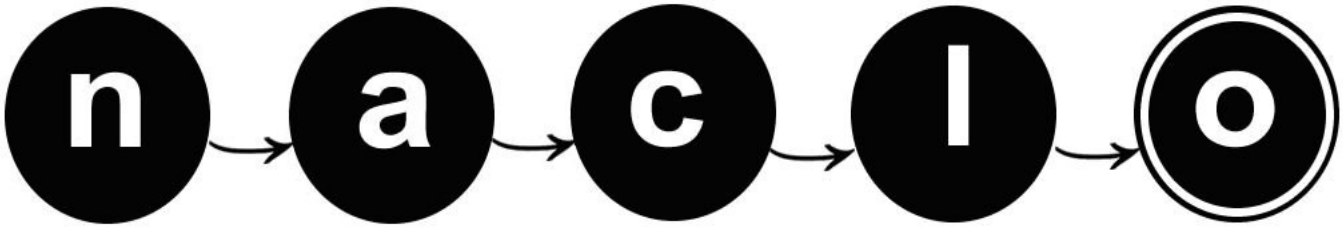
Do not scan this page



mozilla

UNIVERSITY OF MICHIGAN

MASCO

The Association for Computational Linguistics
North American Chapter

Carnegie Mellon

The Seventh Annual

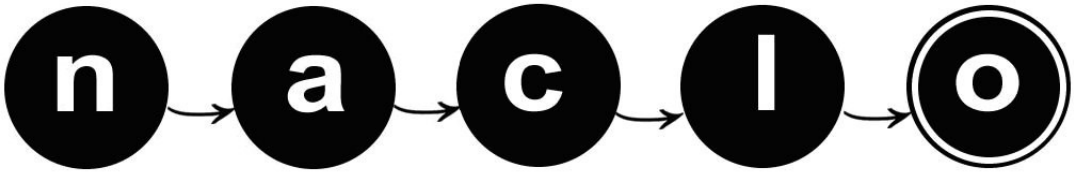
North American Computational Linguistics Olympiad

2013

www.naclo.cs.cmu.edu

Open Round

January 31, 2013



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

Contest Booklet

REGISTRATION NUMBER			

Name: _____

Contest Site: _____

Site ID: _____

City, State: _____

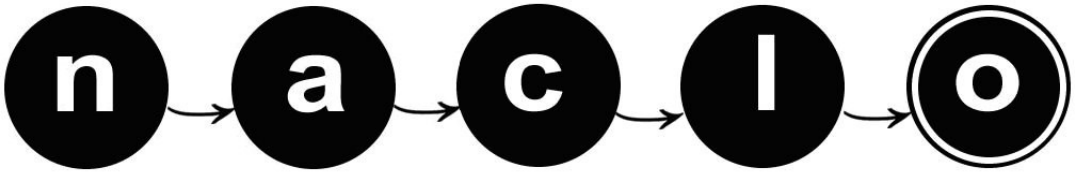
Grade: _____

Start Time: _____
End Time: _____

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 LATE MARCH.

Signature: _____



Welcome to the seventh 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 three hours long and includes eight problems, labeled A to H.
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 & 12.
5. Each problem is worth a specified number of points, with a total of 100 points. In this year's open round, no points will be given for explanations. Instead, make sure to fill out all the answer boxes properly.
6. 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.
7. Write your name and registration number on each page:
   Here is an example: Jessica Sawyer #850
8. The top 100 participants (approximately) across the continent in the open round will be invited to the second round.
9. 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 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. 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 shortly after the competition is finished with instructions on how to fill it out.
12. **DO NOT DISCUSS THE PROBLEMS UNTIL THEY HAVE BEEN POSTED ONLINE! THIS MAY BE SEVERAL WEEKS AFTER THE END OF THE CONTEST.**

Oh, and have fun!

Do not scan this page

NACLO 2013 Organizers

Program committee:

Morris Alper, Massachusetts Institute of Technology
Susan Barry, Manchester Metropolitan University
Jim Bauman, Center for Applied Linguistics
Aleka Blackwell, Middle Tennessee State University
Jordan Boyd-Graber, University of Maryland
Bozhidar Bozhanov, Consultant
Alan Chang, Princeton University
John deNero, Google and University of California, Berkeley
Ivan Derzhanski, Bulgarian Academy of Sciences
Jason Eisner, Johns Hopkins University
Dominique Estival, University of Western Sydney
Eugene Fink, Carnegie Mellon University
Anatole Gershman, Carnegie Mellon University
Marty Hackl, Massachusetts Institute of Technology
Adam Hesterberg, Massachusetts Institute of Technology
Dick Hudson, University College London
Boris Iomdin, Russian Academy of Sciences
Alexander Iriza, Princeton University
Ann Irvine, Johns Hopkins University
Rowan Jacobs, University of Chicago
Wesley Jones, University of Chicago
Mary Laughren, University of Queensland
Lori Levin, Carnegie Mellon University
Patrick Littell, University of British Columbia (co-chair)
Rachel McEnroe, University of Chicago
David Mortensen, University of Pittsburgh
James Pustejovsky, Brandeis University
Dragomir Radev, University of Michigan (co-chair)
Catherine Sheard, University of Oxford
Ben Sklaroff, University of California, Berkeley
Noah Smith, Carnegie Mellon University
Harold Somers, All-Ireland Linguistics Olympiad
Babette Verhoeven, Aquinas College

Problem Credits:

Problem A: Harold Somers
Problem B: Babette Verhoeven
Problem C: John DeNero
Problem D: Patrick Littell
Problem E: Babette Verhoeven
Problem F: Andrea Schalley and Patrick Littell
Problem G: Harold Somers
Problem H: Jason Eisner

NACLO 2013 Organizers (cont'd)

Organizing Committee:
Jim Bauman, Center for Applied Linguistics
Mary Jo Bensasi, Carnegie Mellon University
Aleka Blackwell, Middle Tennessee State University
Josh Falk, Stanford University
Eugene Fink, Carnegie Mellon University
Adam Hesterberg, MIT
Alex Iriza, Princeton University
Ann Irvine, Johns Hopkins University
Wesley Jones, University of Chicago
Lori Levin, Carnegie Mellon University (chair)
Patrick Littell, University of British Columbia
Rachel McEnroe, University of Chicago
Graham Morehead, University of Maine
David Mortensen, University of Pittsburgh
James Pustejovsky, Brandeis University
Amy Troyani, Taylor Allderdice High School
Susanne Vejdemo, University of Stockholm
Julia Workman, University of Montana
Yilu Zhou, George Washington University

Website and Registration:
Graham Morehead, University of Maine

US Team Coaches:
Dragomir Radev, University of Michigan (head coach)
Lori Levin, Carnegie Mellon University (coach)

Canadian Coordinator and Team Coach:
Patrick Littell: University of British Columbia

Do not scan this page

NACLO 2013 Organizers (cont'd)

Contest Site Coordinators:

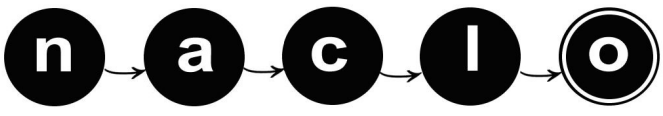
USA

Brandeis University: Rachel Basse, Jasenia Hartman, James Pustejovsky
Brigham Young University: Deryle Lonsdale
Carnegie Mellon University: Lori Levin, David Mortensen, Naomi Saphra
Central Connecticut State University: Matthew Ciscel, Seunghun Lee, Leyla Zidani-Eroglu
College of William and Mary: Ann Reed
Columbia University: Kathy McKeown, Esther Tavares
Cornell University: Abby Cohn, Sam Tilsen
Dartmouth College: Sravana Reddy
Georgetown University: Daniel Simonson
Johns Hopkins University: Mark Dredze
Massachusetts Institute of Technology: Adam Hesterberg, Chelsea Voss
Middle Tennessee State University: Aleka Blackwell
Minnesota State University, Mankato: Rebecca Bates, Dean Kelley
Northeastern Illinois University: Judith Kaplan-Weinger, John Boyle
Ohio State University: Micha Elsner, Julie McGory, Michael White
Princeton University: Alan Chang, Christiane Fellbaum, Alex Iriza
San Jose State University: Hahn Koo, Roula Svorou
Stanford University: Josh Falk
Stony Brook University: Yejin Choi, Kristen La Magna, Lori Repetti
Union College: Kristina Striegnitz, Nick Webb
University at Buffalo: Carl Alphonce
University of Alabama at Birmingham: Tamar Solorio
University of Colorado at Boulder: Silva Chang
University of Great Falls: Porter Coggins
University of Illinois at Urbana-Champaign: Julia Hockenmaier, D'Anne Winston
University of Maine: George Markowsky, Graham Morehead
University of Memphis: Vasile Rus
University of Michigan: Steve Abney, Sally Thomason
University of North Carolina, Charlotte: Wlodek Zadrozny
University of North Texas: Rada Mihalcea
University of Pennsylvania: Cheryl Hickey
University of Rochester: Mary Swift
University of Southern California: David Chiang, Zornitsa Kozareva
University of Texas: Stephen Wechsler
University of Texas at Dallas: Vincent Ng
University of Washington: Jim Hoard
University of Wisconsin: Nathanael Fillmore. Steve Lacy, Benjamin Snyder
University of Wisconsin, Milwaukee and Marquette University: Steven Hartman-Keise, Suzi Loosen, Hanyong Park, Gabriella Pinter, Joyce Tang Boyland
Western Michigan University: John Kapenga
Western Washington University: Kristin Denham

CANADA

Dalhousie University: Magdalena Jankowska, Vlado Keselj, Armin Sajadi
McGill University: Michael Wagner
Simon Fraser University: Maite Taboada, John Alderete
University of Lethbridge: Yllias Chali
University of Toronto: Qin Long

High school sites: Dragomir Radev



Do not scan this page

NACLO 2013 Organizers (cont'd)

Student Assistants:
Rachel Basse: Brandeis University
Tamar Forman-Gejrot: Brandeis University
Matthew Gardner: Carnegie Mellon University
Jasenia Hartman: Brandeis University
Kenneth Lai: Brandeis University
Andrew Lamont: Eastern Michigan University
Adrienne Reed: University of Michigan
Naomi Saphra: Carnegie Mellon University
Miriam Wong: Brandeis University

Booklet Editors:
Andrew Lamont: Eastern Michigan University
Dragomir Radev: University of Michigan

Sponsorship Chair:
James Pustejovsky: Brandeis University

Corporate, Academic, and Government Sponsors:
ACM Special Interest Group on Information Retrieval (SIGIR)
Brandeis University
Carnegie Mellon University
Masco Corporation
The Army Research Lab
The Linguistic Society of America
The Mozilla Foundation
The National Science Foundation
The North American Chapter of the Association for Computational Linguistics (NAACL)
The University of Michigan
The University of Pittsburgh
And many generous individual and corporate donors

Special thanks to:
Tanya Korelsky, NSF
And the hosts of the 90+ University Sites

YOUR NAME: REGISTRATION #

(A) Intuitive Inuit (1/3) [10 points]

The **Inuit** live in the Arctic regions of Canada and Greenland, and speak a language called **Inuktitut**. The Inuit used to be known as Eskimos, but this term is now considered insulting. The writing system used for the Inuktitut language is based on the one devised for writing Cree, a Native American language not related to Inuktitut. The writing system is highly regular and systematic, which should make your task all the easier.

One area of Canada where Inuit is spoken is called Nunavut and its capital is Iqaluit. Here is how these two words are written:

Nunavut	ᓄᓇᕗᑦ
Iqaluit	ᐃᖃᓗᐃᑦ

English has borrowed some words from the Inuit, such as 'igloo' which in Inuktitut is written as ᐃᖢ and is pronounced 'ihlu': the 'hl' is a lateral fricative, like the LL sound in Welsh, a bit like an 's' combined with an 'l'.[1] The 'hl' is treated as a single sound.

Your task is to look at the text below – it's the first part of the Universal Declaration of Human Rights – and to transcribe the underlined words into Roman letters on the basis of the first four words which have been done for you.

ᓯᓚᕐᔪᐊᕐᒥᐅᑦ	ᓇᓗᓇᐃᖅᓯᐅᑎᖕᒐ	ᑭᓇᒃᑯᑐᐃᓐᓇᐃᑦ	ᐱᔪᓐᓇᐅᑎᖕᒐᑦ
silarjuarmiut	nalunaiqsiutingga	kinakkutuinnait	pijunnautinggat

ᑎᓯᐱᕆ 10, 1948 ᑲᑎᒪᕐᔪᐊᕐᓂᕐᒥ ᑲᑐᔾᔨᔨᐅᔪᑦ ᓯᓚᕐᔪᐊᕐᒥᐅᓄᑦ ᐊᑐᓕᖅᑎᑦᑎᓯᒪᓂᐊᑦ ᐊᒻᒪᓗ ᐱᑖᕈᑎᖃᕐᓯᒪᓂᐅᓂᑯ ᓯᓚᕐᔪᐊᕐᒥᐅᑦ ᓇᓗᓇᐃᖅᓯᐅᑎᖕᒐᓂᒃ ᑭᓇᒃᑯᑐᐃᓐᓇᕐᓄᑦ ᐱᔪᓐᓇᐅᑎᖕᖏᓐᓂᒃ ᑕᒪᒃᑭᑦᑎᐊᕐᑐᒥᒃ ᑎᑎᕋᕐᓯᒪᐅᑎᖃᕐᑎᑦᓗᒍ ᓴᖅᑭᑕᐅᓯᒪᓂᑯᑦ ᑭᖑᓪᓖᒃᖢᑎᒃ ᒪᒃᐱᖅᑐᕋᓚᐅᕐᓯᒪᕗᑦ. ᑭᖑᓪᓚᓈᖕᖏᓐᓂ ᑕᒪᑐᒧᖕᒐ ᐅᓂᒃᑲᐅᓯᖕᒐᑕ ᐱᖁᔭᓚᐊᖑᒪᓂᑯ ᑲᑎᒪᕐᔪᐊᖅᑐᓂ ᖃᐃᖅᑯᔭᐅᓯᒪᓂᑰᑎᓪᓗᒋᑦ ᐃᓚᒋᔭᐅᔪᑦ ᓄᓇᖃᑎᒌᖕᖏᑦᑐᕐᔪᐊᓂᑦ ᐅᖃᓛ ᒪᒐᓚᐅᓚᐅᕐᓯᒪᖕᒪᑕ ᑎᑎᕋᕐᓯᒪᔪᖕᖏᓐᓂᒃ ᓇᓗᓇᐃᖅᓯᐅᑎᖕᒐ ᐊᒻᒪᓗ "ᐱᔾᔪᑎᖕᒐᓄᑦ ᓴᐃᒪᖅᑕᐅᓂᖅᖃᖅᑎᑕᐅᕐᑯᓪᓗᒍ, ᑕᑯᖅᑯᔾᔨᔾᔪᑕᐅᖃᓪᓗᒍᓗ, ᐅᖃᓛ ᒪᕐᓂᒃᑯᑦ ᐊᒻᒪᓗ ᐊᓯᖏᑎᒨᖓᔪᒃᑯᑦ ᐃᒃᔪᖅᑐᐃᖁᐅᔪᓂ, ᐅᕝᕙᓘᓐᓃᑦ ᓄᓇᑦᑎᐊᕐᒪᐅᓂ.

[1] This sound is also found in Zulu and the Mexican language Nahuatl, among others.

YOUR NAME: REGISTRATION #

(A) Intuitive Inuit (2/3)

Note that Inuktitut uses the same numerals as English, and has the same punctuation marks (only commas and fullstops in this text) as in English. There is no distinction between upper case (capitals) and lower case letters. A dot above a symbol indicates that the vowel is long, which can be represented in transcription by doubling the vowel letter. The sequence ‘rk’ is used for the [q] sound, a uvular plosive (like a ‘k’, but further back in the throat) and should be transcribed as a ‘q’. The sequence ‘ng’ should be thought of as a single sound (a velar nasal, as in the English word "sing"). The letter ‘j’ represents a ‘y’ sound. (as in "yank").

A1. Write your answers here:

A. ᑲᑐᔨᐳᐅᓪᓕᑦ															
B. ᐱᒋᑭᑎᖏᖅᓯᒪᓂᐅᓂᐃ															
C. ᑭᓐᔪᑦᓱᒃᓂᑎᒃ															
D. ᒪᒃᐱᖏᑐᕆᓕᖅᓯᒪᔭᑦ															
E. ᐱᖅᐃᓴᓕᐊᓐᔪᒪᓂᐃ															
F. ᐅᖏᓕᒻᒪᓕᐅᓪᓕᐅᖅᓯᒪᓐᒪᑕ															
G. ᐃᒃᓪᖏᑐᐃᐱᐅᓪᓂᓂ															
H. ᐅᓴᓰᒻᓄᓐᓂᑦ															

YOUR NAME: REGISTRATION #

(A) Intuitive Inuit (3/3)

A2. Using the information you have extracted from the text, how would you write the following words in the Inuktitut writing system? Let's start with two words for snow.[2] Enter one character in each cell.

A.	qanniq	'snow as it is falling'													
B.	aput	'snow on the ground'													
C.	mukluk	'sealskin boot'													
D.	umiaq	'canoe'													

A3. Finally, can you identify the English word borrowed from Inuktitut in A, and identify the place names in B and C? Enter the English word, one letter in each cell.

A.	ᖃᔭᖅ (a form of transport)								
B.	ᑲᓇᑕ								
C.	ᐊᓚᔅᑲ								

[2]You may have heard that the Inuit (or Eskimos) have lots of different words for 'snow'. In fact this is a kind of urban legend. Inuktitut has two main words for 'snow' although lots of shades of meaning can be expressed by adding endings – you will have noticed that Inuktitut words are very long.

YOUR NAME: REGISTRATION #

(B) A Case of Pali (1/2) [10 points]

Pali is a dead language, like Latin. It was a literary language related to Sanskrit, the ancestor of modern languages spoken in Northern India, such as Hindi. Pali was first written down around 100 BCE in Sri Lanka by Buddhist monks to preserve the teachings of the Buddha. Pali used to be written in the Brāmī script, but it is also written in the Roman alphabet (which we'll be using here). Pali is still used by Buddhist monks and scholars (just as Latin is still used in the Vatican by Catholic priests and theologians).

Pali is a highly inflected language, which means that the main words such as nouns and verbs get a range of endings (called "suffixes") or beginnings (called "prefixes") attached to make it clear what role the word is playing in the sentence.

English also has some inflections, just not as many as Pali. Here are some examples of English inflections:

- *house – houses*. The *–s* added to the end of a noun like "house" indicates that there is more than one you are talking about.
- *John – John's coat*. The *'s* is used after a noun to indicate possession.
- *I walk. You walked. He walks. She has been walking*. The suffixes added to the verb "walk" give you sense of a different person doing the walking, or the walking taking place at a different time – the past as opposed to the present.

Pali has different consonant and vowel sounds, which explains the use of diacritics (special symbols) on particular letters. These do not matter for solving this puzzle. Also, Pali texts do not use capital letters or punctuation.

Here are some sentences in Pali with their English translations:

Pali	**English Translation**
mahāmatto nisīdati	The minister sits down.
mahāmattaṃ upasaṃkamanti	They visit the minister.
samaṇo tathāgato hoti	The philosopher is enlightened.
samaṇe atthaṃ pucchanti	They ask the philosophers the meaning.
upāsako pucchati	The disciple asks.
loko mahāmattassa	The minister's world.

YOUR NAME: REGISTRATION #

(B) A Case of Pali (2/2)

B1. Translate the following English sentences into Pali:

1. The minister asks the philosophers.													
2. The philosopher sits down.													
3. They sit down.													

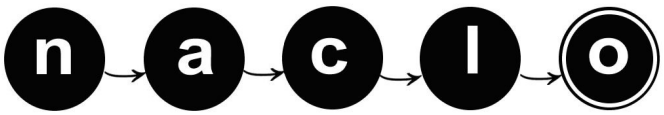
B2. Translate the following into English, using the vocabulary given here in its dictionary form (which is the same as the subject form, without any suffixes):

Pali	**English Translation**
rājo	king
devo	god
gāmo	village

1. rājo nisīdati															
2. rājo gāmassa devo hoti															

B3. Translate the following into Pali, entering one letter in each box, ignoring the diacritics:

1. The minister asks the kings.													
2. The lay disciple's village.													
3. The meaning of the world is god.													



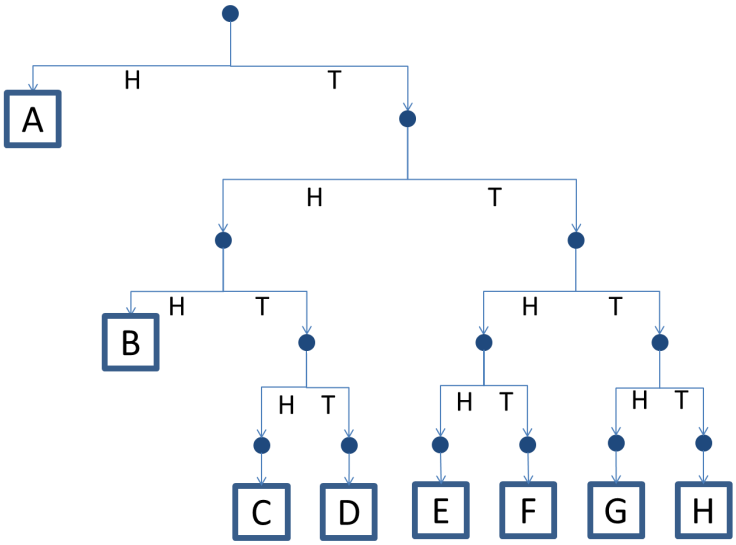
YOUR NAME: REGISTRATION #

(C) The Heads and Tails of Huffman (1/2) [10 points]

When Deb gets mad, she sends her friend Ahab encoded messages using lines of pennies, each of which is either heads up (H) or tails up (T). Example:

THHHTHTT HTTTTHTHH

Deb also sends a decoding tree, which indicates how to read the message encoded by the pennies. A decoding tree starts with two branches, marked (H)eads and (T)ails. Each branch either leads to a letter in the message or another decoding tree. This type of tree is called a Huffman encoding tree, based on the name of its inventor.

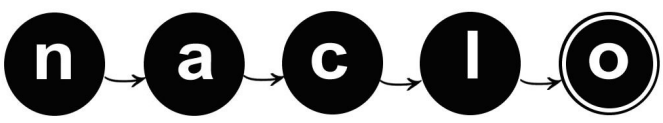


Pennies are read from left to right, and each penny indicates which branch of the decoding tree to follow. Whenever a letter is reached, the next letter is decoded starting back at the top of the decoding tree. For example, the message above reads "BAD AHAB", where individual letters are placed in boxes below:

B	A	D	A	H	A	B
THH	H	THTT	H	TTTT	H	THH

C1. Decode the following messages using the decoding tree shown above:

A.	TTTTTTHHTTHTTTHHTHTTTHHTTHHTTHHTTHT										
B.	HTHTHHTTTHTTHHTHTTTHHTTHHTHTT										



YOUR NAME: REGISTRATION #

(C) The Heads and Tails of Huffman (2/2)

C2. The following English word from Deb is missing a penny somewhere in the middle. Mark the location and orientation (heads or tails) of the missing penny and decode the message.

TTTTTTHHTHTTTTHHTHTT																	

Location of the new penny (counting from the left): ☐ orientation: ☐

C3. Deb doesn't want to spend all of her allowance on messages. Design an encoding tree and write the corresponding encoding for each letter below, such that the encoding requires as few pennies as possible, but still correctly encodes the messages. Assume that the message only contains the letters in the example (e.g., MISP in the first example and ABCDR in the second one). In a Huffman encoding, the encoding of a letter cannot begin with the encoding of another letter. So, for instance, if some letter is encoded as H, then another one cannot be encoded as HT. In fact, if some letter is encoded as H, then the encoding for any other letter must start with T.

The two examples below are independent. There may be more than one optimal encoding per example. You only need to show one of them.

MISSISSIPPI

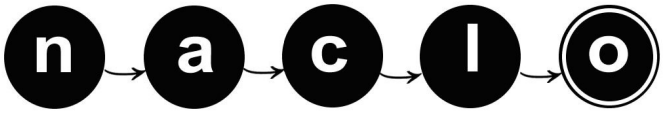
Letter	Code
I	
M	
P	
S	

Total number of coins: _____

ABRACADABRA

Letter	Code
A	
B	
C	
D	
R	

Total number of coins: _____



Do not scan this page

YOUR NAME: REGISTRATION #

(D) Kwak'wala Word Search (1/2) [20 points]

Kwak'wala is the language of the *Kwakwaka'wakw* people, one of Canada's many aboriginal nations. It is spoken on and around Vancouver Island, but only a few hundred fluent speakers remain. We've hidden 30 Kwak'wala words in the puzzle on the following page, horizontally, vertically, or diagonally, but we've only given you 10 of them. Your challenge is to find the remaining 20 words and match them to their meanings.

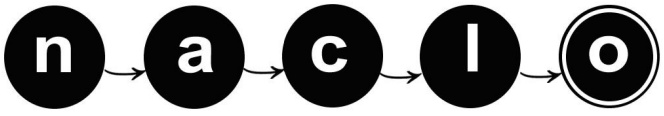
K'W	A	KW	N	I	U	A	KW	A	'M	K'	XW
G	X	T	A	Y	A	Y	A	G	A	Y	U
Y	I	DL	A	X	P'	U	G	D	TL	A	Y
'L	GW	D	U	K'	I	KW	A	DL	I	K'	A
B	A	K	A	Y	U	GW	Y	T'S	DZ	I	B
K'	T'S	'M	A	KW	A	L	A	S	K'	N	I
K	I	DL	A	T'S	I	B	XW	Y	T'L	U	L
Y	'M	K'W	I	P'	A	U	A	T'	U	XW	I
KW	G	A	I	T'S	N	Y	S	'L	B	U	XW
X	TS	N	L	I	P	A	A	A	TS	K	U
D	U	K'	T'	T'S	L	GW	I	G	XW	A	N
XW	S	A	L	A	K	A	Y	A	A	'M	I
I	K'	D	P	G	GW	'M	B	KW	K'W	T'S	T'L
A	K	A	'L	M	U	I	I	A	A	U	I
A	T'S	KW	DZ	A	'L	X	TS	P	T'	K'	K
KW	A	K	I	TL	A	GW	A	K	M	A	TL

YOUR NAME: REGISTRATION #

(D) Kwak'wala Word Search (2/2)

Notes: Some two-character sequences, like *tl*, *dl*, *kw*, *gw*, and *ts*, are treated as single letters in the Kwak'wala alphabet. a represents an "uh" sound. *k* and *g* are pronounced like "k" and hard "g", but with the tongue further back in the mouth. *x* is pronounced like the "h" in "human", and *x* like the "ch" in "Bach". An apostrophe indicates that the sound is pronounced with increased pressure at the back of the throat.
Enter one character as it appears in the word search per cell.

an iron										
berry cakes	'L	A	G	A	KW					
bowl for candlefish oil										
broom	X	I	GW	A	Y	U				
deck of cards	L	I	B	A	Y	U				
dustpan										
envelope										
expert card player										
expert knitter										
fisherman	K	I	T'L	I	N	U	X W			
fishing boat										
food for dipping in oil	T'S	A	P	A	L	A	S			
knitting basket	Y	A	G	A	T'S	I				
knitting needles										
letter	K'	A	D	A	KW					
pen or pencil										
something knitted, such as a sweater										
to be proud, to be a snob	TL	A	M	K	A					
to catch fish with a net										
to dip food in candlefish oil										
to iron something	'M	A	KW	A						
to knit										
to make berry cakes										
to play cards										
to sweep										
to write	K'	A	T	A						
tourist boat, cruise ship, ferry										
wool										
wrinkled clothes										
writer										



YOUR NAME: REGISTRATION #

(E) Shaw Business (1/2) [10 points]

The author George Bernard Shaw (author of *Pygmalion*,[1] which was later adapted as the musical *My Fair Lady*) saw the use of the Latin (or Roman) alphabet for English as a waste of time – the alphabet simply was not suited to write English.

An infamous example of English pronunciation being at odds with its use of the Roman alphabet is the sequence –*ough*. Consider the different pronunciations of *ough* in the following words: *rough*, *through*, *hiccough*, *though*, and *bough*. This should give you an idea why a spelling reform may be needed.

Shaw left money in his will to the inventor of a new (and better) script for English. Kingsley Read, among many others, entered the competition, and his alphabet was chosen as the best response to Shaw's challenge in 1958. Read named his invention the "Shavian Alphabet" in honor of Shaw.

Some of the rules of Shavian:

- The majority of its characters simply represent an individual English sound. However, a few can be used as abbreviations for a word.
- Shavian is based on a Rhotic English accent – an accent in which all the r symbols are pronounced (such as American English where the word "rarer" is pronounced with 3 r-sounds).
- There are no capital "letters" in Shavian.

E1. On the left below are some phrases rendered in Shavian, while on the right are their transliterations in the Roman alphabet, except they have been reordered. Work out which Roman transliterations match the Shavian phrases.

	Shavian alphabet		**Roman Alphabet**
1.	𐑞 𐑒𐑨𐑑 𐑕𐑤𐑧𐑐𐑑	A.	this is Shavian
2.	𐑢𐑰 𐑣𐑨𐑝 𐑒𐑨𐑑𐑕	B.	the cat slept
3.	𐑑 𐑤𐑻𐑯	C.	to learn
4.	𐑞𐑦𐑕 𐑦𐑟 ·𐑖𐑱𐑝𐑾𐑯	D.	we have cats
5.	𐑓 𐑧𐑝𐑼	E.	for ever

[1] *Pygmalion* is about a linguistics professor trying to correct the pronunciation of a Cockney flower seller called Eliza (Cockney being a dialect of British English used by Dick van Dyke in *Mary Poppins*). One of the first artificial intelligence programs, a chatbot, was named Eliza, after the character in *Pygmalion*.

YOUR NAME: REGISTRATION #

(E) Shaw Business (2/2)

E2. Using your knowledge of Shavian so far, transliterate the following English phrases into Shavian alphabet. Instead of writing out the symbols yourself, refer to the table provided (e.g. for 𐑒 enter 3 in one cell):

Eve											
Ian											
turn left to sit											
sleep for Steve											

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
𐑐	𐑑	𐑒	𐑓	𐑔	𐑕	𐑖	𐑗	𐑘	𐑙	𐑚	𐑛	𐑜	𐑝	𐑞	𐑟	𐑠	𐑡	𐑢	𐑣
21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40
𐑤	𐑥	𐑦	𐑧	𐑨	𐑩	𐑪	𐑫	𐑬	𐑭	𐑮	𐑯	𐑰	𐑱	𐑲	𐑳	𐑴	𐑵	𐑶	𐑷
41	42	43	44	45	46	47	48	49											
𐑸	𐑹	𐑺	𐑻	𐑼	𐑽	𐑾	𐑿	𐑀											

You may have noticed that Shavian has characters that are like the Roman alphabet's ascenders (letters that stick out in an upwards direction from the line of writing such as "f" and "h") and descenders (letters that go below the line of writing such as "g"). Ascenders are known as "tall characters" and descenders as "deep characters" in Shavian. Some tall and deep characters make deliberate pairs, such as:

𐑕 and 𐑟

𐑡 and 𐑓

E3. Work out what sounds the symbols in the pairings above represent. How would you write this sound using the Roman alphabet?

𐑕 ☐ 𐑟 ☐ 𐑡 ☐ 𐑓 ☐

E4. What is the Shavian symbol for "b" (enter the code): ☐

YOUR NAME: REGISTRATION #

(F) Grammar Rules! (1/3) [10 points]

One way for computers to understand language is by parsing sentences to figure out the role of each word. A context free grammar (CFG) (also called phrase structure grammar) is a set of rules for forming sentences. Only sentences that can be generated using such a set of rules are then deemed grammatically correct and 'well-formed'. Computer scientists and linguists use CFGs to define and parse languages, where a "language" is defined as any and all sentences that a given CFG can generate. S is the starting symbol for each sentence.

The following rules make up a simple CFG:

S → N V N → children N → squirrels V → sing V → eat

Each rule says that the element to the left of the arrow can be expanded into the elements to the right of the arrow. By repeatedly replacing symbols, this CFG can expand the symbol S into "squirrels sing", "children sing", "squirrels eat", and "children eat". It cannot, however, generate "children eat squirrels" or "squirrels eat children" or just "children" – you can see that there is no possible sequence of replacements that turns S into any of these.

The following is another simple CFG. The rules have been numbered for your convenience, but the numbers are not part of the rules.

1. S → NP VP	2. VP → VP PP	3. PP → P	4. IV → runs
5. NP → N	6. VP → VP CONJ VP	7. PP → P NP	8. C → that
9. NP → D N	10. N → squirrel	11. TV → chases	12. P → in
13. NP → NP CONJ NP	14. N → he	15. TV → eats	16. P → away
17. VP → IV	18. N → John	19. TV → catches	20. CONJ → and
21. VP → IV PP	22. N → Mary	23. TV → tells	24. D → the
25. VP → TV NP	26. N → dog	27. TV → sees	
28. VP → TV C S	29. N → tree	30. IV → sits	

YOUR NAME: REGISTRATION #

(F) Grammar Rules! (2/3)

F1. Here is a simple story. Several of the following sentences are, according to the above CFG, not well formed, meaning they cannot be derived from S by repeated substitution of symbols. List the sentences that the CFG above *can* generate in the box below; ignore the periods.

A. John sees the dog and Mary sees the dog.

B. The dog sees John and Mary.

C. The dog sees a squirrel.

D. The squirrel sits in the tree.

E. That squirrel sees the dog.

F. The squirrel was seen by the dog.

G. The dog runs.

H. The squirrel in the tree runs.

I. The dog chases the squirrel and eats the squirrel.

J. The dog eats.

K. John sees that the dog eats the squirrel.

L. John tells Mary that the dog eats the squirrel.

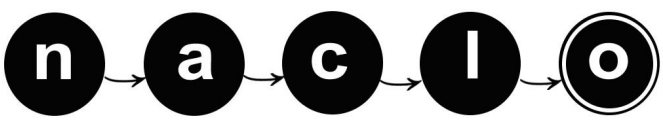
M. The dog sees that John sees that he eats the squirrel.

N. And the dog runs away.

O. Mary and John chase the dog.

P. John chases and catches the dog.

Q. John eats dog.



YOUR NAME: REGISTRATION #

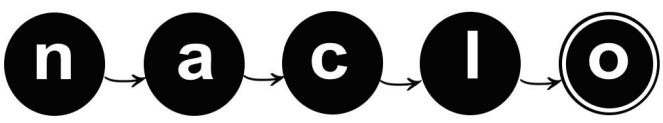
(F) Grammar Rules! (3/3)

F2. Not all of the sentences that this CFG can generate are actually sentences of English. For example, "The dog and the squirrel sits" can be generated but this isn't a correct sentence of English.

Give three more examples of sentences that can be generated by this CFG but are not correct English sentences; enter one word per box. Skip the periods at the end of the sentences.

A.						
B.						
C.						

F3. One of the rules in the CFG above is redundant: any sentence that can be generated using this rule can already be generated by a combination of other rules. Write the number of the redundant rule below.



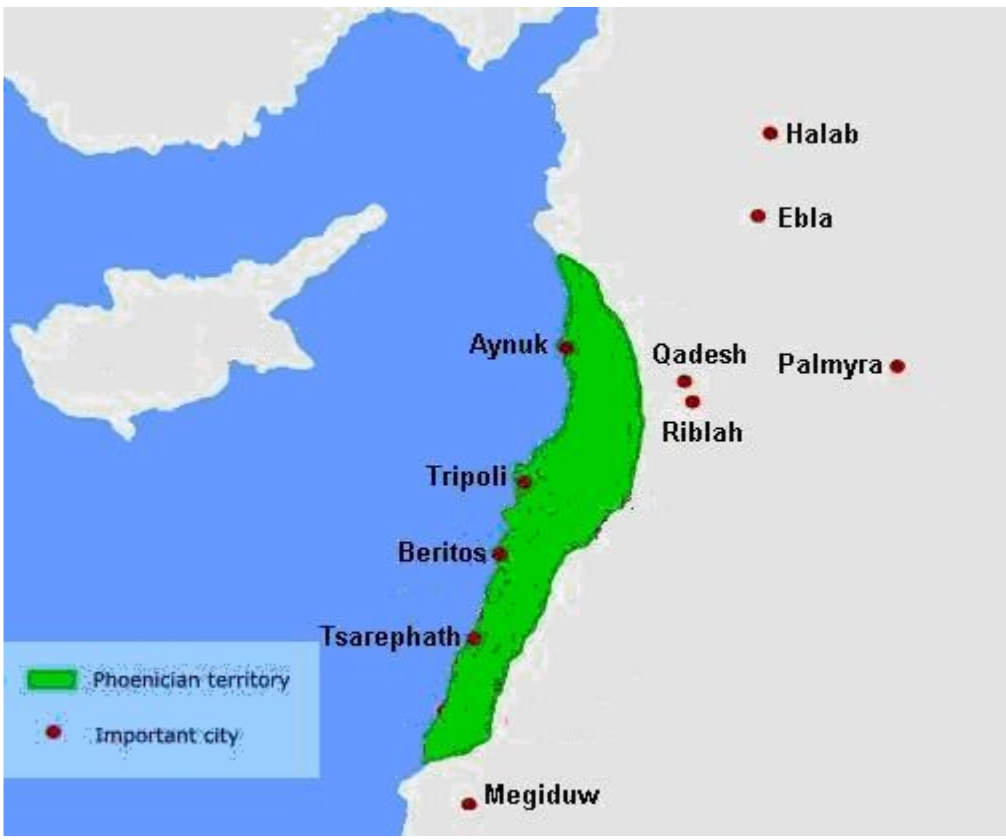
Do not scan this page

YOUR NAME: REGISTRATION #

(G) Phoenician Fun (1/2) [10 points]

The Phoenician script can be dated at around 1050 BCE, and from it the Arabic, Hebrew and by extension Greek, Roman, and Cyrillic scripts evolved.

The Phoenician civilization was centered along the Mediterranean coast in an area known as Cana'an. The map below shows a number of Phoenician cities and nearby cities that were important trading partners. The spellings reflect their pronunciation in Phoenician. However, two of the cities on the map are shown with their modern names which are very different from what they were called in Phoenician times.



On some black and white printers, the colors of the sea and the Phoenician territory are hard to tell apart. The Phoenician territory is the area that includes the following cities: Aynuk, Tripoli, Beritos, and Tsarephath.

G1. Match up the Phoenician names in the list below with the names on the map. Remember, two of the names will not match, so you should have two names left over.

.

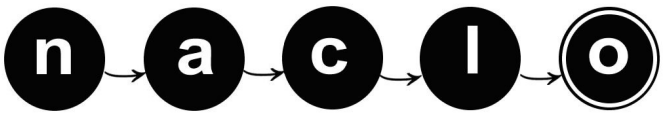
1	𐤑𐤓𐤐𐤈	6	𐤀𐤉𐤍𐤊
2	𐤇𐤋𐤁	7	𐤏𐤁𐤋
3	𐤓𐤁𐤋𐤄	8	𐤀𐤈𐤓
4	𐤌𐤂𐤃𐤅	9	𐤒𐤃𐤔
5	𐤈𐤃𐤌𐤓	10	𐤁𐤓𐤕𐤎

YOUR NAME: REGISTRATION #

(G) Phoenician Fun (2/2)

Fill in the corresponding number, or use the letter X if it is one of the two left-over cities.

Aynuk	
Beritos	
Ebla	
Halab	
Megiduw	
Palmyra	
Qadesh	
Riblah	
Tripoli	
Tsarephath	



Do not scan this page

YOUR NAME: REGISTRATION #

(H) Twodee (1/7) [20 points]

Spoken language is one-dimensional — the words of a sentence are pronounced in some order. Standard writing systems merely record that order. But spoken language is often ambiguous. Written language might be clearer if we used Twodee, a two-dimensional writing system that places words on the cells of a rectangular grid. Here are all the ways to write "Frogs eat flies" in Twodee:

Frogs	eat	flies.

Frogs	eat
	flies.

Frogs
eat flies.

Frogs
eat
flies.

Why so many layout options? If a sentence or other phrase consists of two sub-phrases A and B, where A is *spoken first*, then Twodee lets you *write* A either to the *left* of B (as in ordinary writing) or *above* B. If A or B has multiple words, then it will have sub-phrases of its own. The diagrams below reveal how the four Twodee sentences above were put together.

Frogs	eat	flies.

Frogs	eat
	flies.

Frogs	
eat	flies.

Frogs
eat
flies.

Clear writers of Twodee usually try to choose layout options that will help their readers identify the sub-phrases A and B. Make sure you align the top edges as shown when placing the A and B rectangles side-by-side, or align the left edges as shown when stacking the A and B rectangles top-to-bottom. No extra space (blank columns or rows) is allowed between the two rectangles.

To write Twodee, you do need to know how to divide the phrase you're writing into sub-phrases. That is mostly a matter of common sense, based on the intended meaning of the phrase. There are also some conventions, such as the traditional division of a sentence into subject ("Frogs") and predicate ("eat flies," which describes what frogs do).

While Twodee is intended as a better writing system, it can still be ambiguous. For example, the meaning of

The	fall	
leaves	each	year.

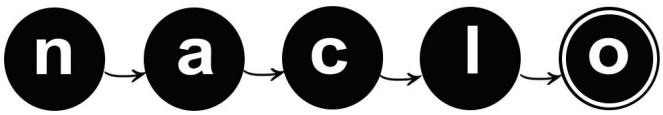
is not clear because the sentence could have been constructed in at least two plausible ways:

The	fall	
leaves	each	year.

(Spoken order: "The leaves fall each year.")

The	fall	
leaves	each	year.

(Spoken order: "The fall leaves each year.")



YOUR NAME: REGISTRATION #

(H) Twodee (2/7)

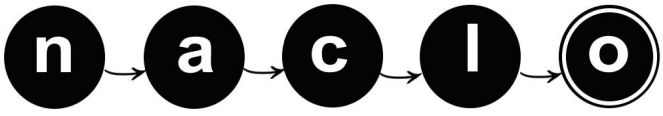
H1. For each Twodee sentence (on the left), use the circles to enter the numbers of all its possible meanings (paraphrased on the right). Some circles may be left blank. **Hint**: You may find it helpful to identify the sub-phrases of a Twodee sentence by drawing boxes as above.

A.

Jack and Jill TM went up the hill.	1. The trademarked group known as "Jack and Jill" ascended the hill.
Jack and Jill TM went up the hill.	2. Jack ascended the hill with his trademarked companion, "Jill."
Jack and Jill TM went up the hill.	

B.

The citizens hugged the and soldiers. cheered	3. The citizens hugged the soldiers and cheered the soldiers.
The citizens hugged and cheered the soldiers.	4. The citizens hugged each other and cheered the soldiers. 5. The citizens hugged the soldiers and cheered.



YOUR NAME: REGISTRATION #

(H) Twodee (3/7)

C.

John didn't marry his sweetheart
because she was rich.

6. It's not because his sweetheart was rich that John married her.

John didn't marry his sweetheart
because she was rich.

7. Because his sweetheart was rich, John didn't marry her.

D.

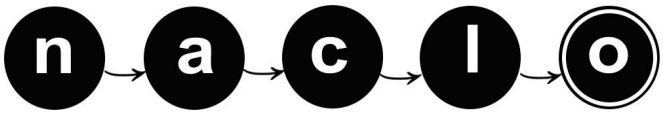
Alice
attacked the
scientists
with the
faulty data.

8. Alice used the faulty data to attack the scientists.

Alice
attacked the with the
scientists faulty data.

9. Alice attacked the scientists who had faulty data.

Alice attacked the
scientists
with the
faulty data.

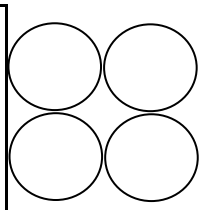


YOUR NAME: REGISTRATION #

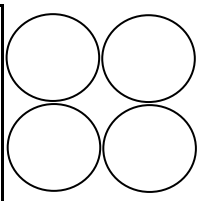
(H) Twodee (4/7)

E.

Heavy	farmers	are	forbidden
	and	on	this
	cattle		ramp.



Heavy	farmers	are	forbidden
and		on	this
cattle			ramp.



10. Heavy farmers and heavy cattle are forbidden on this ramp.

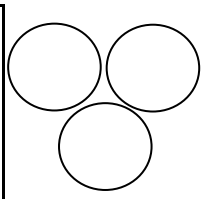
11. Heavy farmers and cattle farmers are forbidden on this ramp.

12. Cattle and heavy farmers are forbidden on this ramp.

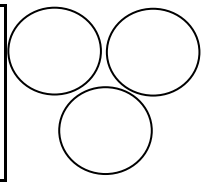
13. Cattle and heavy farmers are on this forbidden ramp.

F.

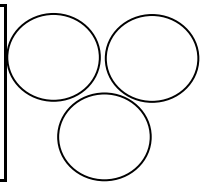
Please	shake	in
		the
		raisins.



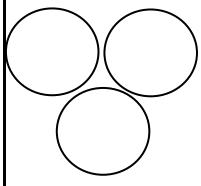
Please	shake	in
	the	raisins.



Please	shake	in
		the raisins.



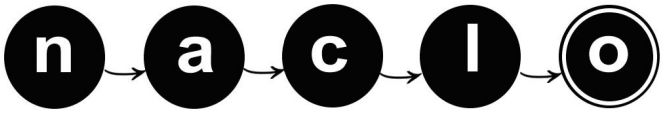
Please	shake	in
	the	
	raisins.	



14. Please add the raisins by shaking them.

15. Please get into the raisins and shake yourself.

16. If you are going to shake yourself, please do it in the raisins.

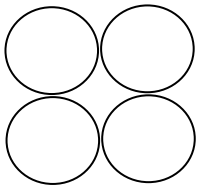


YOUR NAME: REGISTRATION #

(H) Twodee (5/7)

G.

I	was				
	spooked	out			
	by	the	dark	magician's	
				hut.	



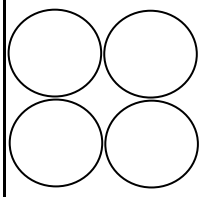
17. The hut of the dark magician spooked me out.

18. The dark hut of the magician spooked me out.

19. Out by the hut of the dark magician, something spooked me.

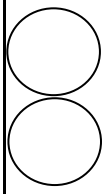
20. Out by the dark hut of the magician, something spooked me.

I	was				
	spooked				
	out	by	the	dark	hut.
				magician's	



H.

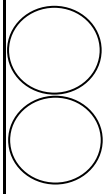
The	sold			
business	to			
	the	bank		
		bought	from	
			the	
			government.	

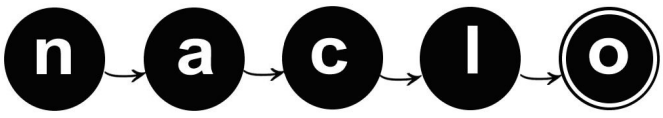


21. The business sold things to the bank that was bought from the government.

22. The business that was sold to the bank bought things from the government.

The			bought	from
business	sold			the
	to			government.
	the	bank		





YOUR NAME: REGISTRATION #

(H) Twodee (6/7)

H2. In April 2012, an entertainment reporter wrote:

> "Zooey Deschanel left her husband Death Cab for Cutie frontman Ben Gibbard."

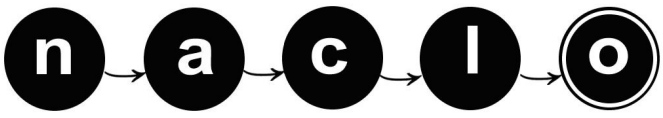
The former husband, Ben Gibbard, was the frontman for a band called "Death Cab for Cutie." However, an editor for ABC News incorrectly clarified the sentence and published the following:

> "Deschanel filed for divorce from husband Death Cab, throwing him over for Cutie frontman Ben Gibbard."

(The editor went on to add that "Deschanel and Cab were married for a little more than three years.")

By using Twodee, how could the reporter have written the original sentence to avoid this confusion? Place the words of the original sentence in 12 different cells of this two-dimensional grid (starting with "Zooey" at the top left).

Zooey								



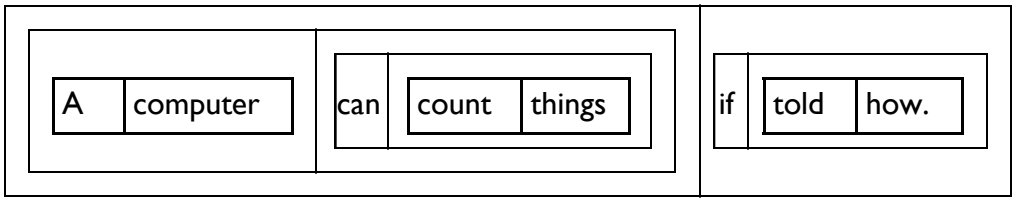
YOUR NAME: REGISTRATION #

(H) Twodee (7/7)

H3. Recall that multiple Twodee layouts were possible for "Frogs eat flies." Here is a completely horizontal Twodee sentence:

A	computer	can	count	things	if	told	how.

An editor is considering alternative arrangements of the words that might make the intended meaning more apparent. *How many ways* can this sentence be written in Twodee? Consider only ways that preserve the intended division into phrases and sub-phrases, for instance:



Answer: _____

H4. Here are some Twodee sentences whose spoken order is not obvious, as in the "leaves fall"/"fall leaves" example. In each case, how many different spoken orders are possible? (We have just used letters to stand for the words, since the particular words don't affect the answer.)

a	b	c
e		g
i	j	k

A. _____

a	b	c
e	f	g
i	j	k

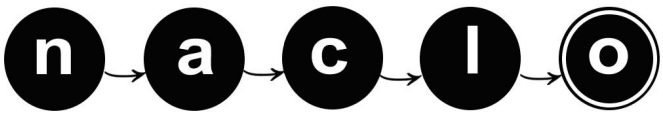
B. _____

a	b	c	d
e	f	g	h
i	j	k	l

C. _____

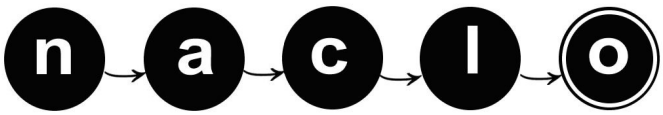
a	b	c	d
e	f	g	h
i	j	k	l
m	n	o	p

D. _____



YOUR NAME: REGISTRATION #

Extra Page - Enter the Problem Name Here: _____



Do not scan this page

NACLO 2013 Sites

As well as more than 90 high schools throughout the USA and Canada