chapter

Carnegie Mellon

The Fourth Annual

North American Computational Linguistics Olympiad

2010

www.naclo.cs.cmu.edu

Invitational Round
March 10, 2010

The North American Computational Linguistics Olympiad
www.naclo.cs.cmu.edu

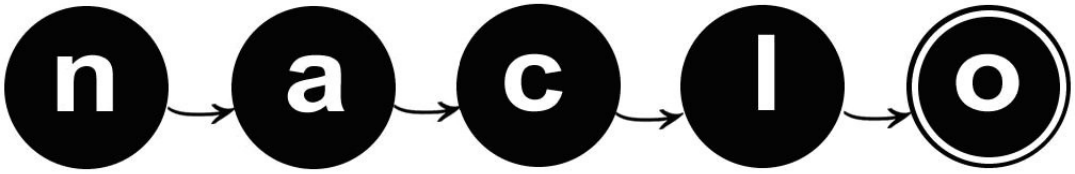
Contest Booklet

Your Name: _____
Registration Number: _____
Your School: _____
City, State, Zip: _____
Your Grade: _____

Start Time: _____
End Time: _____

Your Teacher's Name: _____

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



Welcome to the fourth annual North American Computational Linguistics Olympiad! 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 five hours long, including a break, and has nine problems (H—P).
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. Each problem is worth a specified number of points, with a total of 100 points. **In this year's open round, a number of points will be given for explanations**.
5. We will grade only work in this booklet. All your answers should be in the spaces provided in this booklet. DO NOT WRITE ON THE BACK OF THE PAGES.
6. Write your name and registration number on each page:
   Here is an example: Jessica Sawyer #850
7. Each problem has been thoroughly checked by linguists and computer scientists as well as students like you for clarity, accuracy, and solvability. Some problems are more difficult than others, but all can be solved using ordinary reasoning and analytic skills. You don't need to know anything about linguistics or about these languages in order to solve them.
8. 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.
9. If you have any comments, suggestions or complaints about the competition, we ask you to remember these for the web based evaluation. We will send you an e-mail after the competition is finished with instructions on how to fill it out.
10. The top 8 scorers on the invitational round will be invited to represent the US at the International Linguistics Olympiad in Sweden. In case of ties, the NACLO organizers reserve the right to consider other factors, such as open round score and the quality of solutions to specific problems.
11. **DO NOT DISCUSS THE PROBLEMS UNTIL THEY HAVE BEEN POSTED ONLINE! THIS MAY BE 3-4 WEEKS AFTER THE END OF THE CONTEST.**

Oh, and have fun!

PART I

(problems HIJKLM)

(before the break)

Do not work on this part after the break. You have three hours for this part.

YOUR NAME: REGISTRATION #:

(10 points)

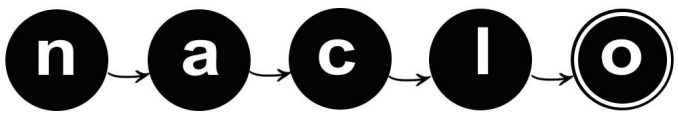
(H) Ardhay Uzzlepay (1/4)

Minangkabau is an Austronesian language spoken by about 7 million people around the West Sumatran city of Padang in Indonesia. Its speakers generally also speak Indonesian but Minangkabau is a distinct language.

Minangkabau has a number of 'play languages' that people use for fun, like Pig Latin in English. Ordinary language words are changed into play language by following just a few rules. One of these 'play languages' is called *Sorba* while another is called *Solabar.*

Here are some examples of standard Minangkabau words and their Sorba play language equivalents:

Standard Minangkabau	**Sorba**	**English Translation**
raso	*sora*	'taste, feeling'
rokok	*koro*	'cigarette'
rayo	*yora*	'celebrate'
susu	*sursu*	'milk'
baso	*sorba*	'language'
lamo	*morla*	'long time'
mati	*tirma*	'dead'
bulan	*larbu*	'month'
minum	*nurmi*	'drink'
lilin	*lirli*	'wax, candle'
mintak	*tarmin*	'request'
cubadak	*darcuba*	'jackfruit' (a large tropical fruit)
mangecek	*cermange*	'talk'
bakilek	*lerbaki*	'lightning'
sawah	*warsa*	'rice field'
pitih	*tirpi*	'money'
manangih	*ngirmana*	'cry'
urang	*raru*	'person'
apa	*para*	'father'
iko	*kori*	'this'
gata-gata	*targa-targa*	'flirtatious'
maha-maha	*harma-harma*	'expensive'
campua	*purcam*	'mix'



YOUR NAME: REGISTRATION #:

(H) Ardhay Uzzlepay (2/4)

H1 (2 points). Using the same rules that you have discovered from examining the words in the Table above, write the Sorba equivalents of the following standard Minangkabau words in the Table below.

Standard Minangkabau	**Sorba**	**English**
rancak		'nice'
jadi		'happen'
makan		'eat'
marokok		'smoking'
ampek		'hundred'
limpik-limpik		'stuck together'
dapua		'kitchen'

H2 (2 points). If you know a Sorba word, can you work backwards to standard Minangkabau? Demonstrate with the Sorba word *lore* which means 'good'.

YOUR NAME: REGISTRATION #:

(H) Ardhay Uzzlepay (3/4)

H3 (4 points). The other 'play language' is called *Solabar*. The rules for converting a standard Minangkabau word to *Solabar* can be worked out from the following examples:

Standard Minangkabau	**Solabar**	**English**
baso	*solabar*	'language'
campua	*pulacar*	'mix'
makan	*kalamar*	'eat'

What is the Solabar equivalent of the Sorba word *tirpi* 'money'? How many different possible answers are there based on the evidence that you have? List all of these hypotheses, from most likely to least likely.

What one or two other words in Minangkabau would you like to see translated to Solabar in order to rule out all of these hypotheses except for one? The remaining hypothesis may or may not be the likeliest one that you selected above.

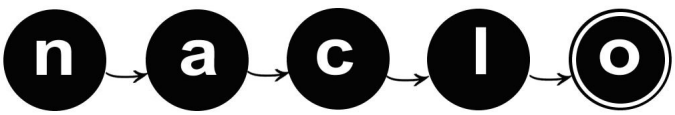
YOUR NAME: REGISTRATION #:

(H) Ardhay Uzzlepay (4/4)

H4 (2 points). In writing Minangkabau does the sequence 'ng' represent **one** sound or **two** sounds?

_____

Provide evidence that supports your answer.

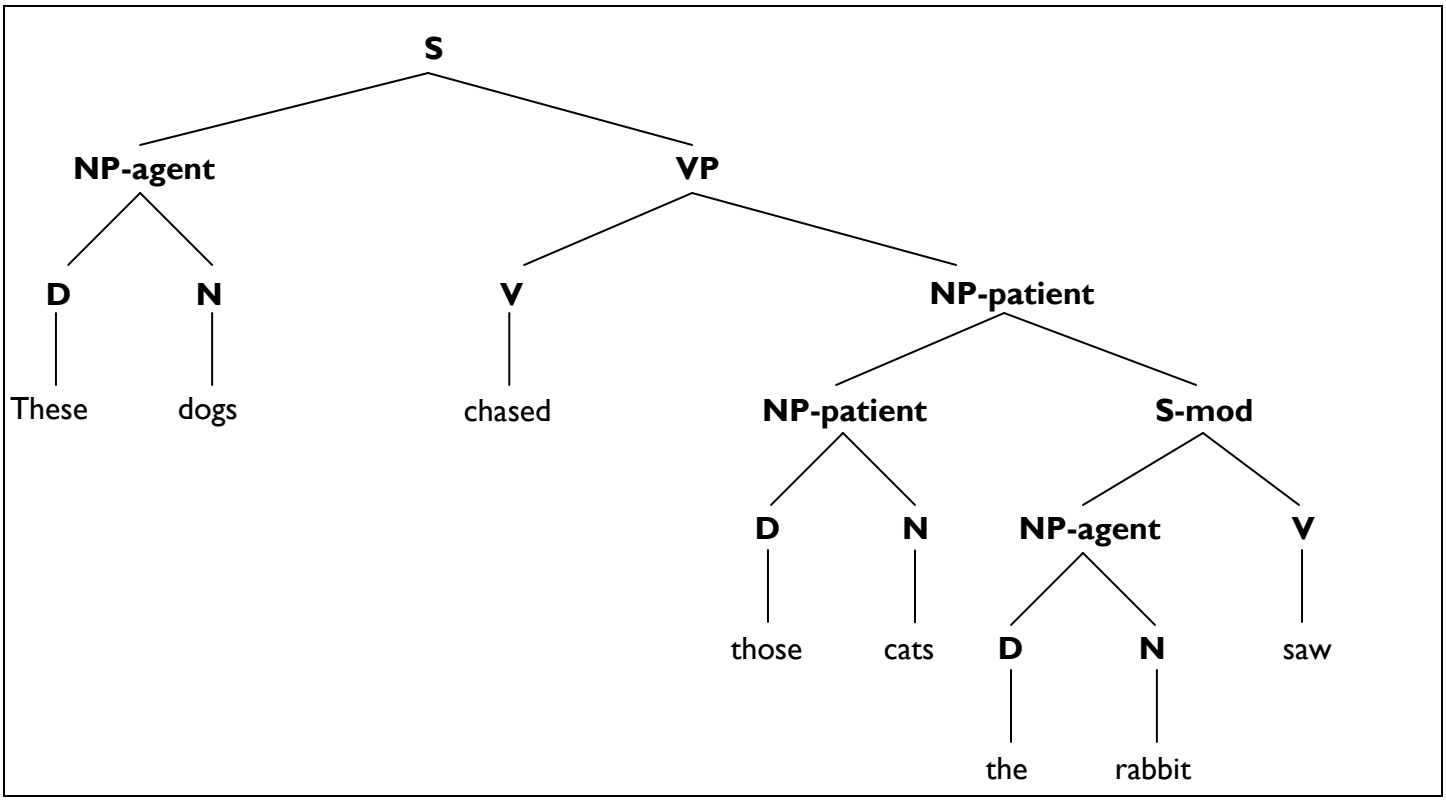


YOUR NAME: REGISTRATION #:

(5 points)

(I) Dogs and cats on trees (1/5)

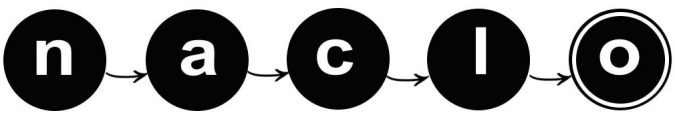
Linguists use diagrams called trees to represent the grouping of words within sentences. Here is a simple example from English:



The tree diagram shows that in the sentence "These dogs chased those cats the rabbit saw", *these* is most closely related to dogs, *those* most closely related to cats etc.

The abbreviations S, NP-agent, VP, etc. stand for different types of words or groups of words. These abbreviations and a few others we will use in this problem are spelled out here:

S: sentence
S-mod: sentence which functions as a modifier
NP-agent: noun phrase denoting the agent (initiator) of an action
NP-patient: noun phrase denoting the patient (undergoer) of an action
NP-location: noun phrase denoting the location of an action
N-agent: noun denoting the agent of an action
N-patient: noun denoting the patient of an action
N-location: noun denoting the location of an action
V: verb
V-mod: verb in a sentence which functions as modifier
VP: verb phrase

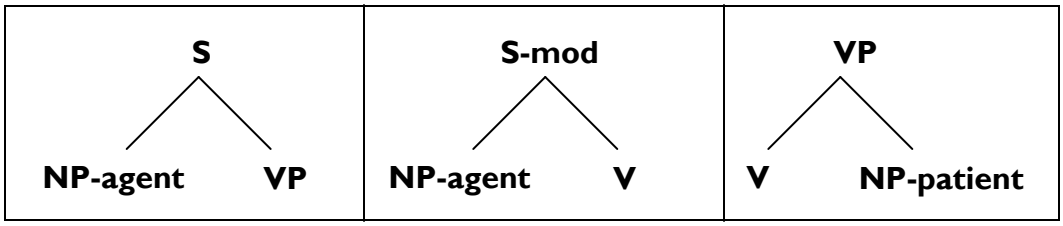


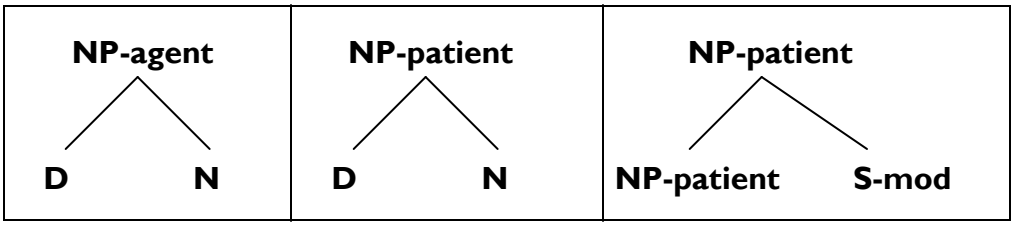
YOUR NAME: REGISTRATION #:

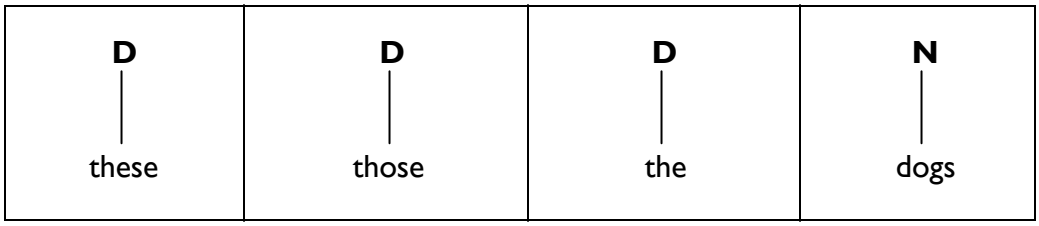
(I) Dogs and cats on trees (2/5)

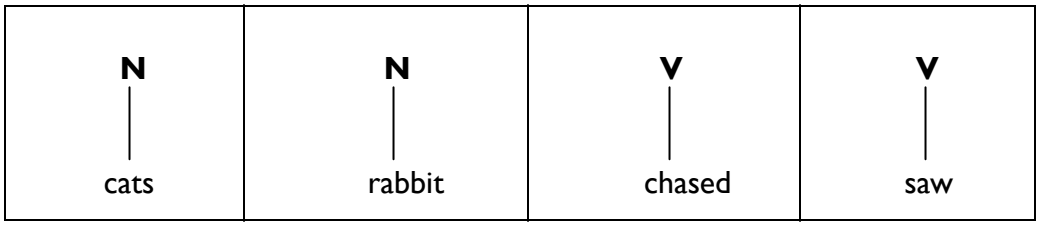
These labels give information about the part of speech of a word or group of words (e.g., noun, verb etc) as well as the role that that word or group of words plays in the meaning of the sentence.

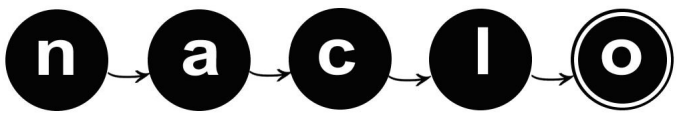
When working with trees, linguists write systems of rules (called 'grammars') which describe sets of trees. Each rule in the system is a building block. Any tree which can be constructed out of those building blocks is in the set of trees described by the grammar. For example, the tree given above for *These dogs chased those cats the rabbits saw.* requires the following building blocks or rules:











YOUR NAME: REGISTRATION #:

(I) Dogs and cats on trees (3/5)

I1 (3 points). Your first task is to translate the following sentences from Malayalam, a Dravidian language spoken by about 37 million people, primarily in India. There are two sources of information provided to you: a list of translations of the Malayalam words and a small grammar (set of rules) in the style above for Malayalam. Note that the set of abbreviations used in the Malayalam grammar is not the same as the set used in the English grammar. This is due to grammatical differences between the languages.

There is one twist, however. Some of these sentences are not actual Malayalam sentences. Use the grammar to figure out which ones they are.

For any sentence that is not an actual Malayalam sentence, you should not provide a translation. Write 'Not a Malayalam sentence' instead.

1. ആന സിംഹത്തെ ഓടിച്ചു
2. ആനയെ സിംഹം ഓടിച്ചു
3. ആനയെ സിംഹത്തെ ഓടിച്ചു
4. സിംഹം ഓടിച്ച ആനപ്പുറത്ത് ബാലൻ സഞ്ചരിച്ചു
5. ബാലൻ ഓടിച്ചു ആനപ്പുറത്ത് സിംഹം

YOUR NAME: REGISTRATION #:

(I) Dogs and cats on trees (4/5)

I2 (1 point). Draw the tree for any sentence that uses the V-mod rule. (You may use the English translations in place of the Malayalam words at the bottom of the tree.)

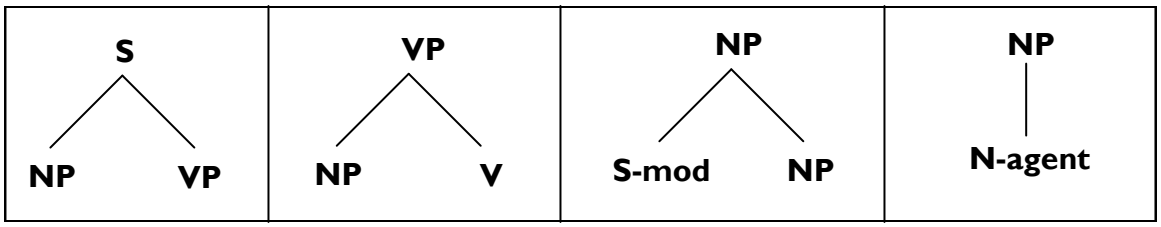
I3 (1 point). Explain what is wrong with the examples that are not actual sentences of Malayalam.

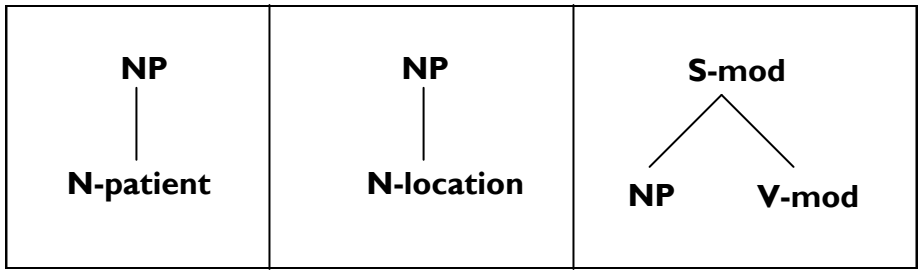
YOUR NAME: REGISTRATION #:

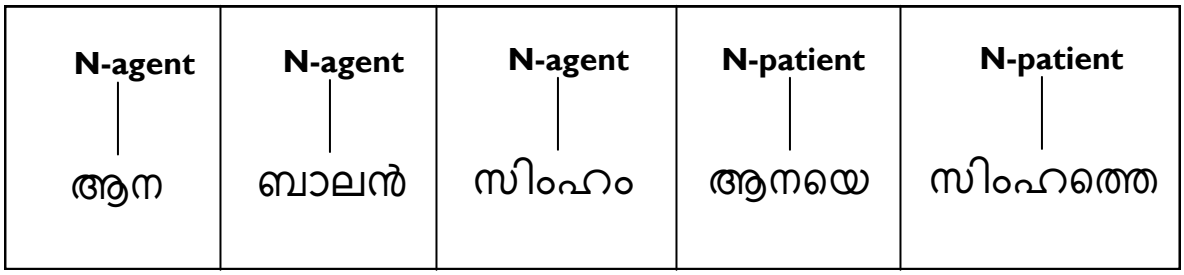
(I) Dogs and cats on trees (5/5)

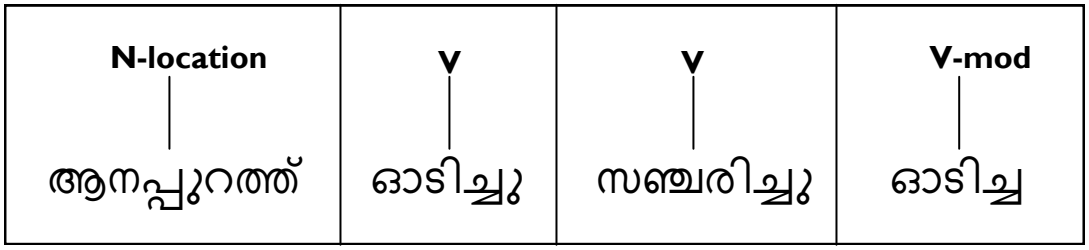
Word translations:

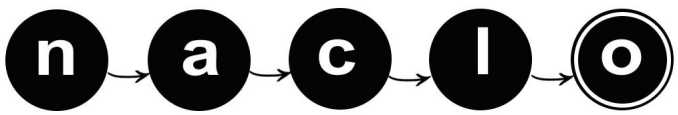
ആന	elephant	ബാലൻ	boy
ആനയെ	elephant	ഓടിച്ചു	chased
ആനപ്പുറത്ത്	elephant's back	ഓടിച്ച	chased
സിംഹം	lion	സഞ്ചരിച്ചു	rode











YOUR NAME: REGISTRATION #:

(10 points)

(J) Plains Cree (1/3)

Cree is the most widely spoken of the Canadian aboriginal languages, with about 117,000 people speaking one of its many varieties. Here are six words in Plains Cree (Nēhiyawēwin), a dialect spoken across much of the Western Canadian prairie and in parts of Minnesota, written using the Roman alphabet:

tehtapiwin	"chair"	*mistikwan*	"head"
iskwahtem	"door"	*tipahikan*	"hour"
sakahikan	"nail"	*astotin*	"hat"

J1 (1 point). Below are six related words, meaning "little hat", "little nail", "little door", "little head", "minute", and "little chair". Which means which?

cipahikanis _____

miscikwanis _____

cehcapiwinis _____

sakahikanis _____

ascocinis _____

iskwahcemis _____

YOUR NAME: REGISTRATION #:

(J) Plains Cree (2/3)

J2 (4 points). Although Cree can be written in the Roman alphabet, it is more frequently written in a writing system known as "Syllabics". This writing system has been adopted by speakers of other Canadian aboriginal languages as well; Inuktitut Syllabics are in wide use, and speakers of Ojibwe (Anishinaabemowin), Blackfoot, and Carrier (Dakelh) have also written their languages in Syllabics.

The twelve words provided above in the Roman alphabet are given below (in random order) in Syllabics. Write their Roman alphabet equivalents in the blanks next to each word.

a. ᑎᐸᐦᐃᑲᐤ ______________________

b. ᐊᐢᒍᒋᓂᐢ ______________________

c. ᓴᑲᐦᐃᑲᐤ ______________________

d. ᒥᐢᑎᑲᐧᐤ ______________________

e. ᒋᐸᐦᐃᑲᓂᐢ ______________________

f. ᐃᐢᑲᐧᐦᑫᒥᐢ ______________________

g. ᓴᑲᐦᐃᑲᓂᐢ ______________________

h. ᑌᐦᑕᐱᐃᐧᐤ ______________________

i. ᒥᐢᒋᑲᐧᓂᐢ ______________________

j. ᐃᐢᑲᐧᐦᑌᑦ ______________________

k. ᐊᐢᑐᑎᐤ ______________________

l. ᑫᐦᒐᐱᐃᐧᓂᐢ ______________________

Notes on pronunciation: When writing Cree in the Roman alphabet, the letter <c> represents the [ts] sound.

YOUR NAME: REGISTRATION #:

(J) Plains Cree (3/3)

J3 (5 points). Explain your answer.

YOUR NAME: REGISTRATION #:

(10 points)

(K) F u cn rd ths (1/4)

Abbreviations are hard. We are used to thinking of standard abbreviations like lb, CA, Mr or Blvd. But in fact people make up new abbreviations all the time, if they are under time pressure (e.g. instant messaging) or if they have severe space limitations (e.g. classified ads in a printed newspaper).

One place where you find lots of abbreviations is the notes taken by the overworked people who staff call centers. They have to record what was discussed, but they don't have the time to type everything out. So you often get things that look like this, from the logs of a call center run by a major telecommunications company:

cust rcvd lttr cncrng local srvc

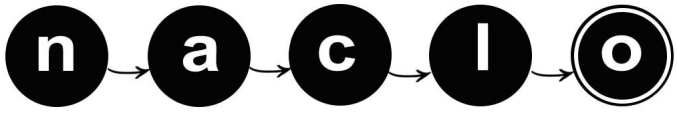
which of course is supposed to mean

customer received letter concerning local service

Let's say you are designing a computer program to try to do this kind of `normalization' automatically. You can't just have a fixed list of abbreviations: the set is pretty open ended. But what you can do is try to look at the whole corpus of data, and hope that someone somewhere has spelled out the complete words. So if for example I am looking at *rcvd lttr*, and somewhere else in the database someone has done us the favor of reporting on a different call, and used fully spelled phrase *received letter*, then we have a chance of guessing the expansion of *rcvd lttr*. That is, *rcvd* is a plausible abbreviation of *received*, *lttr* is a plausible abbreviation of *letter*, and the two occur together in the right order.

Of course, you know English, so you could have figured this out anyway. But the computer really doesn't. To the computer the problem looks as follows:

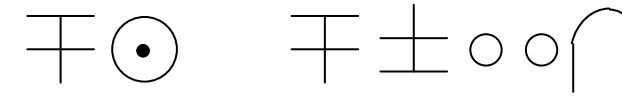
You have a bunch of abbreviated phrases (some of the words are not abbreviated, in fact), written in a bunch of symbols (remember the computer doesn't know English and to it, the strings are ultimately just a bunch of numbers anyway):

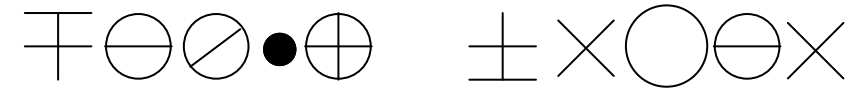


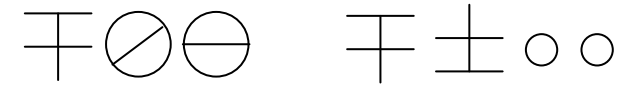
YOUR NAME: REGISTRATION #:

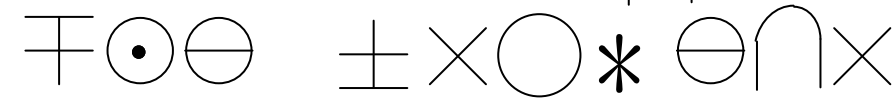
(K) F u cn rd ths (2/4)

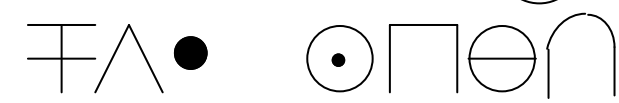
A. 

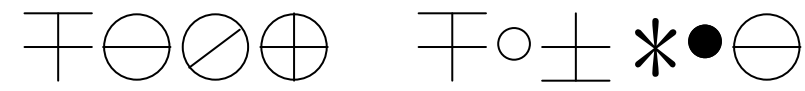
B. 

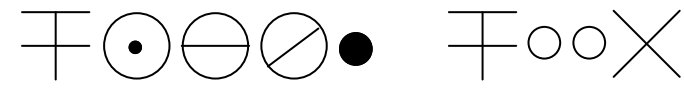
C. 

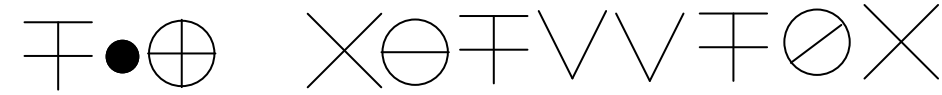
D. 

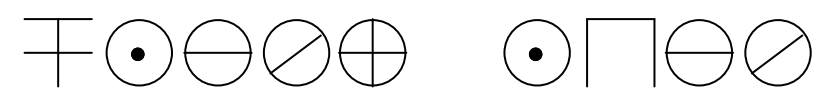
E. 

F. 

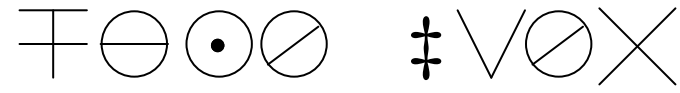
G. 

H. 

I. 

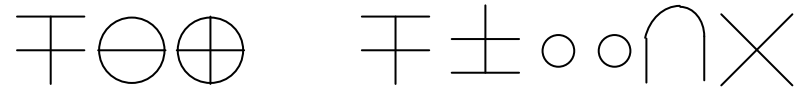
J. 

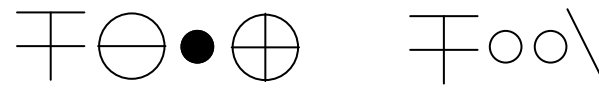
K. 

L. 

M. 

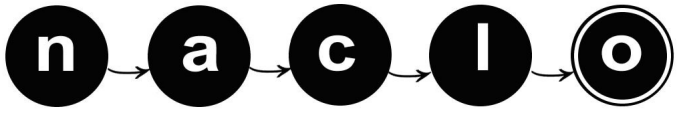
N. 

O. 

P. 

Q. 

R. 



YOUR NAME: REGISTRATION #:

(K) F u cn rd ths (3/4)

And you want to match with full phrases from elsewhere in the corpus:

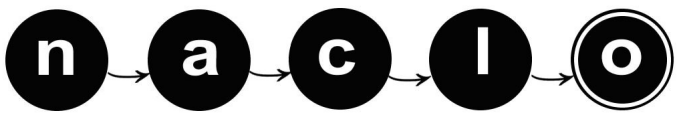
I	customer advised
II.	customer advised
III.	customer call
IV.	customer called
V.	customer called
VI.	customer called
VII.	customer called
VIII.	customer calling
IX.	customer calling
X.	customer care
XI.	customer claims
XII.	customer disconnected
XIII.	customer likes
XIV.	customer needs
XV.	customer request
XVI.	customer says
XVII.	customer understood
XVIII.	customer upset
XIX.	customer upset
XX.	customer wanted
XXI.	customer wants

There are two caveats:

1. When you are under time pressure, you make mistakes. There are actually three typos in the abbreviations—typos in that all the letters are there, but they are out of the expected order, and therefore are not strictly speaking reasonable abbreviations for the words.
2. There are three phrases that are not found in the abbreviations.

K1 (7 points). Match the encoded abbreviations from the previous page to the phrases above.

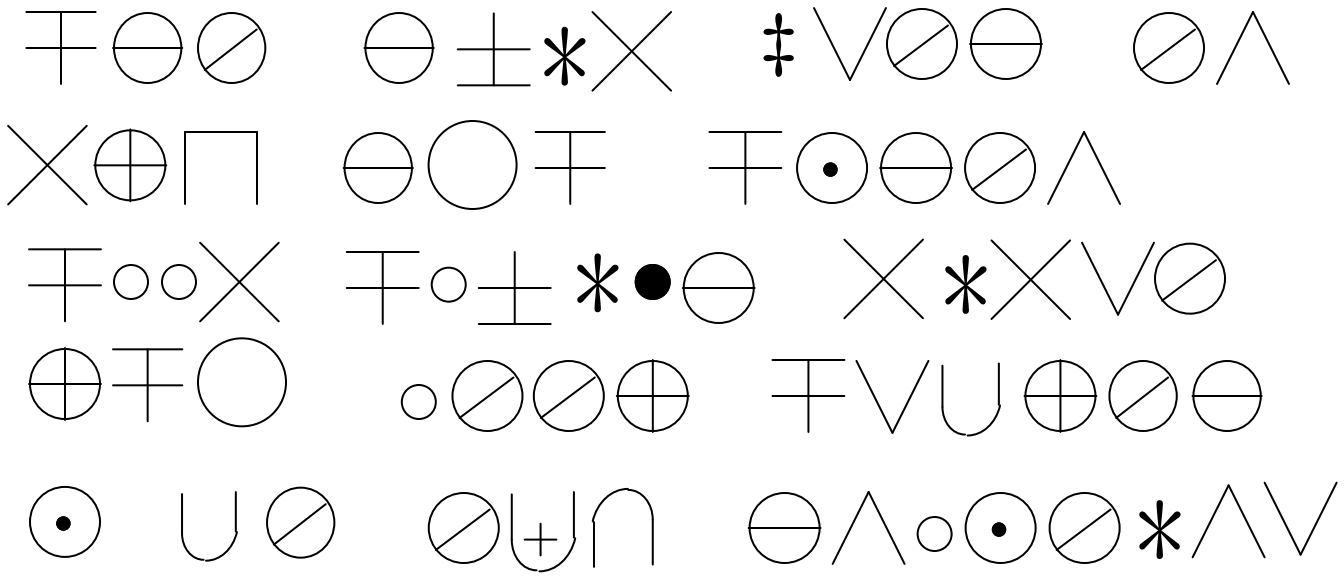
I.	IV.	VII.	X.	XIII.	XVI.	XIX.
II.	V.	VIII.	XI.	XIV.	XVII.	XX.
III.	VI.	IX.	XII.	XV.	XVIII.	XXI.

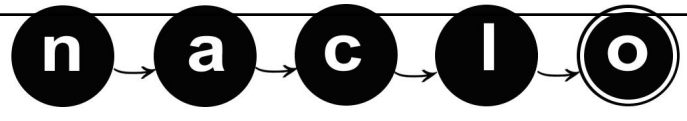


YOUR NAME: REGISTRATION #:

(K) F u cn rd ths (4/4)

K2 (3 points). Now, what phrase is abbreviated in the symbols below? Place your answer in the box at the bottom of the page.





YOUR NAME: REGISTRATION #:

(15 points)

(L) Real Money (1/2)

Languages often have special systems for counting specific sorts of objects – and money is no exception! Speakers of Cuzco Quechua, a widely-spoken indigenous language of Peru, employed a money-counting system still based on the old colonial Spanish and Peruvian coins the *real* and the *medio* (worth half a *real*).[1] Although Peru hasn't issued a coin based on the *real* in almost 150 years – the current Peruvian currency, the *nuevo sol* (notated *Sl.*), divides not into *reales* but into 100 *céntimos* – the counting system depicted below was still in use in recent times.

L1 (8 points). The following is a conversation between a shopkeeper *(qhatuq)* and a series of customers about the price of various tubers[2]. Knowing that the prices of potatoes, cassavas, and ocas at this market are Sl 0.05, Sl 0.10, and Sl 0.15 each (but not knowing which costs which), fill in the missing questions and answers. We've translated the first question as a guide.

Q: ¿Hayk'apaqmi huh lumu, huh papa, kinsa uqa ima?
("How much for one cassava, one potato, and three ocas?")
A: Pisqaralpaqmi.

Q. ¿Hayk'apaqmi iskay papa, huh lumu ima?
A. Iskaral miyunpaqmi.

Q. ¿Hayk'apaqmi suqta papa?
A. Kinsaralpaqmi.

Q. ¿Hayk'apaqmi iskay lumu, iskay uqa, huh papa ima?
A. Pisqaral miyunpaqmi.

Q. ¿Hayk'apaqmi pisqa uqa, kinsa papa ima?
A. Suqtaral miyunpaqmi.

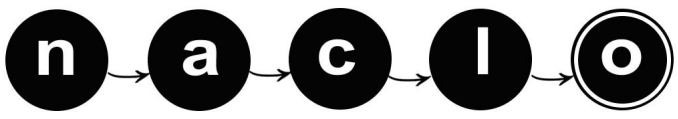
Q. ¿Hayk'apaqmi suqta uqa?
A. _____

Q. ¿Hayk'apaqmi iskay lumu, huh papa ima?
A. _____

Q. _____
A. Miyunpaqmi.

[1]Historical footnote: eight Spanish *reales* made up a *peso de a ocho* or *real de a ocho*. In English these were known as "pieces of eight" or "Spanish doubloons", and in parrot talk as "Awk! Pieces of Eight! Awk!".

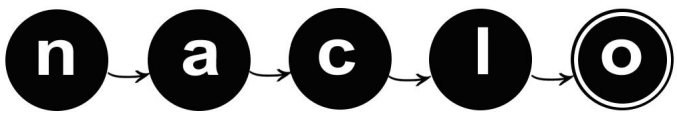
[2] Potatoes were first domesticated in South America, and the Quechua people have cultivated hundreds of species (and thousands of varieties) of potatoes and other tubers.



YOUR NAME: REGISTRATION #:

(L) Real Money (2/2)

L2 (7 points). Explain your answer.



YOUR NAME: REGISTRATION #:

(10 points)

(M) No smoke without fire (1/3)

Think about the meaning of the following sentence:

(1) The 2010 Winter Olympics were in Canada.

Assuming that we only know sentence 1 to be true, is sentence 2 necessarily true?

(2) The 2010 Winter Olympics were in Vancouver.

The answer is no. Assuming we only know sentence 1 to be true, the 2010 Winter Olympics could have taken place in any Canadian city, but not necessarily in Vancouver.

Now examine the relationship between sentences 3 and 4. Assuming sentence 3 is true, is sentence 4 now necessarily true?

(3) The 2010 Winter Olympics were in Vancouver.
(4) The 2010 Winter Olympics were in Canada.

Now the answer is yes. Since Vancouver is a Canadian city, any event which occurs in Vancouver necessarily occurs in Canada.

The logical relationship which holds between sentences 3 and 4 is called an *entailment*. In formal terms, sentence A entails sentence B if whenever A is true, B is necessarily true. The entailment relationship is typically represented graphically this way: A **||-** B.

Here are some more examples of the entailment relationship between sentences:

(5) Shaun White is a Winter Olympian **||-** Shaun White is an Olympian
(6) Shaun White is an Olympian **||-** Shaun White is an athlete
(7) Shaun White won a gold medal **||-** Someone won a gold medal

Notice that the entailment relationship must hold in the specified direction but will not necessarily hold in both directions. So, sentence 3 entails sentence 4 even though sentence 4 does not entail sentence 3.

YOUR NAME: REGISTRATION #:

(M) No smoke without fire (2/3)

Now examine the relationship between sentences 8 and 9.

(8) I did not see Shaun White win the gold medal in the 2010 Winter Olympics.
(9) Shaun White won the gold medal in the 2010 Winter Olympics.

Sentences 8 and 9 illustrate a relationship called *presupposition*. In this pair of sentences, the information presented in sentence 9 is what the speaker assumes (or presupposes) to be the case when uttering sentence 8. That is, to say *"I did not see Shaun White win the gold medal"* assumes the belief that Shaun White won a gold medal. In formal terms, sentence A presupposes sentence B if A not only implies B but also implies that the truth of B is somehow taken for granted. A presupposition of a sentence is thus part of the background against which its truth or falsity is judged. The presupposition relationship is typically represented graphically this way: A >> B

Here are some more examples of presuppositions (where the first sentence in each pair presupposes the second):

(10) I regret not seeing Shaun White's gold medal run >> Shaun White had a gold medal run
(11) Shaun White continues to rule the halfpipe >> Shaun White had been ruling the halfpipe
(12) Snowboarding is now an Olympic sport >> Snowboarding was once not an Olympic sport

YOUR NAME: REGISTRATION #:

(M) No smoke without fire (3/3)

M1. For any given pair of sentences, the entailment and presupposition relationships may or may not hold, together or separately.

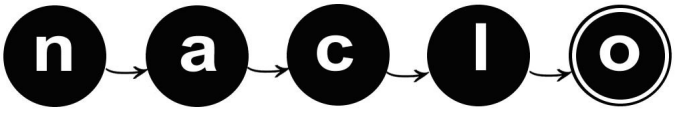
For each of the following possible combinations, your task is to provide one example of a pair of sentences with an explanation of your reasoning for proposing your pair of sentences as a valid and convincing example in each case.

a. A pair of sentences in which sentence A **neither entails nor presupposes** sentence B.

b. A pair of sentences in which sentence A **entails and presupposes** sentence B.

c. A pair of sentences in which sentence A **presupposes but does not entail** sentence B.

d. A pair of sentences in which sentence A **entails but does not presuppose** sentence B.



PART II

(problems NOP)

(after the break)

Do not work on this part before the break. You have two hours for this part.

YOUR NAME: REGISTRATION #:

(15 points)

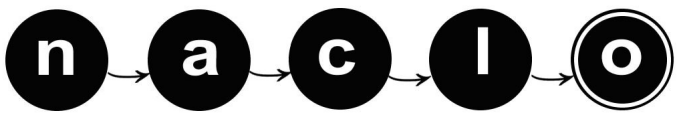
(N) Tale of Kieu (1/3)

Vietnamese is now written using a writing system consisting of Roman letters—just like English, but with lots of special markers or "diacritics" to show many distinctions between the sounds of many different vowels and six different tones. That writing system, called Quốc Ngữ, was developed by European missionaries, which is why Vietnamese is written with the same letters as European languages.

However, writing first came to Vietnam from China. At first, Vietnamese scholars actually wrote in Chinese, and because of the status of Chinese—as a language of government, literature, and culture—many Chinese words were borrowed into Vietnamese. When Vietnamese scholars started writing their own language, it was easy to see how to write these borrowed words: just use the same character that was used when writing Chinese. The following table gives a number of such characters used to write borrowings from Chinese into Vietnamese, their pronunciations in Vietnamese, and their approximate translation in English:

天 ***thiên*** sky; heaven; god	木 ***môc*** tree; lumber; wood; wooden
上 ***thuong*** top; highest; go up	見 ***kiên*** see, observe, perceive
工 ***gông*** labor; work; laborer	告 ***cáo*** tell; announce; inform; accuse
南 ***nam*** south	弄 ***lòng*** do; play or fiddle with; alley
病 ***bênh*** illness; sickness	豆 ***đâu*** peas; beans
沖 ***trong*** pour; infuse; wash out	沐 ***múc*** bathe; cleanse; wash
年 ***nên*** year; person's age	心 ***tâm*** heart; mind; intelligence; soul
糸 ***mich*** silk	皮 ***bì, bề*** skin; hide; fur
人 ***nhân*** man; human; mankind	

In order to write native Vietnamese words, however, these writers had to invent new characters. They did this by using a strategy that was already used, within the Chinese writing system, for creating new characters out of existing characters. In the Chinese writing system, new characters can be made by combining two or more simpler characters. These components provide hints regarding either the meaning of a character or its pronunciation. Components may be stacked on top of one another, place beside one another, or even placed so one surrounds another. While most of the characters given above are simple characters (with only one component) a few are complex characters (with more than one component). The writing system in which Chinese characters and components are used to write Vietnamese words is called Chữ Nôm. It was through the spread of this Chữ Nôm that Vietnamese literature finally came into its own. In fact, the Vietnamese national epic, the Tale of Kieu, was composed in Chữ Nôm.



YOUR NAME: REGISTRATION #:

(N) Tale of Kieu (2/3)

Here is a translation of the first six lines of the Tale of Kieu in English. Beneath it, but out of order, are the same lines in Vietnamese, both in Chữ Nôm (a-f) and in Quốc Ngữ (I-VI).

N1 (11 points). Show which lines from the two Vietnamese versions are translated by each line in the English version. We've given you one correspondence to get you started.

English	Chữ Nôm	Quốc Ngữ
1. A hundred years—in this life span on earth	______	______
2. talent and destiny are apt to feud.	______	______
3. You must go through a play of ebb and flow	______	______
4. and watch such things that make you sick at heart.	______	______
5. Is it so strange that losses balance gains?	______	VI
6. Blue Heaven's habit is to strike a rose from spite.	______	______

a. 𡦂才𡦂命窖羅恄饒

b. 邏之彼嗇私豐

c. 𡗶青慣退 𦟐紅打慳

d. 仍調𪾺𧡊罵忉疸𢚸

e. 𠪸戈没局𣷭㮣

f. 𤾓𢆥𥪞𡎝𠊛嗟

I. ***Trải qua một cuộc bể dâu***
II. ***Trăm năm trong cõi người ta***
III. ***Trời xanh quen thói má hồng đánh ghen***
IV. ***Những điều trông thấy mà đau đớn lòng***
V. ***Chữ tài chữ mệnh khéo là ghét nhau***
VI. ***Lạ gì bỉ sắc tư phong***

Translation by Huynh Sanh Thong

YOUR NAME: REGISTRATION #:

(N) Tale of Kieu (3/3)

N2 (4 points). Explain your answer.

YOUR NAME: REGISTRATION #:

(15 points)

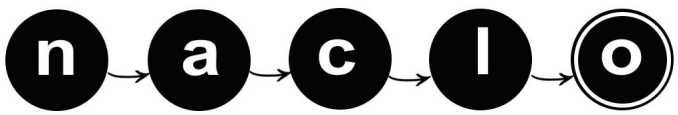
(O) Possessed in Vanuatu (1/3)

Vanuatu is a South Pacific country with 74 populated islands and more than 100 languages belonging to the Oceanic language family made up of languages spoken from Papua New Guinea to Hawaii to Easter Island. In Vanuatu, speakers of these languages have developed interesting ways of saying that something belongs to someone. You are invited to examine some examples from the language spoken on the island of Tanna.

Take a look at the examples of how possession is expressed in this language (given on the next page) and then answer the questions which follow.

NOTE: [ə] *represents a sound like the last sound of 'the' in 'the book'.*
[ŋ] *represents a sound like the last sound of 'hang'.*

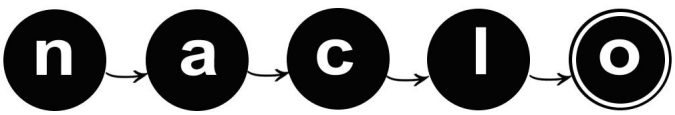
Map courtesy of Cindy Schneider, U. of New England (Australia)



YOUR NAME: REGISTRATION #:

(O) Possessed in Vanuatu (2/3)

	TANNA	ENGLISH TRANSLATION
1	*ralah neŋow*	their canoe
2	*rahan nasumien*	his garden
3	*raham nima*	your house
4	*nepikə kahaw*	rat's tail
5	*nəməm nəkawə*	your (=one person's) kava to drink
6	*netetamlaw*	your child (speaking to mother and father of child)
7	*niŋlaw nahwel*	their laplap pudding (a food) (for both of them)
8	*nenien raha enteni*	Tanna's speech (*enteni* 'earth' = Tanna)
9	*ratah naŋhatien*	our language (=yours (one person) and mine)
10	*narmen*	his image
11	*rahak nien*	my coconut (that I'm selling)
12	*rahak sot*	my shirt
13	*narfu tem*	man's belly
14	*neiwok mil*	my two female cousins
15	*pukah asoli*	big pig
16	*niŋək nien*	my coconut (for eating)
17	*nelkak*	my leg
18	*piam*	your (=one person) same sex sibling [*sibling* is a brother or sister]
19	*nisiməteliŋəm*	your (=one person) ear-wax
20	*narunien raha Tjotam*	Jotham's knowledge
21	*niŋlah kuri*	their dog (for eating)
22	*niŋən nawanien*	his food
23	*nepiken*	his tail
24	*ratalaw jow*	their turtle *(*belonging to both of them)
25	*rahak jerehi*	my lobster
26	*nisin*	his excrement
27	*nentowi jow*	turtle's neck
28	*nerow raha jow*	the turtle's spear
29	*nelka pukah*	pig's leg
30	*nakale naw mil*	The two edges of the knife OR The two knives' edges
31	*nisi kunget*	louse excrement
32	*nəmtalaw nəkawə, ian mwamnəm*	As for the kava (drink) belonging to you two, go and drink it!
33	*ratamlaw kuri ije?*	Where is your dog (belonging to the two of you)?
34	*niŋək kuri u, ojakawan*	My dog here, I'm going to eat (it).
35	*rahak nima takaku*	My house is small



YOUR NAME: REGISTRATION #:

(O) Possessed in Vanuatu (3/3)

O1 (6 points). Using the examples above as your model, translate each of these five expressions into the Tanna language.

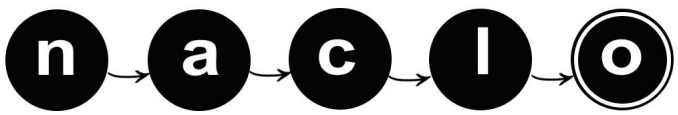
1. rat's ear	
2. my two dogs (that I own)	
3. their bellies	
4. a brother of those two (men)	
5. our child (= yours (1 person) and mine)	

O2 (6 points). Now see if you can translate these five expressions into the Tanna language.

1. Jawkelpi's house	
2. the pig's canoe	
3. My picture of you (=the one that I own that is an image of you)	
4. The house belonging to you two is big	
5. Where is my lobster (that I am going to eat)?	

O3 (3 points). There are several ways of saying 'their' in Tanna. List those found in the Tanna examples and explain the differences in meaning they express.

'Their' in Tanna	**Used when....**

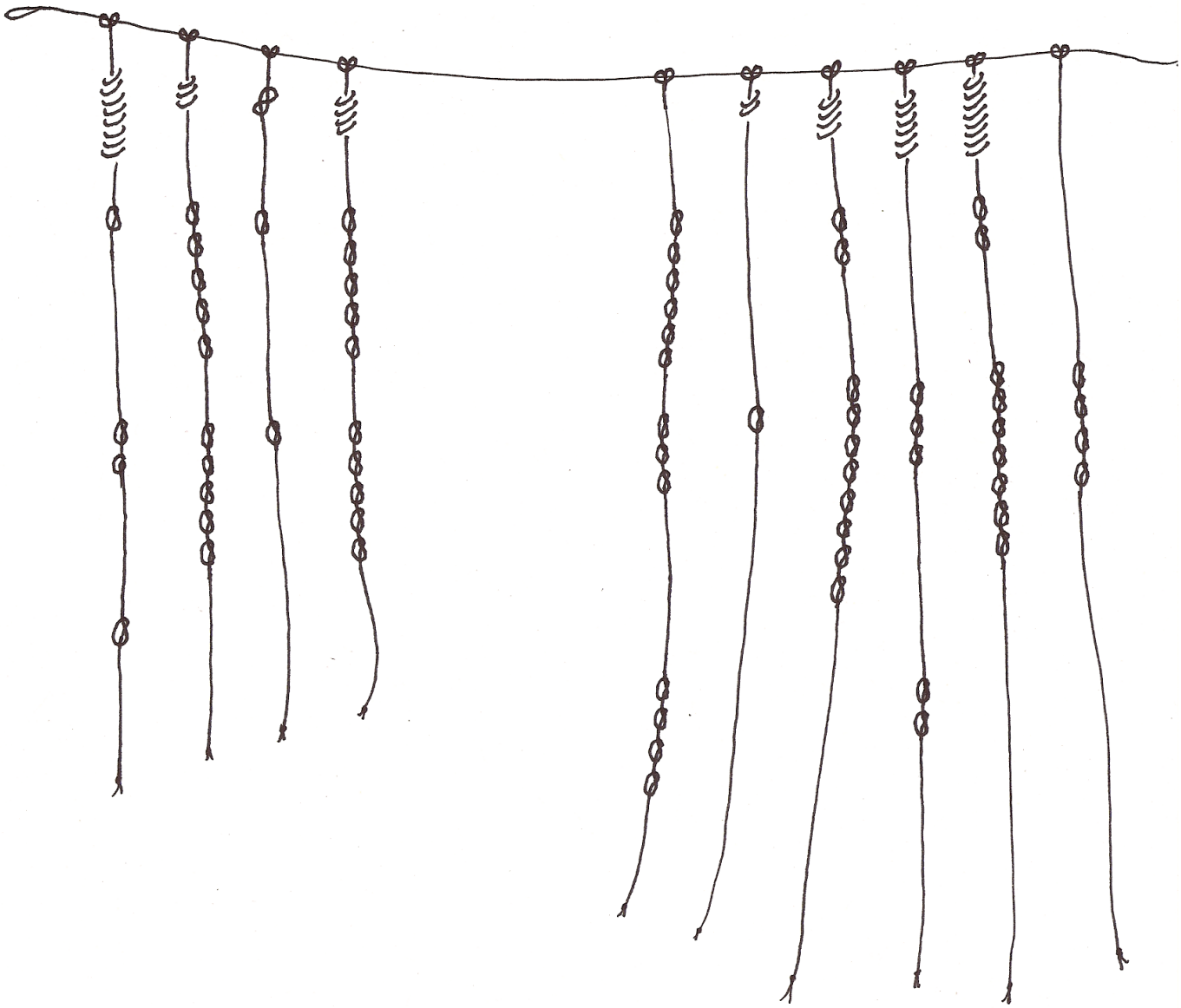


YOUR NAME: REGISTRATION #:

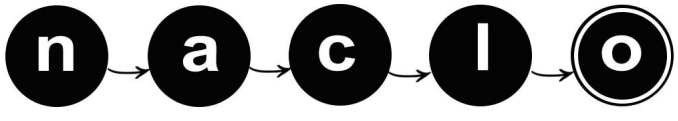
(10 points)

(P) Khipu (1/3)

Of all the major Bronze Age civilizations, there is only one for which we cannot (yet) find evidence of writing: the Inca Empire. Instead, a scribe-like class called the *khipukamayuq* kept records on collections of intricate knotted strings called *khipu*.



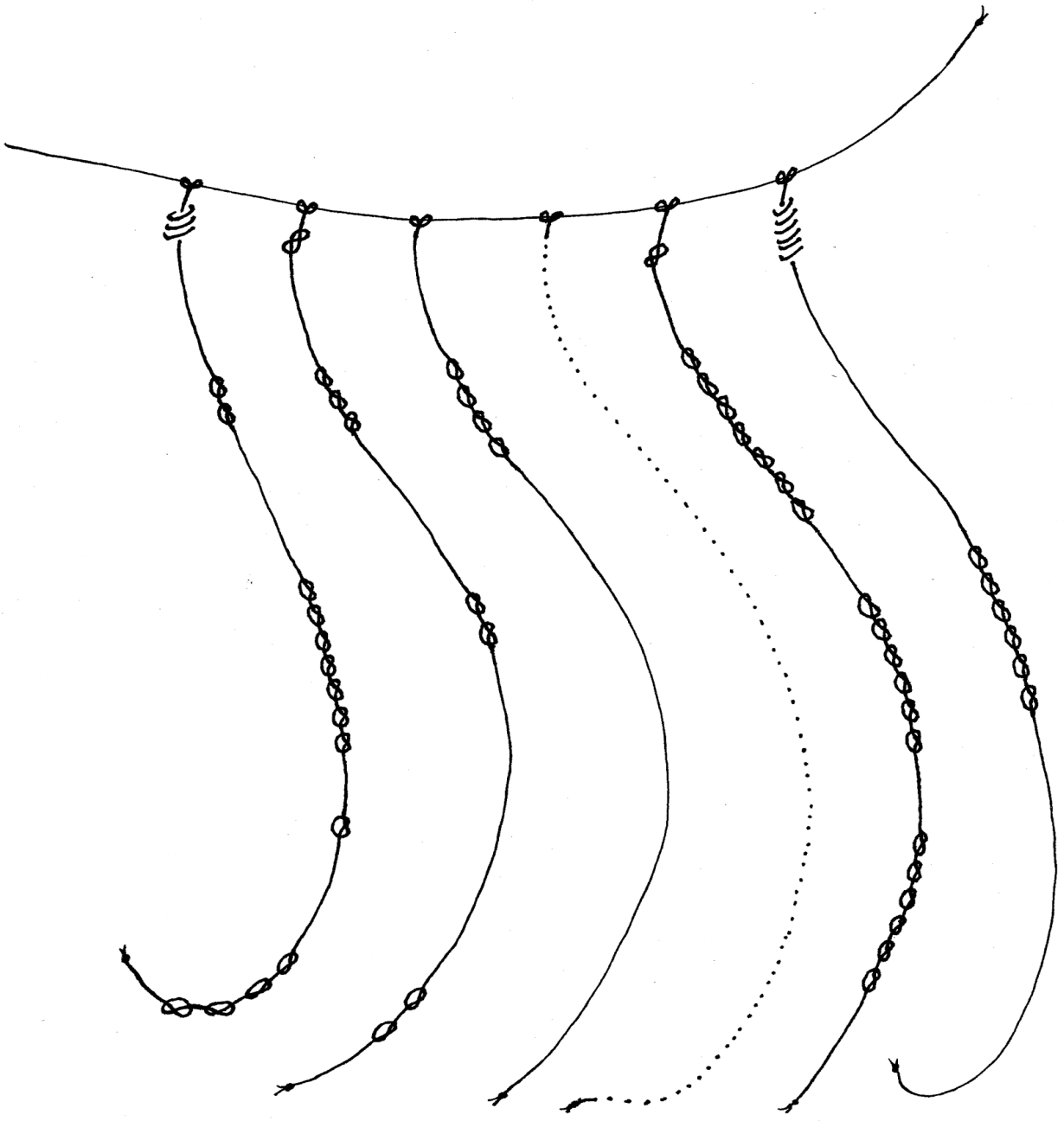
These knots are not random: there is a meaningful pattern that you can discover if you examine them closely. Each of these three groups of strings (two on this page, one on the next) is independent, but the same pattern is used in each. Patterns similar to this are frequent on real khipu, but only about 2/3 of the "khipu code" has been deciphered. The rest remains mysterious, and linguists, mathematicians, and computer scientists are still trying to uncover their secrets.



YOUR NAME: REGISTRATION #:

(P) Khipu (2/3)

P1 (6 points). This khipu has lost one of its strings. Can you figure out what was on it? Draw the missing string where the dotted line is.



YOUR NAME: REGISTRATION #:

(P) Khipu (3/3)

P2 (4 points). Explain your answer.

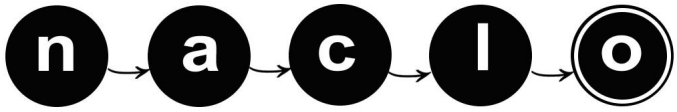
NACLO 2010 organizers

General chair:
Lori Levin, Carnegie Mellon University

Program committee chair:
Dragomir Radev, University of Michigan

Program committee:
Emily Bender, University of Washington
John Berman, Massachusetts Institute of Technology
Steven Bird, University of Melbourne
Aleka Blackwell, Middle Tennessee State University
Bozhidar Bozhanov, Bulgaria
Eric Breck, Cornell University
Ivan Derzhanski, Bulgarian Academy of Sciences
Jason Eisner, Johns Hopkins University
Dominique Estival, Australia
Eugene Fink, Carnegie Mellon University
Adam Hesterberg, Princeton University
Richard Hudson, University College London
Anatole Gershman, Carnegie Mellon University
Boris Iomdin, Russian Academy of Sciences
Rebecca Jacobs, University of Chicago
Joshua Katz, Princeton University
Mary Laughren, University of Queensland
Lori Levin, Carnegie Mellon University
Patrick Littell, University of British Columbia
Scott Mackie, University of British Columbia
Ruslan Mitkov, University of Wolverhampton
K P Mohanan, National University of Singapore
Helen Mukomel, Carnegie Mellon University
David Mortensen, University of Pittsburgh
Ani Nenkova, University of Pennsylvania
Barbara Partee, University of Massachusetts
James Pustejovsky, Brandeis University
Nathan Schneider, Carnegie Mellon University
Catherine Sheard, Yale University
Harold Somers, Dublin City University
Ekaterina Spriggs, Carnegie Mellon University
Richard Sproat, Oregon Health and Science University
Amy Troyani, Taylor Allderdice High School
Susanne Vejdomo, Eastern Michigan University
Richard Wicentowski, Swarthmore College
Xiaojin "Jerry" Zhu, University of Wisconsin

Administrative assistant:
Mary Jo Bensasi, Carnegie Mellon University



NACLO 2010 organizers (cont'd)

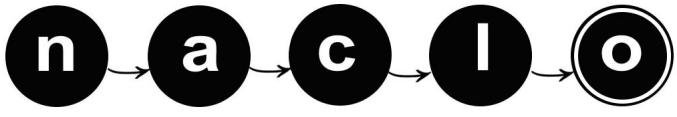
Problem credits:
Problem H: John Henderson with the assistance of
Sophie Crouch, University of Western Australia.
Based on Crouch (2008, 2009) and data from the MPI EVA Minangkabau corpus
Problem I: Emily Bender
Problem J: Patrick Littell and Julia Workman
Problem K: Richard Sproat
Problem L: Patrick Littell
Problem M: Aleka Blackwell
Problem N: David Mortensen and Patrick Littell
Problem O: Jane Simpson, University of Sydney and
Jeremy Hammond, Max Planck Institute for Psycholinguistics
Problem P: Patrick Littell and Erin Donnelly

Other members of the organizing committee:
Mary Jo Bensasi, Carnegie-Mellon University
Aleka Blackwell, Middle Tennessee State University
Josh Falk, Stanford University
Eugene Fink, Carnegie Mellon University
Katy Gann, Boeing
Adam Hesterberg, Princeton University
Lori Levin, Carnegie-Mellon University
Patrick Littell, University of British Columbia
James Pustejovsky, Brandeis University
Dragomir Radev, University of Michigan
Amy Troyani, Taylor Allderdice High School
Susanne Vejdemo, Eastern Michigan University
Michael White, Ohio State University
Julia Workman, University of Pittsburgh
Yilu Zhou, George Washington University

Web site and registration:
Adam Emerson, University of Michigan

US Team coaches:
Dragomir Radev, University of Michigan, head coach
Lori Levin, Carnegie Mellon University, coach
Adam Hesterberg, Princeton University, assistant coach

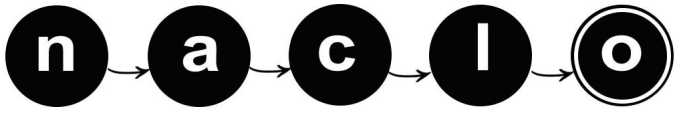
Canadian coordinator:
Patrick Littell, University of British Columbia



NACLO 2010 organizers (cont'd)

Contest site coordinators:

Brandeis University: James Pustejovsky
Carnegie Mellon University and University of Pittsburgh: Lori Levin and David Mortensen
Central Connecticut State University: Seunghun Lee
Columbia University: Kathy McKeown
Dalhousie University: Connie Adsett
Georgetown University: Graham Katz
Indiana University: Markus Dickinson and Sandra Kuebler
Johns Hopkins University: Mark Dredze
Middle Tennessee State University: Aleka Blackwell
Minnesota State University, Mankato: Rebecca Bates
Northeastern Illinois University: Judith Kaplan
Ohio State University: Michael White
Princeton University: Christiane Fellbaum and Adam Hesterberg
Queens College, CUNY: Heng Ji, Matt Huenerfauth, Andrew Rosenberg, Crystal Slaughter, Xiuyi Huang
San José State University: Roula Svorou
Simon Fraser University: John Alderete, Cliff Burgess, and Maite Taboada
Stanford University: Josh Falk, Spence Green, Dan Jurafsky, and Kyle Noe
University at Buffalo: Carl Alphonce
University of Great Falls: Porter Coggins
University of Illinois: Roxana Girju and Julia Hockenmaier
University of Illinois, Chicago: Barbara di Eugenio
University of Memphis: Vasile Rus
University of Michigan: Sally Thomason and Steve Abney
University of North Texas: Rada Mihalcea
University of Pennsylvania: Mitch Marcus
University of Rochester: Mary Swift
University of Southern California: David Chiang and Liang Huang
University of Texas at Dallas: Vincent Ng
University of Washington: Jim Hoard
University of Wisconsin: Nathanael Fillmore and Xiaojin Zhu
High school sites: Dragomir Radev

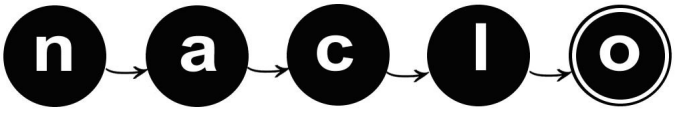


NACLO 2010 organizers (cont'd)

Student assistants and graders:
Marcus Berger, University of Michigan
Reed Blaylock, University of Michigan
Adam Emerson, University of Michigan
Amy Hemmeter, University of Michigan
Ridley Jones, University of Michigan
Nate LaFave, University of Michigan
Andrew Lamont, University of Michigan
Carrie Pichan, University of Michigan
David Ross, University of Michigan
Andrea Sexton, University of Michigan
Samuel Smolkin, University of Michigan
Laine Stranahan, University of Michigan

Grading software:
Adam Hesterberg, Princeton University

Graders:
Adam Hesterberg, Princeton University
Kapil Thadani, Columbia University
Laura Furst, Columbia University
Tracy Copeland, Georgetown University
David Mortensen, University of Pittsburgh
Lori Levin, Carnegie Mellon University
Eugene Fink, Carnegie Mellon University
Helen Mukomel, Carnegie Mellon Univeristy
David Elson, Columbia University
Josh Gordon, Columbia University
Sidharatha Nallu, Columbia University
Mohamed Altantawy, Columbia University



NACLO 2010 sponsors

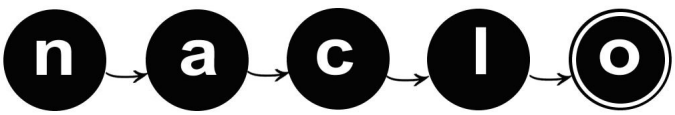
Booklet editors:
Nate LaFave, University of Michigan
Dragomir R. Radev, University of Michigan

Sponsorship chair:
James Pustejovsky, Brandeis University

Corporate, academic, and government sponsors
National Science Foundation
The North American Chapter of the Association for Computational Linguistics (NAACL)
Carnegie Mellon University's Language Technologies Institute
University of Michigan
Brandeis University

Special thanks to:
Tanya Korelsky, NSF
More than 70 high school teachers from 25 states and provinces

And many other individuals and organizations



NACLO 2010 sites

as well as more than 70 high schools throughout the USA and Canada